2016

Self-Assembling Peptide Nanomaterials: Molecular Dynamics Studies, Computational Designs And Crystal Structure Characterizations

Huixi Zhang
University of Pennsylvania, z.huixi@gmail.com

Recommended Citation

Zhang, Huixi, "Self-Assembling Peptide Nanomaterials: Molecular Dynamics Studies, Computational Designs And Crystal Structure Characterizations" (2016). *Publicly Accessible Penn Dissertations*. 2662.
https://repository.upenn.edu/edissertations/2662

Self-Assembling Peptide Nanomaterials: Molecular Dynamics Studies, Computational Designs And Crystal Structure Characterizations

Abstract

Peptides present complicated three-dimensional folds encoded in primary amino acid sequences of no more than 50 residues, providing cost-effective routes to the development of self-assembling nanomaterials. The complexity and subtlety of the molecular interactions in such systems make it interesting to study and to understand the fundamental principles that determine the self-assembly of nanostructures and morphologies in solution. Such principles can then be applied to design novel self-assembling nanomaterials of precisely defined local structures and to controllably engineer new advanced functions into the materials. We first report the rational engineering of complementary hydrophobic interactions to control β-fibril type peptide self-assemblies that form hydrogel networks. Complementary to the experimental observations of the two distinct branching morphologies present in the two β-fibril systems that share a similar sequence pattern, we investigated on network branching, hydrogel properties by molecular dynamics simulations to provide a molecular picture of the assemblies. Next, we present the theory-guided computational design of novel peptides that adopt predetermined local nanostructures and symmetries upon solution assembly. Using such an approach, we discovered a non-natural, single peptide tetra-helical motif that can be used as a common building block for distinct predefined material nanostructures. The crystal structure of one designed peptide assembly demonstrates the atomistic match of the motif structure to the prediction, as well as provides fundamental feedback to the methods used to design and evaluate the computationally designed peptide candidates. This study could potentially improve the success rate of future designs of peptide-based self-assembling nanomaterials.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Chemistry

First Advisor
Jeffery G. Saven

Subject Categories
Chemistry

SELF-ASSEMBLING PEPTIDE NANOMATERIALS: MOLECULAR DYNAMICS STUDIES, COMPUTATIONAL DESIGNS AND CRYSTAL STRUCTURE CHARACTERIZATIONS

Huixi Violet Zhang

A DISSERTATION

in

Chemistry

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

Dr. Jeffery G. Saven

Professor of Chemistry

Graduate Group Chairperson

Dr. Gary A. Molander

Hirschmann-Makineni Professor of Chemistry

Dissertation Committee

Dr. Tobias Baumgart, Associate Professor of Chemistry

Dr. Ivan J. Dmochowski, Professor of Chemistry

Dr. Amish J. Patel, Reliance Industries Term Assistant Professor of Chemical and Biomolecular Engineering

ACKNOWLEDGMENT

Throughout the past five years, I have come to realize that the journey of a Ph.D. is a daunting one for the most part. It is full of frustrations, struggles and loneliness. Yet it is an invaluable experience that makes one grow in all aspects and evolve into a new person who thinks deeply and fundamentally about the knowledge in this world. This journey would be much worse, or at least greatly different, without many people. I am not confident that I could make this far if these people haven't appeared and shared a part of the past five years with me. I am greatly thankful for their companion, support and guidance along the way.

I would like to thank foremost my research advisor, Professor Jeffery G. Saven, who is the most brilliant and integrative scientist I personally know. He is such an inspiration and a role model for me as a researcher and a person. I am deeply influenced by his fundamental and precise way of thinking. He gave me the freedom to explore the scientific challenges in my research project, yet offered much-needed advices and wisdom when I felt stuck and wasn't able to move forward myself. He trained me with high standards to be an independent researcher who could think through a detailed problem and understand the broad impact of the research at the same time. I am extremely blessed to be the graduate student of such a nice and smart person.

I am also very grateful for my wonderful experimental collaborators at the University of Delaware, Professor Darrin Pochan, Professor Kristi Kiick, Dr. Frank Pochan, Michael Haider, Yu Tian and Dr. Sameer Sathaye. They had faith and trust in my work. Their expertise in material characterizations contribute tremendously to the successful publications of the collaborative research presented in the thesis.

Additionally, I greatly appreciate the time and advices my thesis committee: Prof. Ivan Dmochowski, Prof. Tobias Baumgart and Prof. Amish Patel have provided for my research projects. I also thank Ivan for the opportunity of my 1st-year rotation in his lab and the recommendation letters for various travel award applications.

I feel so lucky to be a part of the Saven group. My fellow group members are so friendly and kind to me all the time. Particularly, I would like to thank Dr. Christopher MacDermaid, who was my mentor when I first joined and passed along a lot of helpful research experience to me, as well as Matthew (Will) Eibling, who is a very capable experimentalist and played a critical role in getting the crystal structures of the peptides that I designed. My dear friends, Jose Villages and Krisha Vijayendran, who have endured my occasional complaints about the Ph.D misery and always cheered me up.

Last but not least, I want to thank my friends and family for the emotional support and love that keep me fighting in this journey and survive until now. When I am stressed, my caring Christian friends always send prayers and let me feel God's covenants and almighty love. Their comforting words have been a powerful source of positive energy for me to stay out the dark holes of depression. My loving parents, although they don't understand at all my research, they have always believed I should get the best education as possible and have provided unconditional support for that, which I am deeply grateful. My husband, John, who himself has a doctorate degree as well, can understand all my pains and struggles throughout the journey, and has offered most helpful, constructive, yet loving suggestions and endless support for me. I want to dedicate my thesis to all these people.

ABSTRACT

SELF-ASSEMBLING PEPTIDE NANOMATERIALS: MOLECULAR DYNAMICS STUDIES, COMPUTATIONAL DESIGNS AND CRYSTAL STRUCTURE CHARACTERIZATIONS

Huixi Violet Zhang

Jeffery G. Saven

Peptides present complicated three-dimensional folds encoded in primary amino acid sequences of no more than 50 residues, providing cost-effective routes to the development of self-assembling nanomaterials. The complexity and subtlety of the molecular interactions in such systems make it interesting to study and to understand the fundamental principles that determine the self-assembly of nanostructures and morphologies in solution. Such principles can then be applied to design novel self-assembling nanomaterials of precisely defined local structures and to controllably engineer new advanced functions into the materials. We first report the rational engineering of complementary hydrophobic interactions to control β-fibril type peptide self-assemblies that form hydrogel networks. Complementary to the experimental observations of the two distinct branching morphologies present in the two β-fibril systems that share a similar sequence pattern, we investigated on network branching, hydrogel properties by molecular dynamics simulations to provide a molecular picture of the assemblies. Next, we present the theory-guided computational design of novel peptides that adopt predetermined local nanostructures and symmetries upon solution assembly. Using such an approach, we discovered a non-natural, single peptide tetra-helical motif that can be used as a common building block for distinct predefined material nanostructures. The crystal structure of one designed peptide assembly demonstrates the atomistic match of the motif structure to the prediction, as well as provides fundamental feedback to the methods used to design and evaluate the computationally designed peptide candidates. This study could potentially improve the success rate of future designs of peptide-based self-assembling nanomaterials.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 General background

1.1.1 Existence of biomolecular-derived assemblies in nature

Natural biomolecular assemblies and derivatives are highly-ordered materials that are complex in structure and rich in function. For example, ferritin is a hollow protein assembled from 24 subunits, and is critical in iron storage and mineralization in living systems(*1*); viruses use highly-symmetrical, protein-coated vesicles for encapsulation and protection of genetic information, immunological evasion, target binding, and oligonucleotide delivery(*2*); S-layer proteins assemble as a major cell-wall component for protein protection, scaffolding and nutrient uptake in many bacteria(*3*); laminins undergo receptor-directed assembly on cell surfaces that are involved in activities such as cell differentiation, movement and signaling and are important for tissue survival(*4*, *5*); assembly of nacre defines growth and ordering of inorganic/organic hybrid phases essential for the superior mechanic and color features of shells(*6*). Despite the large variety and unique material properties, assembling processes in natural systems oftentimes are dependent on the environment or require external energy input. Synthetic systems can be engineered to overcome the barriers and requirements of assembling processes in nature. Therefore, it is of great interest to study and engineer biological and synthetic self-assemblies that are robust in spontaneous assembling processes for nanomaterial applications.

1.1.2 Overview and history of designed self-assemblies

Engineered nanomaterial self-assemblies are targeted to achieve ordered systems from disordered components with minimal human or machine intervention(*7*). Design approaches have employed synthetic molecules and inorganic-organic hybrids as building blocks targeting a broad range of applications in nanoelectronics, nanomachines, photovoltaics, molecular-level data storage and catalysis(*8*, *9*). Although numerous examples of self-assembling nanomaterials have been discovered since as early as 1950s, the design of complex, multi-dimensional, highly-ordered self-assemblies has become successfully realized only in the past two decades. The Murray group pioneered the design of diverse self-assembling binary and ternary nanoparticle superlattices via synthetic small-molecules(*10–14*). Additionally, synthetic polymers are largely employed in the designs of hierarchical self-assemblies(*15*), such as dendronized polymers and block copolymers amphiphiles. The Percec group are experts in dendronized polymer self-assemblies and have discovered the isomeric libraries of quasi-equivalent primary structures composing self-assembling dendrons and dendrimers that decode the 3D assembly architectures(*9*, *16*, *17*). The "self-assembly" of block copolymers is usually realizable through the careful tuning of the assembling environment such as solvents/non-solvents(*18*), pH and temperature, as well as block composition and length. For example, successful approaches to achieve block copolymer assemblies include "phase inversion"(*19*), living crystallization-driven self-assembly (CDSA)(*18*, *20*), and kinetic controls(*21*, *22*).

However, synthetic molecules, especially block copolymers, rarely have controllable sequences to precisely define structure and chemistry. Biomolecules, such as

DNA, RNA, peptides, and proteins, naturally have the three-dimensional structures and functions decoded in their primary sequences, therefore providing a sequence-controlled, alternative approach for novel designs of advanced self-assembling nanomaterials. DNA tiles and origami are established as a nanomaterials field, where both empirical knowledge and computer prediction allow the design of proper base pairing and hybridizations to build targeted, complex nanostructures(*23–25*). Although the DNA nanostructures can be highly-ordered, there are still limitations in their finite assembly size and dimensionalities in all cases(*23*). Only the DNA tensegrity triangle has produced substantial 3D crystals thus far(*26*). DNA's primarily function in nature is to encode genetic information. This relatively low-level chemical and functional complexity is another limitation of DNA self-assemblies compared to proteins- and peptides-based assemblies. Additional attachments to DNA self-assemblies are necessary to achieve functional diversity. For example, the Mirkin group has further functionalized DNAs into spherical nucleic acid (SNA) self-assemblies through multivalent hybridization with nanoparticles, creating superlattices of various architectures(*27–29*), and further producing precisely engineered optical properties(*30*).

Because of such functional limitations, nature chooses proteins, rather DNAs, for most of its functional molecular assemblies. Proteins and peptides have an expanded building-block library of twenty natural amino acids rather than four bases used in DNAs. For the past decade, more research efforts have been spent in the area of de-novo design of peptide- and protein-based self-assemblies than that of DNA assemblies. The designed assemblies span various geometrical, mechanical, and functional properties, and range from hydrogel networks(*31*) to solid biomaterials(*7*), and can form various hierarchical nanostructures, such as one-dimensional tubes(*32*), rods(*33*) and fibrils(*31*, *34*), two-dimensional sheets (*35–38*) and three-dimensional cages(*39–43*) and crystals (*44–46*). Because proteins are rich in functions, it has become possible to utilize well-designed complex nanostructures to spatially distribute chemical functionalities in the controlled amino acid sequences, although the process remains more challenging for synthetic polymers and DNAs. Recent advances include a self-assembling biomolecular hydrogenase enzyme that has improved catalytic activity and resistance to protease, heat and oxygen due to the ordered encapsulation(*47*). Additionally, controlled and precise patterning of hybrid organic–inorganic structures has been successfully realized via designed solution-phase protein self-assemblies, where metal ions(*48*), and functional small molecules(*45*) co-assemble into hierarchical, highly-ordered nanostructures. Lastly, the Mayo group and the Mirkin group reported designed self-assembly hybrids between proteins and DNAs(*49*, *50*) that lay the foundation for the engineering of a new class of bio-nanomaterials.

1.2 Motivations of designing peptide-based self-assemblies

Proteins and peptides are structurally-complex building blocks that nature provides for self-assembling nanomaterial. Their structural complexity comes from the rich primary sequence of amino acids, well-defined secondary structures such as α-helices and β-sheets, specific intramolecular interactions in tertiary structures, and from the intermolecular interactions which guide the ordered, sometimes symmetrical, quaternary structures. Such innate structural complexity of proteins and peptides offers advantages for constructing self-assembling nanomaterials.

1.2.1 Advantages of designed peptide self-assemblies

Protein- and peptide-based self-assemblies and their derivatives have advantages over synthetic organic polymers and DNA in multiple aspects: 1) controlled structural output through primary sequence; 2) a large pool of amino acid monomers, including unnatural amino acids; 3) accessibility to post-assembling chemical modifications; 4) high bio-compatibility in-vivo. For peptides, which typically have a short primary sequence, there are additional advantages, including straightforward synthesis through standard solid-phase techniques, easier to design into self-assemblies than large proteins, and highly compatible to nonbiological modifications.

Given these advantages, a variety of biomedical, chemical and energy related applications of peptide-based self-assemblies have been reported. Hydrogel networks cross-linked by peptide assemblies are promising artificial extracellular matrices for cell culture, tissue engineering and regenerative medicine(*31*, *51*, *52*). These hydrogel networks could also serve as carriers for controlled delivery and release of drugs(*53*, *54*) and functional proteins(*55*). Peptide-based self-assemblies rich in arginine and lysine residues have also been proved to have antimicrobial or antibacterial effects, which could be applied to treat infections(*56*, *57*). Enzyme-responsive peptide self-assemblies have shown efficacy in cancer therapeutics(*58*–*60*). Fusion peptide self-assemblies with enzymes have assisted in improving protease and esterase catalytic activity(*61*) and offered a simple and cost-effective strategy to stabilize and reuse carbonic anhydrase(*62*). Besides that, peptide-inorganic hybrid self-assemblies have been designed towards the development of more efficient nano-electronics and solar cells, such as peptide-

DDP(Zn)-TiO2 hybrid assemblies capable of photoinduced charge separation with spatial control of the donor-acceptor pair by tuning the peptide sequence(*63*).

Many past successful designs of highly-ordered self-assemblies rely on scaffolds from natural proteins, where the interior/core structure of the protein is adopted to preserve its globular fold (tertiary structure), and only interfaces between proteins are designed and modified(*35–38*, *43*, *46*, *64*). The *de novo* designs that use arbitrary protein backbones and non-natural sequences without existing references to nature still remain very challenging for predetermined molecular- and nano-structures. The Baker group tackles the problem in the realm of novel isolated proteins such as tandem protein repeat architectures and the TIM-barrel fold through geometric and chemical rules(*65*, *66*). Peptides have fewer degrees of freedom than proteins due to short primary sequences and are a reasonable starting point to engineer novel tertiary folds and quaternary assembly architectures from scratch, and design sequences accordingly using rational and computational approaches. However, the reported successful designs of ordered peptide self-assemblies with high control and precision are limited to peptide oligomers, like helical barrels(*67*), and a cage-shaped nanostructure(*68*). Therefore, there is still large room of improvement regarding designing novel and non-natural peptide-based self-assemblies with predefined nanostructures. Some of the chapters in this thesis will present results to address the engineering of lattice-forming peptides of non-natural sequences and backbones by computational design approaches. Next, we will review rational and computational approaches used in the designs that will appear in the later chapters of the thesis.

1.2.2 Computational design vs. rational design

The design methodology for protein- and peptide-based self-assemblies can be grouped into two categories: rational approaches and computational approaches. Rational designs combine biomimicry and chemical intuitions to engineer self-assembling biomaterials. In some cases, natural self-assembling sequence patterns are adopted, such as amyloid beta and precursor proteins(*61*, *69*, *70*), leucine-zippers(*71*, *72*) and laminin- and collagen-derived motifs(*73*, *74*). Other times peptide amphiphiles are designed based on well-understood chemical interactions between the polar and nonpolar parts of the sequences(*75*, *76*). Self-assemblies designed by those strategies usually adopt irregular nanostructures lack of long-term periodicity because of the less-controlled, nonspecific associations between the flexible peptide building blocks. These structures are typically observed in peptide-based hydrogels. To have a better control of the assembly nanostructures and to engineer defined long-term order, an alternative rational design strategy employs natural proteins of known structure and symmetry, and fuse them together through mutagenesis(*77*)'(*39*), or through simple covalent linkages such as disulfide bonds. Although these attempts have successfully produced the defined nanostructures, the selection of assembly building blocks is still constrained by a small portion of known proteins in nature. Thus the accessible nanostructures are limited. On the other hand, the computational approach, guided by theory and biomolecular models of the defined assembly nanostructures such as rods, sheets, cages and lattices, can achieve precise control and fine tuning of the building-block geometry and orientations. Furthermore, it is possible to achieve atomistic-level precision with the designed assembly nanostructures using a computational design approach(*40*, *44*, *46*, *66*). More specifically, given an assembly nanostructure, computational design identifies energy minima in a large conformational and sequence landscape where noncovalent, interatomic interaction energetics are calculated within the building-block proteins and at the protein-protein interfaces. However, such interactions are subtle to design as the energetics are approximated. Therefore, finding sequences that will fold and self-assemble into a particular nanostructure with predetermined order and symmetry in the conformational and the sequence energy-landscape is a non-trivial task.

One critical step in the computational design of protein- and peptide-based self-assemblies is to find such primary sequences able to fold into target tertiary structures and assemble into targeted nanostructures of higher order. The target tertiary structure can be taken from the backbone of natural proteins or created using mathematical models(*78, 79*). Additional constraints can be imposed on the design, such as symmetric approximations(*80*) between the neighboring monomer building blocks, to improve the computational efficiency when designing infinite lattices or symmetrical cages. There is typically no unique answer to what such sequences should be, since nature gives many examples where sequences sharing no similarities fold into nearly identical structures(*81*)’(*82*). In order to find those sequences, approximations of folding free energy, derived from appropriate physical models, can be used as an objective function and minimized by optimization techniques. Due to the tremendous number of amino acid combinations possible in only a medium-sized protein sequence, and the various conformations each amino acid could possess, it is crucial to have computationally efficient methods to solve the functions that model the energetics of designed structures accurately. Lazaridis & Karplus(*83*) and Mendes et al(*84*). have categorized the energy models into three types: 1) Statistical effective energy functions, such as the environmental energy model(*85*, *86*) and the helix propensity model(*87*); 2) Empirical effective energy functions achieved by machine-learning(*88–92*); 3) Physical effective energy functions, which use atom-level force field such as CHARMM(*93*), AMBER(*94*) and OPLS(*95*). A combination of these three types of energy models are usually incorporated in the computational approach at different design stages.

Another important aspect of computational design is the balance between computational efficiency and sampling thoroughness when facing the huge phase space of possible sequences. Heuristic techniques are efficient for sampling the whole structural and/or sequence space; for example, Monte Carlo simulated annealing(*96*)(*97*), and genetic algorithms(*98*, *99*) have been widely used in computational designs. The drawback of heuristic search is that finding the global minimum of the energy objective function is not guaranteed. Instead, the objective function could be trapped in one of the local minima. To overcome this issue, another set of algorithms applying dead-end elimination or its generations(*100–102*) have been developed to search through a subset of the whole phase space and to identify the designed sequence at global minimum. However, one has to keep in mind that dead-end elimination is very computationally expensive. An alternative to explicit sampling is to apply computational design via a statistical-mechanics based probabilistic approach, which characterizes the energetics of an ensemble of designed sequences in a given nanostructure using the mean-field energy theory(*103*, *104*). In this approach, the probabilities of amino acids at the designed peptide sites are estimated by minimizing the free energy or maximizing the entropy of the sequence ensemble using nonlinear optimization techniques that identifies local energy minima. Additional Monte Carlo sampling or the grid search techniques for backbone structures with different nanostructure configurations can be combined into the probabilistic approach to design peptide assemblies of defined symmetry and crystalline order(*44*).

1.3 Thesis Contributions

This section is a summary of the set of key results and findings of the following chapters: Chapters 2, 3, and 4, in this thesis.

In section 2.1-2.2, we will introduce *de-novo*, rationally designed, self-assembling peptides MAX1 and LNK1. MAX1 and LNK1, despite their similar sequence patterns, are capable of self-assembling into distinct β-fibril nanostructures respectively and form hydrogel networks. In section 2.3, experimental characterizations for the two peptide self-assemblies are detailed. In section 2.4, computational and modeling details for the two peptide self-assemblies are described. Section 2.5-2.6 shows the key results of the computational and experimental work followed by discussion. In brief, transmission electron microscopy (TEM) reveals branched fibrils of MAX1 and rigid rod-like fibrils of LNK1 with no branches (section 2.5.1). MAX1 and LNK1 hydrogel assembly exhibit very different rhelogical behaviors (section 2.5.2). We constructed molecular models of the two peptide self-assemblies and use molecular dynamics (MD) simulations to obtain dynamic and molecular structural information and to understand the driving force for their distinct assembly features. Each peptide is a single β-hairpin with the same hydrophobic patterning featuring a hydrophobic face and a hydrophilic face, only the identities of the hydrophobic amino acids differ. MD simulations and quantitative analysis of the fibril models are presented to provide molecular insights into the fibril structures and the fluctuations of the two self-assemblies (section 2.5.3), which can inform the rational design of new peptide-based materials.

Section 3.1-3.6 presents a set of computationally designed peptide self-assemblies with four predetermined nanostructures, an isolated helical bundle and three crystalline nanosheets with different local symmetries. Section 3.3.1 describes the computational design details of the 29-residue peptides guided by statistical-mechanics based theory. We encoded the information required for multiple, distinct nanostructures into a novel sequence of a single bundle motif as a common building block for four distinct, predetermined material nanostructures. These predetermined structures were specified by the design of the exterior sequences of the helical bundles motifs. Section 3.3.2 -3.3.3 shows the TEM and small angle neutron scattering (SANS) characterizations of the nanomaterials composed of the designed peptides upon solution assembly. The peptides self-assemble into distinct nanostructures (non-assembling bundles and nanosheets) with various degrees of agreement to the predefined local symmetry and repeating-unit size. Furthermore, the material solution assembly process is robust with respect to both variation of solution conditions (pH and temperature) and covalent modification of the computationally designed peptides. Section 3.4 presents additional modeling to address and discuss the possible lattice configurations formed in two of the nanosheet materials that appear to deviate from the designed local structures in the solution assembly.

Lastly, in section 4.1-4.7, we expand our study on one of the computationally designed self-assembling peptides from Chapter 3, and report the crystal structure and more solution phase characterizations of the lattice nanostructure. Section 4.4.1 recaps the computational design strategy in brief, followed by detailed description of X-ray crystallographic and small angle x-ray scattering (SAXS) characterizations of the peptide lattice in section 4.4.2. In section 4.4.3, we compare the crystal structure of the peptide lattice to the design model. The crystal structure reveals that the helical bundle motif matches to the design with atomistic precision. Such motif is applicable as a common building block to design different peptide lattices. Section 4.4.4 investigates the lattice structures assembled under different solution conditions and shows the consistency between these structures and robustness of the assembly. Additional molecular modeling and energetics calculated on peptide assembly consistent with the crystal structure configuration are reported in section 4.4.5. We discuss in section 4.5 the successful design of the helical bundle motif, and the possible causes of disagreement between the designed and the observed lattice structure from a thermodynamic perspective. In particular, we emphasize the critical role of specific hydrophobic interactions in stabilizing the peptide bundle-bundle interfaces in the tightly-packed crystal assembly. This discussion also suggests fundamental ways to potentially improve the success rate of future computational design of protein- and peptide- based lattices.

List of Publications:

H. V. Zhang*, F. Polzer*, M. J. Haider, Y. Tian, J. A. Villegas, K. L. Kiick, D. J. Pochan, J. G. Saven, "Computationally designed peptides for self-assembly of nanostructured lattices". Science Advances. (2016). Under review. * these authors contributed equally

H. V. Zhang, M. Eibling, Y. Tian, K. L. Kiick, D. J. Pochan, J. G. Saven, "Crystal Structure of a Computationally Designed Lattice-forming Peptide". In preparation. (2016)

S. Sathaye*, H. V. Zhang*, C. Sonmez, J. P. Schneider, C. M. MacDermaid, V. D. Christopher, J. G. Saven, D. J. Pochan, "Engineering Complementary Hydrophobic Interactions to Control β-Hairpin Peptide Self-Assembly, Network Branching, and Hydrogel Properties", *Biomacromolecules*. **15**, 3891–900 (2014). * these authors contributed equally

CHAPTER 2 Engineering Complementary Hydrophobic Interactions to Control β-Hairpin Peptide Self-Assembly, Network Branching, and Hydrogel Properties

Adapted with permission from *Biomacromolecules*. 15, 3891–900 (2014). Copyright (2014) American Chemical Society.

2.1 Abstract

The MAX1 β-hairpin peptide (VKVKVKVK-V^{D}PPT-KVKVKVKV-NH_2) has been shown to form nanofibrils having a cross-section of two folded peptides forming a hydrophobic, valine-rich core, and the polymerized fibril exhibits primarily β-sheet hydrogen bonding(*105–111*). These nanofibrils form hydrogel networks through fibril entanglements as well as fibril branching(*112*). Fibrillar branching in MAX1 hydrogel networks provides the ability to flow under applied shear stress and immediately reform a hydrogel solid on cessation of shear. New *β*-hairpins were designed to limit branching during nanofibril growth because of steric specificity in the assembled fibril hydrophobic core. The nonturn valines of MAX1 were substituted by 2-naphthylalanine (Nal) and alanine (A) residues, with much larger and smaller side chain volumes, respectively, to obtain LNK1: (Nal)K(Nal)KAKAK-V^{D}PPT-KAKAK(Nal)K(Nal)-NH_2. LNK1 was targeted to self-associate with a specific "lock and key" complementary packing in the hydrophobic core in order to accommodate the Nal and Ala residue side chains. The experimentally observable manifestation of reduced fibrillar branching in the LNK1 peptide is the lack of solid hydrogel formation after shear in stark contrast to the MAX1 branched fibril system. Molecular dynamics simulations provide a molecular picture of

interpeptide interactions within the assembly that is consistent with the branching propensity of MAX1 vs. LNK1 and in agreement with experimental observations.

2.2 Introduction

Specific molecular recognition interactions within peptide and protein molecules have been used widely for designing smart, responsive hydrogel materials. Prominent examples of such efforts include materials based on specific interactions between coiled coil domains such as leucine zipper domains, interactions between ditryptophan (WW) and proline-rich domains, standard linear peptides based on purely α-helical structures in addition to tetratricopeptide repeat (TPR)−peptide interactions(*113–118*). A specific type of protein−protein interaction, named the “Lock and Key” mechanism, involves recognition between specific molecules with complementary steric binding domains. This specific steric packing has been studied extensively in proteins but not toward designed materials development(*119–122*). The lock and key analogy was first put forward by Emil Fischer more than 100 years ago specifically to describe specificity in enzyme−substrate interactions(*123*). An example of widely studied, shape-dependent lock and key type interactions is the ligand protein interaction between the vitamin biotin and the egg white glycoprotein avidin, which is of tremendous interest in biotechnological applications(*124*). Similarly, Holzinger et al. have reported complexation between biotin and β-cyclodextrin as a representation of a new bio-receptor immobilization affinity system(*125*). Among the related shape-dependent recognition patterns involving proteins, interactions between proteins and DNA based on DNA local shape variations (individual base pair and minor double helix region) and DNA global shape variations (various helical topologies and deformations) have also been widely reported(*126*). Solution-assembled fibrillar networks based primarily on hydrophobic and hydrogen bonding interactions between β-hairpin peptides have been studied extensively(*105*, *108*, *110*). In this work, we discuss experimental and simulation studies of the hydrogel network behavior of β-hairpin peptide-based hydrogels and the impact of designed hydrophobic and steric interactions at the cores of the fibrils forming the networks.

Hydrogen bonding-dominated assemblies of peptides into linear nanostructures include natural nanostructures formed by amyloid and amyloid-like assembly of proteins and polypeptides, as well as synthetic peptides demonstrating uniform, linear, and unbranched morphologies such as nanofibrils, nanotapes, nanoribbons, nanobelts, nanotubes and many hydrogel networks based on such morphologies(*127–135*). A prominent example of hierarchical fibrillar self-assembly of peptides based on hydrophobic interactions and hydrogen bonding has been reported by the Pochan and Schneider research groups. These groups have studied the self-assembly of MAX1 (VKVKVKVK-V^{D}PPT-KVKVKVKV-NH$_2$) and related peptide sequences(*105*, *108*, *110*). MAX1 is an amphiphilic 20 amino acid residue peptide with alternating hydrophobic valine and hydrophilic lysine residues with a -VDPPT- turn sequence in the middle. Found in random coil conformations in aqueous solution with neutral to low pH due to repulsion between positively charged lysine side chains, folding into the β-hairpin conformation, and consequent intermolecular assembly can be triggered by modulating solution conditions such as increasing pH (pH ~ 9)(*105*, *136*, *137*), increasing ionic strength (e.g., ~ 150 mM NaCl)(*108*, *111*, *112*), and increasing temperature (*T* =37 °C)(*106*). The higher pH and high ionic strength conditions deprotonate lysines or screen interactions between lysines, respectively, thus allowing the turn sequence to force the peptide arms to arrange in an antiparallel conformation, the *β*-hairpin conformation. The *β*-hairpin is stabilized by significant hydrogen bonding between the beta-strands of the peptide as well as the conformation of the turn sequence that anchors the arms. Additionally, folding and assembly can be affected by a rise in temperature that induces the hydrophobic interactions both within and between the peptides both promoting folding and intermolecular assembly. The hierarchical self-assembly of these *β*-hairpins into uniform fibrils takes place due to several interactions. Facial hydrophobic interactions between the valine faces of two hairpins collapsed together form the cross-section of a growing fibril (Figure 2.1). Additionally, lateral intermolecular hydrogen bonding and additional hydrophobic interactions between folded hairpins define the axis of the growing fibrils(*110*). After assembly, MAX1 forms self-standing hydrogel networks that are purely physically cross-linked. The facial hydrophobic collapse at the core of the growing fibrils sometimes results in formation of a defect characterized by potentially incomplete burial of the hydrophobic valine side chains and sliding of the layers at the bilayer interface in a manner that disrupts extension of the linear fibril. Such defects lead to the nucleation of a branch point in the fibril growth leading to two daughter fibrils extending from the branch point(*112*). These branch points contribute to physical cross-linking of the hydrogel network in addition to fibrillar entanglement. The defects in hydrophobic face packing of folded, opposing hairpins in a fibril, and the consequent branch point/new crosslink point that is formed, can be partially attributed to the lack of specificity in the facial hydrophobic interactions between peptides due to the uniform steric volumes of the valine side chains. In this paper, we attempt to introduce "lock and key" type specificity in the facial hydrophobic interactions of the MAX1 peptide in an attempt to significantly limit the formation of branching cross-links formed as a result of nonspecific hydrophobic collapse.

Several variants of MAX1 have been designed and studied. These variants have different primary sequences and have been developed to incorporate different functionalities such as faster gelation kinetics(*136*, *137*), photo-cross-linkable hydrophilic side chains(*138*), inherently antibacterial properties(*139*), and swapped positions of hydrophobic and hydrophilic residues(*110*). They all undergo hierarchical self-assembly in a manner very similar to MAX1 resulting into a uniform fibrillar nanostructure. These functional variants have been designed by varying the hydrophilic side chains of MAX1. Each of these peptides has a nonspecific valine hydrophobic face like MAX1. Thus, designed modifications to the hydrophobic face of MAX1 offer relatively unexplored, fertile ground to the study of self-assembly and network behavior of the resulting peptides.

Herein a designed peptide, (Nal)K(Nal)KAKAK-V^{D}PPT-KAKAK(Nal)K(Nal)-NH_2, is presented and studied, wherein four valine residues of MAX1 (VKVKVKVK-V^{D}PPT-KVKVKVKV-NH_2) have been replaced with non-natural 2-naphthylalanine (Nal) amino acid residues, whose side chains possess larger steric volumes than valine. The middle, nonturn valines of MAX1 have been replaced by alanine (A) residues, which has a smaller side-chain steric volume than valine. Thus, the hydrophobic surfaces of two LNK1 hairpins can pack specifically into a lock and key type structure in the hydrophobic core. Such complementary steric interactions are often associated with well-defined structures in proteins (Figure 2.1). This is in stark contrast to the simple hydrophobic collapse in MAX1 and the less specific interactions associated with an interface that comprises only valine residues. Thus, fibrils formed from LNK1 peptide self-assembly are intended to be unbranched (Figure 2.1) as compared to the branched fibrils of MAX1 formed from incomplete, defective collapse of non- specific valine faces during β-hairpin assembly. We hypothesize that LNK1 fibrils form percolated networks only by fibril entanglement as opposed to the hydrogel networks of MAX1 that form a network due to fibril branching as well as entanglement. Due to the hypothesized severe inhibition in the branching in the LNK1 fibrillar hydrogel networks, these materials are expected to have a significantly different response to shear treatment as compared to the MAX1 network hydrogels. Specifically, the LNK1 hydrogel networks are expected to undergo flow and fibril fracture and disentanglement when subject to shear treatment but lack hydrogel reformation ability post-shear due to fibril collapse and permanent loss of a percolated network. We report on the assembly of the LNK1 peptide, the local fibril nanostructure, and ultimate hydrogel network structure via a combination of (i) physical characterization techniques such as circular dichroism (CD), transmission electron microscopy (TEM) and oscillatory rheological measurements as well as (ii) molecular dynamics simulations to complement the experimental observations.

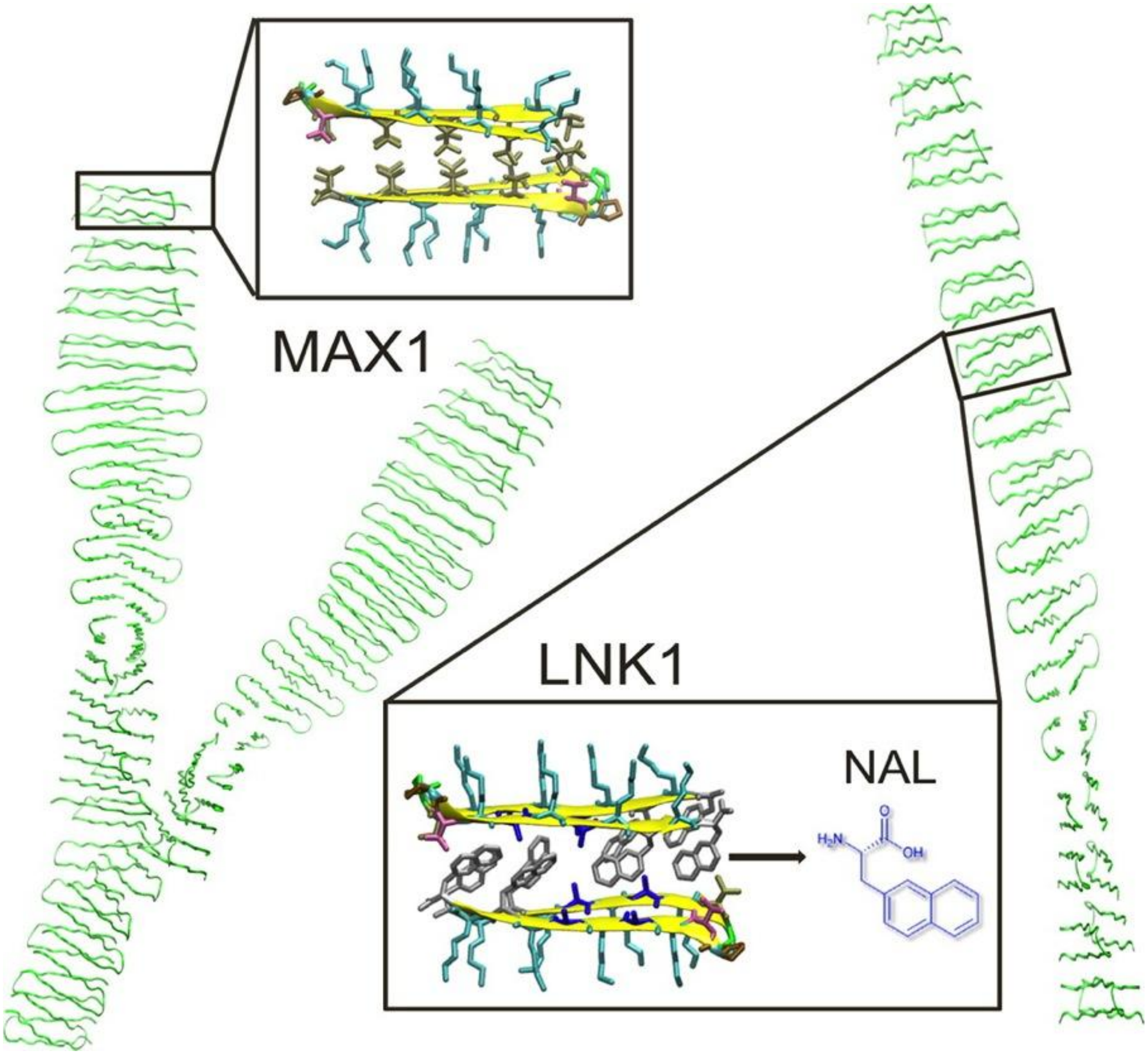


Figure 2.1. Ribbon representations of MAX1 and LNK1. MAX1 can form branched nanofibrils. LNK1 is restrained to form only linear fibrils. The inserts are views of fibrillar cross sections. In MAX1, valine residues form a nonspecific, “flat” hydrophobic interface. In LNK1, specific hydrophobic steric interactions appear as well as the interior hydrophobic interface between complementary naphthylalanine and alanine side chains. Adapted with permission from *Biomacromolecules* (2014).

2.3 Materials and methods

Peptide Synthesis. Peptides MAX1 and LNK1 were designed at the University of Delaware and synthesized and purified by New England Peptide, LLC (Gardener, MA, USA) where matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectral analysis data confirming molecular weights of purified peptides were obtained. Both peptides were prepared using Fmoc-based solid phase peptide synthesis as described elsewhere(*140*).

Hydrogel Preparation. A 1 mg aqueous solution of either MAX1 or LNK1 peptide in 100 μL of deionized chilled water (5 °C) is prepared leading to a 1% (w/v) aqueous solution. An equal amount of chilled (5 °C) buffer solution pH 9 (250 mM boric acid, 20 mM NaCl) is added to both aqueous solutions to give buffered solutions of either MAX1 or LNK1. Then the temperature of the solutions is raised to 30 °C to obtain 0.5% (w/v) hydrogels with effective solution conditions pH 9 (125 mM boric acid, 10 mM NaCl). The same procedure is used to obtain networks from MAX1 and LNK1 peptides using a different solution condition using a pH 7 buffer (100 mM bis-tris propane, 300 mM NaCl), ultimately leading to pH 7 (50 mM bis-tris propane, 150 mM NaCl). Briefly, an equal volume of chilled (5 °C) pH 7 (100 mM bis-tris propane, 300 mM NaCl) buffer solution is added to a chilled (5 °C) 1% (w/v) aqueous solution of either MAX1 or LNK1. The temperature is immediately raised to 30 °C to obtain 0.5% (w/v) hydrogels.

Circular Dichroism. CD spectra were collected using an AVIV Model 420 (AVIV Biomedical, Inc. Lakewood, NJ, USA) CD spectrophotometer. Solutions of MAX1 and LNK1 (150 μM) at pH 9 and 10 mM NaCl were prepared by adding equal volumes of chilled (5 °C) buffer solution of pH 9 (250 mM boric acid, 20 mM NaCl) to 300 μM deionized peptide solution. The random coil to β-hairpin folding transition temperatures were determined by scanning temperatures from 15 to 60 °C keeping the wavelength of incident radiation fixed at 218 nm, the signature wavelength at which a significant drop in mean residue ellipticity indicates the formation of β-sheet secondary structure. Mean residue ellipticity $[\theta]$ was calculated from the equation $[\theta]= \theta_{obs}/(10l \times c \times n)$, where θ_{obs} is the measured ellipticity (millidegrees), l is the path length of the cell (cm), c is the peptide concentration (molar), and n is the number of residues on the peptide sequence. Temperature scans were performed with 2 °C increments and 5 min equilibration time at each temperature.

Oscillatory Rheology. Oscillatory rheology measurements were performed on an ARG2 rheometer (TA Instruments, New Castle, DE, USA) using 20 mm diameter stainless steel parallel plate geometry. For the initial time sweep measurements, the samples for both LNK1 and MAX1 were prepared as follows. Buffered peptide solutions were prepared in ice-chilled conditions by adding 120 μL of chilled (5 °C) pH 9 (250 mM boric acid, 20 mM NaCl) buffer to 120 μL of 1% (w/ v) of peptide solution in chilled (5 °C) deionized water. The chilled (5 °C) buffered peptide solution was quickly transferred to the Peltier plate of the ARG2 rheometer equilibrated at 5 °C, and the upper plate was lowered to a gap height of 500 μm. The upper plate and the Peltier plate were equilibrated to 35 °C prior to carrying out the rheological experiments. Oscillatory time sweep measurement steps carried out for 90 min each, before and after subjecting the hydrogel networks to a steady-state shear of 1000/s for 120 s, were carried out for both MAX1 and LNK1. Throughout the oscillatory time sweep measurements, the oscillatory frequency was maintained at 6 rad/s and oscillatory strain at 1%. The gap height was maintained at 500 μm for both the steady-state shear and oscillatory measurements. Further characterization of network properties of both MAX1 and LNK1 networks was carried out using oscillatory frequency sweep measurements at a constant 1% oscillatory strain. Prior to the frequency sweep measurements both hydrogel networks were allowed to assemble inside syringes by pulling buffered solutions (pH 9 125 mM boric acid, 10 mM NaCl) of both peptides into syringes then maintained at 35 °C. The hydrogels were then subjected to a multiple injection treatment that involved a sequential injection of MAX1 or LNK1 networks formed inside a syringe, seven times through a 27−1/ 2 G needle. For future reference to this method within this thesis, it shall be referred to as the "Injection Shear" treatment.

Transmission Electron Microscopy. Transmission electron microscopy was carried out on a 120 kV Tecnai-12 Electron Microscope (FEI Company, Hillsboro, OR, USA). MAX1 or LNK1 hydrogel was prepared at 0.5% (w/v) with final hydrogel conditions of pH 7 (50 mM bis-tris propane, 150 mM NaCl). To observe the fibrillar width, particularly the local nanostructure of LNK1 vs MAX1 networks, 10 μL of gel was diluted to a concentration of 0.1% (w/v), and a drop was placed on a 300 mesh copper-coated grid (Electron Microscopy Sciences, Hatfield, PA, USA) held by a pair of tweezers. Excess fluid was blotted off with filter paper. Then, immediately, 3 μL of a 1% (w/v) of uranyl acetate solution in water was placed on the grid and blotted off after 40 s. The grid was left to dry for an hour and used for imaging. For the preparation of the sample for MAX1 and LNK1 networks after being subject to the injection shear treatment as described above, a small piece of the treated hydrogel without dilution was placed on a 300 mesh copper-coated grid, and the excess volume of gel was blotted off. Three microliters of a

1% (w/v) of uranyl acetate aqueous solution was placed on the grid and blotted off after 40s to stain the sample. This method was applied to the network samples after injection shear treatment to capture their morphology without dilution and additional mixing effects.

2.4 Computational modeling

2.4.1 Preparation of initial peptide and fibril structures

The initial structures of both MAX1 and LNK1 were constructed as octamers of β-hairpins, forming two layers of β-sheets (a bilayer) with hydrophilic lysine residues exposed on the fibril exterior and the hydrophobic residues (valine in MAX1; 2-naphthylalanine and alanine in LNK1) buried within the bilayer at the interface between the two β-sheets. The construction of the initial model involved three-steps. First, the coordinates of the backbone atoms of two 8-residue β-strands were generated de novo, consistent with trans amide bonds and a pleated β-sheet ($\varphi = -135°$, $\psi = 135°$). Individual antiparallel β-strands were then positioned at hydrogen-bonding distance from each other at heavy atom donor−acceptor distances of 3.1 Å. A β-turn (VDPPT) between the two beta strands was added using cyclic-coordinate descent algorithm for loop modeling(*141*), complemented by a neighbor-dependent Ramachandran distribution(*142*). The turn contained a trans peptide bond that connected the two proline residues. An amide capping group was added to the C-terminus. In the next step, amino acid side chains were added to the constructed backbones. Two hairpins were positioned such that their hydrophobic faces were in contact and the total energy was minimized (using NAMD and the CHARMM22 molecular potential, see below) with respect to variation of the rigid body translation of the parallel hairpins with respect to one another. Side chain conformations where determined as the most probable amino acid conformations identified by statistical sequence design algorithm(*44*, *143–145*); Nal side chain conformations (rotamers) were adapted from those associated with phenylalanine. In the last step, the complete octamer structures were created by replicating the hairpin bilayer pairs along the fibril growth direction. Two initial structures of LNK1 were generated: (1) the neighboring LNK1 hairpins were positioned such that neighboring beta strands were at hydrogen bonding distance, and (2) this structure from (1) was energy minimized interbilayer distance while simultaneously solving for the most likely side chain conformations using computational design methods identify the most probable Nal conformations(*44*, *143–145*).

2.4.2 Molecular simulations

Molecular dynamics simulations were performed using the NAMD2 package(*146*) with the modified CHARMM22 force field(*93*) with TIP4P water model(*147*). Energy minimization by conjugate gradient and line search were performed for 100 to 500 steps on all initial peptide structures. All peptide models were solvated using the SOLVATE module in VMD(*148*). A rectangular simulation box was chosen such that the minimum initial separation between any peptide atom and the nearest boundary of the solvent box was 15 Å. The appropriate choice of water model can affect the production of correct conformations of small peptides(*149*), and four-site water models such as TIP4P, which can better populate fully coordinated water configurations than three-site water models, have been suggested for simulating such systems(*150*).

Sodium and chloride counter-ions were added to neutralize the protonated Lys residues, and the ionic strength was set to 0.1 M. All Lys residues are protonated to maintain the equivalence of Lys residues in the fibril. Interestingly, this choice also demonstrates the robustness of the β-fibril assemblies to the electrostatic repulsive interactions among these exterior Lys residues. Each system (MAX1: 30,136 atoms, LNK1: 33,132 and 35,738 atoms) was then subjected to a constant 1000-step energy minimization prior to the simulations.

The NPT ensemble was applied to all systems with constant pressure at 1 atm and constant temperature at 310 K for MAX1 and 320 K for LNK1 to match experimental conditions. Constant pressure was maintained by the Nosé−Hoover Langevin piston method, and constant temperature was controlled by Langevin dynamics. Electrostatic interactions were evaluated fully by particle-mesh Ewald (PME) method at 1.0 Å grid spacing under periodic boundary conditions. Nonbonded interactions were gradually cut off from 10 to 12 Å with the pair-list interactions truncated at 14 Å. The SHAKE algorithm(*151*) was employed to preserve rigid bonds involving hydrogen atoms. An initial 300 ps of solvent relaxation was performed for each model system with protein atoms fixed to their initial coordinates. For the MAX1 octamer, five separate simulations were performed with different starting solvent configurations and random initial velocities; these initial configurations were sampled at 100, 150, 175, 200, and 250 ps from the solvent-relaxation trajectory. For the LNK1 octamer, three simulations were performed for each of two distinct initial protein structures as described above. For each structure, three initial solvent configurations were sampled from the solvent-relaxation trajectories at 100, 150, and 200 ps. Each system was then subject to another 1000-step energy minimization following by a 100 ps preproduction simulation. During these solvent relaxation and preproduction simulations, the bonded and van der Waals interactions were calculated at every 1 fs time step, and the long-term full electrostatics were computed every other step. Subsequently, for the production simulations, a 2 fs time step was used, and configurations were sampled for analysis every 0.01 ns.

2.5 Results and discussion

2.5.1 MAX1 and LNK1 hydrogel assembly

For the same solution conditions (pH 9, 125 mM boric acid, 10 mM NaCl), the CD data (mean residue ellipticity as a function of temperature) (Figure 2.2) reveal similar folding transition temperatures (T_f ~ 30 °C) from random coil to the β-sheet secondary conformation for both MAX1 and LNK1 (full wavelength spectra for MAX1 and LNK1 from which Figure 2.2 was constructed are shown in the Supporting Figure 2.3). The local nanostructures for both MAX1 and LNK1 fibrils are similar, in particular the fibril thickness as observed by TEM (Figure 2.4); both peptides assemble into fibrils with uniform width of approximately 3 nm. This similarity in the MAX1 and LNK1 fibril morphology and the width of the fibril is consistent with each peptide assembling into a fibril whose cross-section involves two stacked β hairpins that contact one another via a hydrophobic core.

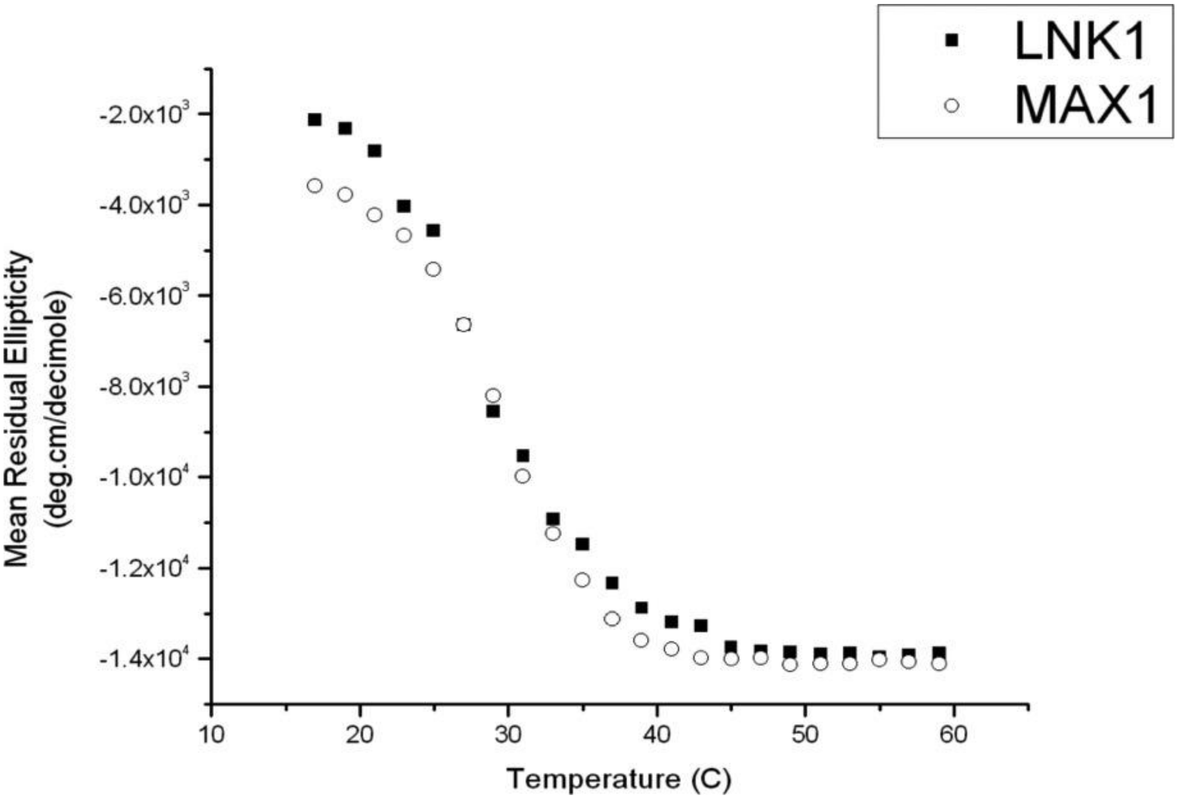
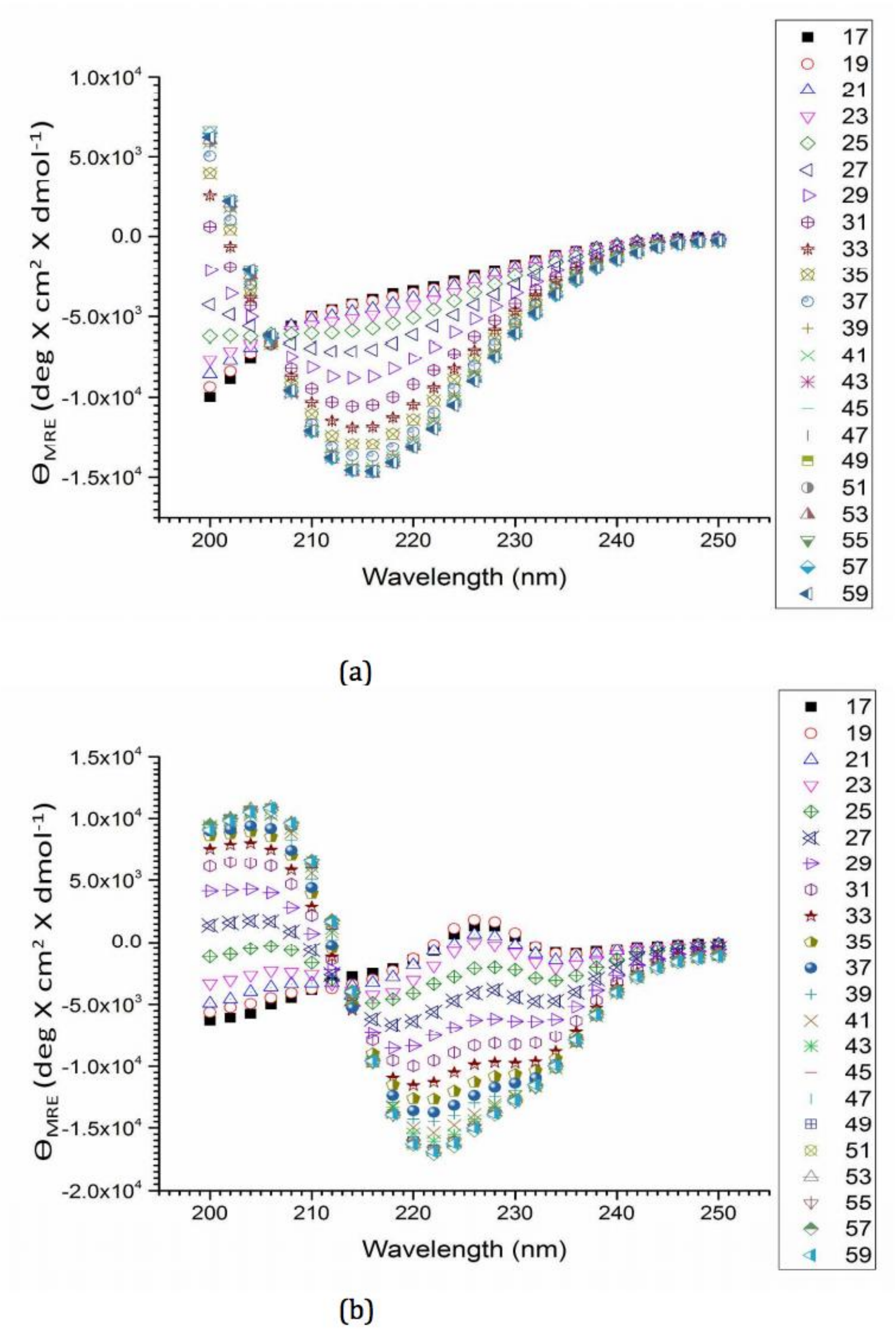


Figure 2.2. Circular dichroism data (mean residue ellipticity in deg·cm/decimole at 218 nm v/s temperature °C) showing approximately the same folding transition temperature (~30 °C) from random coil to β-sheet secondary conformation for both MAX1 and LNK1. Adapted with permission from *Biomacromolecules* (2014).



(a)

(b)

Figure 2.3. Mean residual ellipticity values as a function of incident wavelength (200-250 nm) at different temperatures (°C), indicated in the column to the right for **(a)** MAX1 **(b)** LNK1 both at 150 µM concentration at solution conditions pH 9 (125 mM boric acid, 10 mM NaCl). Adapted with permission from *Biomacromolecules* (2014)

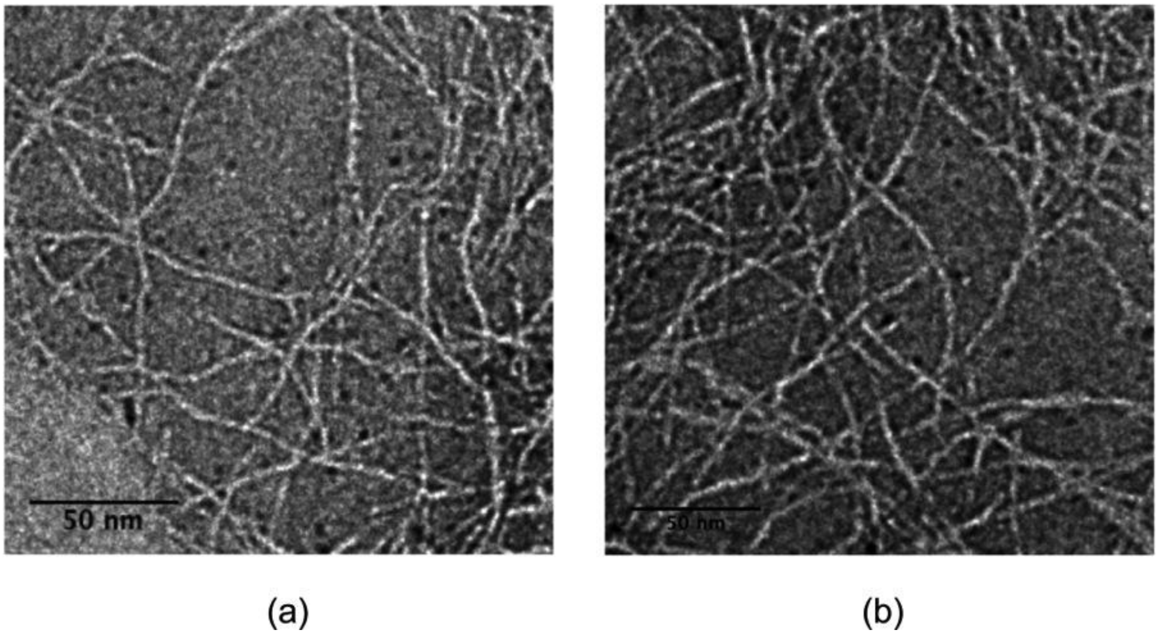


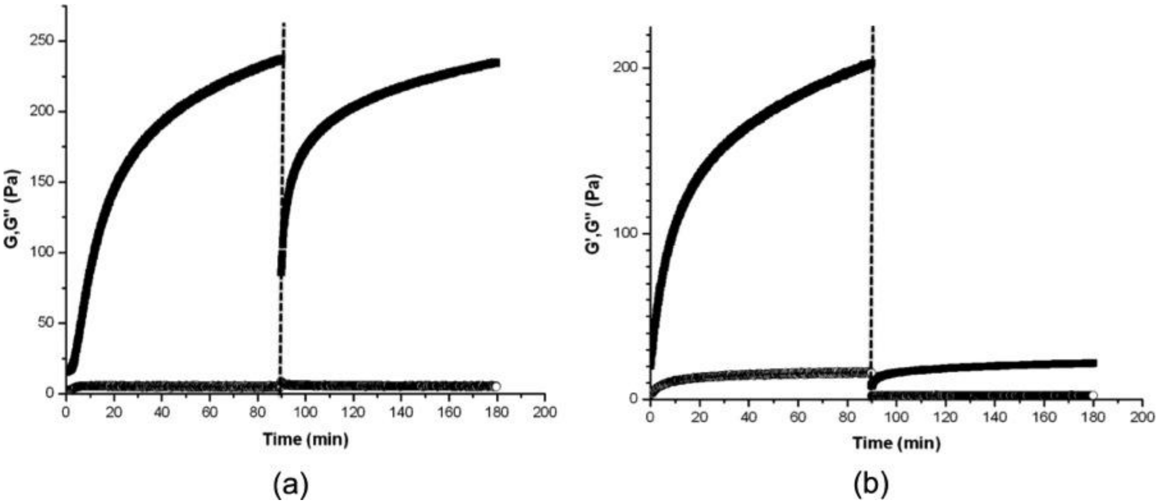
Figure 2.4. Transmission electron micrographs of **(a)** MAX1 and **(b)** LNK1. Samples were prepared at pH 7, 50 mM bis-tris propane, 150 mM NaCl buffer, and stained with 1% (w/v) uranyl acetate in deionized water. Adapted with permission from *Biomacromolecules* (2014).

2.5.2 Rheological behavior of hydrogel assembly

Solid MAX1 hydrogels exhibit a unique property of undergoing shear thinning and flow under an applied shear stress (outside of the material linear viscoelastic regime) but immediately recovering into solid gels on cessation of shear. An earlier study by Yan et al.(*152*) exploring the hydrogel behavior during and after flow indicated the fracture of the gel networks into domains much larger than the length scale of individual fibrils in order to flow. The network morphology within the gel domains during flow was structurally identical to the parent network at rest; the peptide fibrils displayed the same cross-section, the same physical cross-linking points of fibrillar entanglement and branching, and the same porosity. On cessation of shear, the large gel domains immediately percolate and form a bulk, hydrogel network. This shear-thinning and rehealing behavior of MAX1 would not exist if the network disassembled into individual fibrils during flow since there would be no immediate mechanism for the fibrils to recross-link and percolate into a bulk network. Our hypothesis has been that it is the frequent and fast fibrillar branching during MAX1 assembly that is key to this shear-thinning but immediate network reformation behavior. If the network were composed of fibrils with only physical entanglements for cross-links, the shear flow would disentangle the fibrils, thus completely disrupting the network after shear. However, the branching causes the network to fracture into large domains of intact network structure in response to shear that does not allow the simple disentanglement of peptide fibril physical cross-links during shear flow.

If the fibril branching in MAX1 is responsible for the observed shear-thinning and immediate gel reformation behavior, then ridding the system of most fibril branching should significantly affect the hydrogel flow properties. For MAX1, the putative fibril interior interface between peptides contains solely valine residues. Such a “flat,” featureless, hydrophobic interface is expected to tolerate fluctuations in the relative orientations of peptides within the fibrillar assembly and potentially lead to fibril branching. As mentioned earlier, the design of steric specificity in the hydrophobic core of the LNK1 fibrils was an attempt to produce lock and key type interactions and preclude fibrillar branching. Thus, a very different shear response is expected when LNK1 hydrogel networks, presumably held together with primarily physical entanglements as cross-links, are subject to the exact same shear treatment as the MAX1 networks.

To explore the rheological response of MAX1 and LNK1 networks to shear and flow, self-assembled hydrogels from LNK1 or MAX1 were produced at a concentration of 0.5% (w/ v) at pH 9 (125 mM boric acid, 10 mM NaCl). Under these conditions, both hydrogels show similar preshear behavior with $G' \sim 250$ Pa $\gg G'' \sim 20$ Pa for the MAX1 hydrogel and $G' \sim 200$ Pa and $G'' \sim 20$ Pa for the LNK1 hydrogel, where G' and G'' are the storage and loss moduli, respectively. Figure 2.5a shows the shear-thinning and recovery character of MAX1 hydrogels in which a MAX1 gel was subjected to steady-state shear rate of 1000/s for 120 s. Upon cessation of shear, the hydrogel immediately showed solid gel properties ($G' \sim 75$ Pa $\gg G''$) and quickly recovered to almost the same value of storage modulus of the preshear, original MAX1 network ($G' \sim 250$ Pa) after several hours. In stark contrast, when an LNK1 network, formed with the same solution conditions as the MAX1 hydrogel network, was subject to the identical shear treatment, it immediately displayed very weak hydrogel network properties ($G' \sim 5$ Pa $> G''$) and failed to recover to even 10% of its original modulus value after several hours (Figure 2.5b). The LNK1 design was meant to prevent branching of the peptide fibrils during assembly. The response to shear was consistent with this absence of branching, and the shear treatment destroyed most physical entanglements between LNK1 fibrils that were unable to reform in any significant way on cessation of shear. This lack of rehealing upon cessation of shear is a clearly different shear response by the LNK1 hydrogel network, relative to MAX1, that will be discussed later in this section.



(a)

(b)

Figure 2.5 Oscillatory time sweep measurements before and after application of steady-state shear (1000/s for 120 s, indicated by dotted line) on 0.5% (w/v) **(a)** MAX1 and **(b)** LNK1 networks under the same solution conditions (pH 9 125 mM boric acid 10 mM NaCl). Solid squares indicate *G′* (Pa) and open circles *G″* (Pa). Adapted with permission from *Biomacromolecules* (2014).

After a simple steady shear treatment of LNK1 inside the rheometer (Figure 2.5b), the oscillatory shear data indicate a significant reduction in the hydrogel storage modulus *G′,* consistent with a strong reduction of network-like properties of LNK1. In order to more closely mimic conditions of potential clinical usage, both LNK1 and MAX1 hydrogels were subject to the syringe injection shear treatment as described in the Materials and Methods section. The oscillatory frequency sweep data in Figure 2.6b reveal a complete elimination of hydrogel network properties of LNK1 networks post injection shear treatment. The LNK1 sample shows a greater value of the loss modulus, *G″,* as compared to the storage modulus, *G′,* with *G″* > *G′* at all frequencies. This is a clear signature of a material that is not a percolated hydrogel network but, rather, is a particulate suspension or molecular solution. In stark contrast, MAX1 materials retain hydrogel network properties even after the injection shear treatment with G′ ≫ G″ at all frequencies. The transmission electron micrograph in Figure 2.6d shows the morphology of the LNK1 networks at the end of the syringe injection shear treatment. Observed fibrillar bundle-like nanostructures are much wider (~15−20 nm) and more nonuniform than LNK1 fibrils observed prior to injection (Figure 2.6b). The existence of these bundled structures can be attributed to fibrillar stacking within LNK1 samples as a result of the injection shear treatment that caused disentanglement of the original percolated LNK1 network. Once in these stacks, the fibrils no longer contribute to network properties that confer large storage modulus and no longer form physical cross-links through entanglements. In contrast, the MAX1 local nanostructure at the end of the exact same shear injection treatment (Figure 2.6c) is similar to the MAX1 fibrils seen pretreatment in Figure 2.4. The oscillatory frequency sweep measurements carried out on the MAX1 networks clearly indicate a gel-like response from MAX1 networks postshear injection treatment. MAX1, with a hydrophobic face composed entirely of valine side chains that have the same side chain volume, demonstrates fibrillar branching and, thus, a bulk hydrogel network of MAX1 subject to shear treatment reheals into a fully percolated network when shear is stopped. Even the nanostructure of the MAX1 network before and after shear treatments is the same. In the case of the LNK1 peptides with a designed specificity in the hydrophobic face, fibrillar branching is severely limited, and the rehealing properties of bulk LNK1 hydrogels subject to shear are eliminated. Indicative of this rheological behavior is the significant morphology change in LNK1 from individual fibrils in the initial hydrogel network to a multi-fibrillar bundled structure after flow, presumably due to LNK fibrils collapsing together, thus eliminating network properties. Considering the experimental data together, designed hydrophobic specificity clearly affects the self-assembled hydrogel properties.

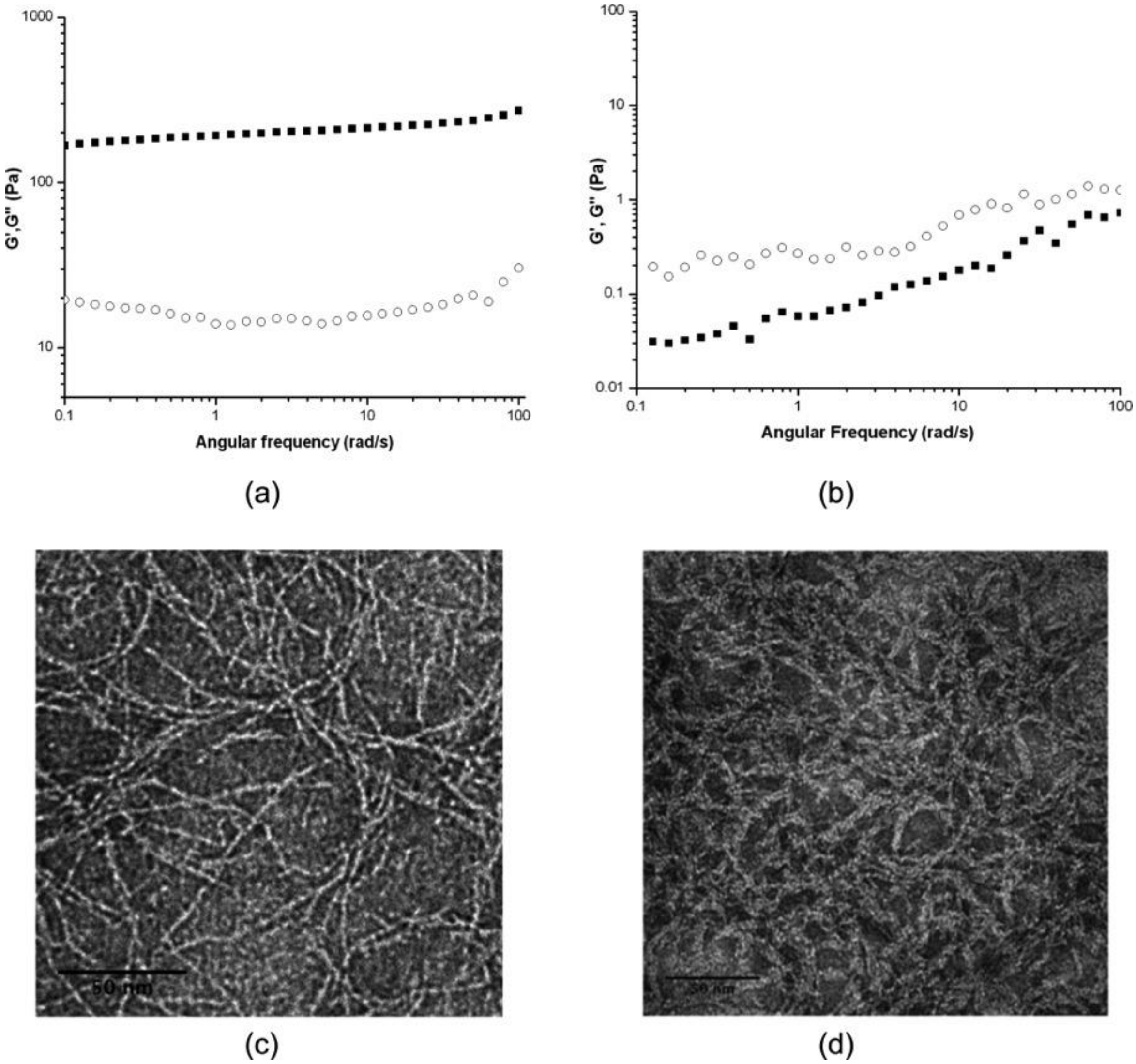


Figure 2.6 Oscillatory frequency sweep **(a)** MAX1 and **(b)** LNK1 measurements after application of injection shear treatment to both networks formed under the same solution conditions (pH 9 125 mM boric acid, 10 mM NaCl). Solid squares indicate *G′* (Pa) and open circles *G″* (Pa). Transmission electron micrographs of **(c)** MAX1 and **(d)** LNK1 post injection shear treatment. Adapted with permission from *Biomacromolecules* (2014).

2.5.3 Simulation trajectories of MAX1 and LNK1: Local molecular structures

The experimental results suggest that LNK1 molecules, due to the hydrophobic specificity in the resultant fibril cores, significantly limit fibril branching and thus greatly affect bulk hydrogel response to shear. Here, MD simulation is used to provide molecular insights into why MAX1 and LNK1 β-fibrils result in different nanostructures (i.e., lack of branching in LNK1 vs MAX1) and consequent hydrogel properties.

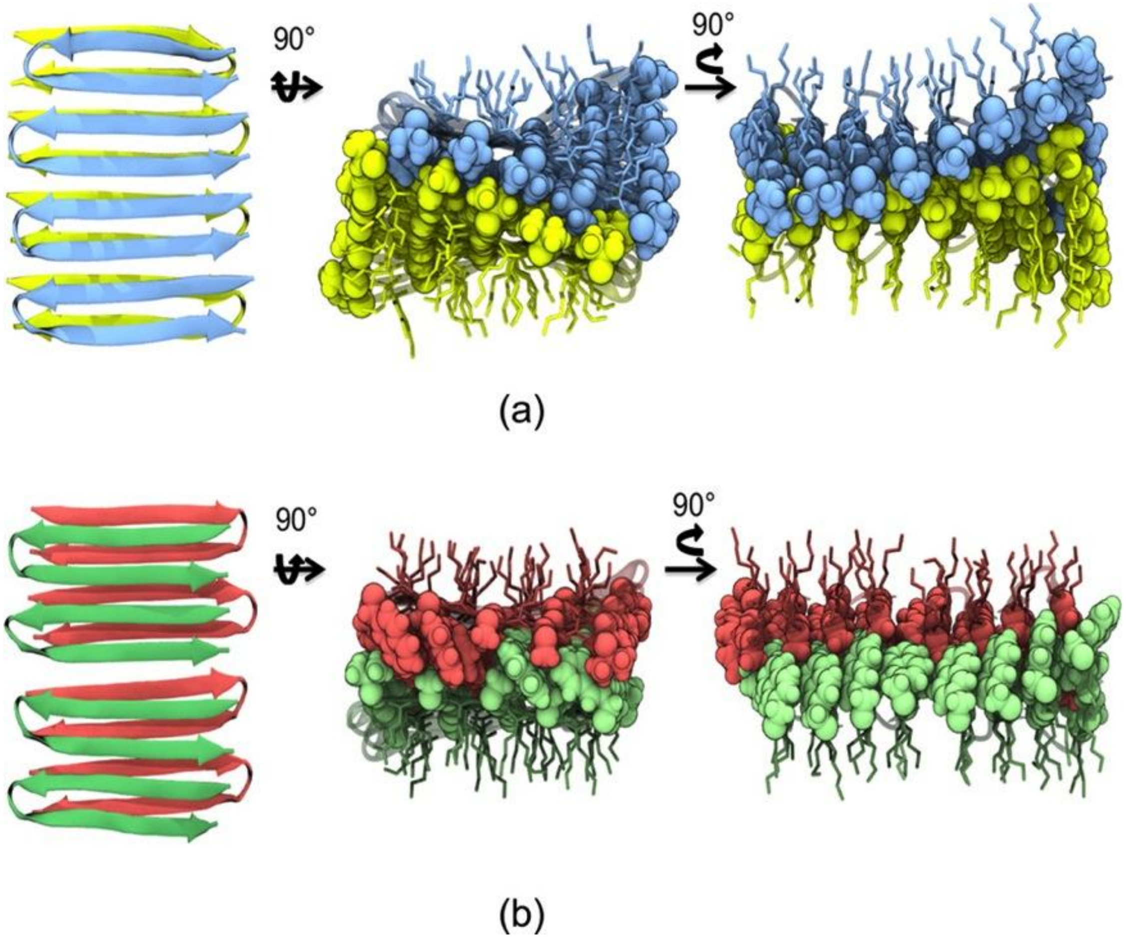


Figure 2.7 Renderings of representative equilibrium structures of **(a)** MAX1 octamer after 70 ns and **(b)** LNK1 octamer after 70 ns. Orthogonal views are shown for each. Coloring is used to distinguish layers in bilayer structure of each fibrillar assembly. The molecular structure of LNK1 presents a specific “lock and key” packing style between size-complementary Nal and Ala residues. “Aromatic ladder” of Nal residues is shown in the panel on the right in (b) (Thr and Val residues are omitted from this figure for clarity). Adapted with permission from *Biomacromolecules* (2014).

Simulations of MAX1 and LNK1 octamers in explicit solvent were performed for 70 ns. The initial structure consisted of ideal, coplanar β sheet structures. The initial configuration contained antiparallel beta strands throughout each layer and had a 2-fold rotational symmetry along the fibril axis. This was termed the "anti" configuration. In preliminary simulations of MAX1 and LNK1, this anti configuration was found to be more persistent and robust than the "syn" fibril configuration, wherein the beta hairpin turns are in proximity and are on the same side of the peptide bilayer. A syn relationship is adopted between turns within distinct monolayers of the bilayer (along the fibril axis). This is based on recent 2D ^{13}C–^{13}C fpRFDR and PITHIRDS-CT NMR data that supported this nanostructure and showed evidence for a syn relationship in MAX8 with possible coexistence of an anti-orientations of the hairpins(*153*). Later, the exact nature of the turn placement on MAX1 was discovered by solid-state NMR studies and in agreement with our model(*154*). Over the course of repeated simulations, MAX1 was observed to adopt a variety of fibrillar configurations achieved through interlayer sliding and twisting of the β strands (Figure 2.8a). On the other hand, simulations of LNK1 exhibited conformational rigidity and little sliding of β strands, even for different initial structures. Throughout the LNK1 simulations, the complementary packing of napthyl side-chains was retained (Figure 2.7b).

The site fluctuations of the interior hydrophobic amino acids were considered for each peptide. In LNK1, the shape- complementary design involving large Nal side chains opposite small Ala residues assists to stabilize and rigidify the self- assembled structure, largely restricting LNK1 fibril to linear growth. In comparison, MAX1 has a relatively flat and featureless hydrophobic interior interface comprising only Val. Side-chain mobility was examined at these interior interfaces. To characterize such fluctuations, the side chain dihedral angle $\chi 1$ was defined using four atoms (Nal: N-CA-CB- CG, Val: N-CA-CB-CG1). This angle $\chi 1$ is the interior torsional angle for rotation about the bond connecting Cα and Cβ atoms of the amino acid side chain. Such internal angles are expected to populate energy basins centered near of 60°, 180°, and 300°. For the LNK1 simulations, Nal residues were observed to populate values of $\chi 1$ at either 180° or 300°, and none of the residues were observed to sample both of these basins; each residue was essentially locked in a particular value of $\chi 1$ and impeded from rotating due to the tight packing of the bilayer's hydrophobic interior. In contrast, Val residues in MAX1 were observed to explore $\chi 1$ values near 60°, 180°, and 300°, with $\chi 1$ = 180° being the most populated (85%). Individual side chains were observed to transition between these values of the angles. Thus, in MAX1, the bilayer interface is not rigid, and Val side chains can rotate to potentially accommodate lateral sliding of the two layers.

The oligomer simulations can be used to estimate the larger scale structure of the fibril, such as the twist of the fibril. The longitudinal twist is characterized by a twisting angle α, defined as the effective angle between adjacent β-strands on the same β sheet (same layer). A vector associated with each β-strand was defined as $\boldsymbol{v}_i = r_{08} - r_{02}$ or $\mathbf{v}_i = r_{13} - r_{19}$, where r_n is the position of the Cα atom of residue n (the same convention as used in the cross-angle calculation described below). The twisting angle α was calculated using $\cos\alpha = (\boldsymbol{v}_i \times v_{i+1}) / |\boldsymbol{v}_i| |\boldsymbol{v}_{i+1}|$ and 10 interior pairs of adjacent β-strands of each sampled configuration; the coordinates of the four edge β-strands were not included since these are not representative of the interior structure of the fibril. Larger α denotes a structure with more longitudinal twist and smaller overall pitch. MAX1 exhibited $\alpha = 6.75 \pm 0.77°$, whereas LNK1 was flatter and less twisted with a value of $\alpha = 4.54 \pm 1.39°$ (uncertainties are one standard deviation). The predicted corresponding pitch in each case is 53 β-strands (25 nm) for MAX1 and 80 β-strands (40 nm) for LNK1. Previous simulations of amyloid peptides have also observed that (a) longer fibrils exhibit a higher propensity for introduction of defects with increasing twist and (b) fibrils with complementary packing of residues along an interior interface are less twisted(*155*). Larger twist values require the sliding of the layers at the interior hydrophobic interface. The complementary interior packing of LNK1 prevents this, whereas MAX1 with its flatter, more mobile interior interface can accommodate such sliding.

To quantify the structure and fluctuations of the peptide bilayer, a crossing angle θ was defined, which is the effective angle between β-strands on opposing β-sheets within the fibril. A vector v_1 describing the orientation of a β-strand is represented by the difference in the coordinates ($r_{08} - r_{02}$, $r_{19} - r_{13}$) of the Cα atoms on residues Lys08 and Lys02 / Lys13 and Lys19, $\boldsymbol{v_1} = r_{08} - r_{02}$ and $\boldsymbol{v_2} = r_{19} - r_{13}$. For a pair of β-strands opposite each other within the fibril oligomer, the crossing angle can be calculated using $\cos\theta = (\boldsymbol{v_1} \cdot \boldsymbol{v_2}) \ / \ |\boldsymbol{v_1}| \ |\boldsymbol{v_2}|$. If two strands are parallel, then $\theta = 180°$. θ was then averaged across all eight pairs in the structure for each sampled configuration. MAX1 exhibited a range of crossing angles, and representative configurations are rendered in Figure 2.8a. As can be seen in Figure 2.8a, the closer the crossing angle is to $\theta = 180°$, the more aligned the β-strands are on the opposite sheets. The distribution of this crossing angle was calculated using configurations sampled after the first 20 ns of the simulations. Despite initial conditions near *θ* = 180° for each peptide, the simulations sample a wide range of crossing angles for MAX1, *θ* = 159° − 178°. Analogous simulations of LNK1 have a narrow range of *θ* and a single peak centered at θ = 175° corresponding to a well-aligned cross-beta conformation that could extend linearly during fibril growth. On the other hand, MAX1 explores a much broader range of *θ* and has two maxima (Figure 2.8c) at *θ* = 176° and *θ* = 167°, respectively. The large fluctuations and the presence of the second peak at *θ* = 167° are consistent with a significant population of MAX1 conformations that could potentially form branch points where the two leaflets of the fibril bilayer separate thus exposing valine side chains. For such values of *θ*, the resulting *β*-strands at the ends of the oligomer, their partially exposed valine residues and available backbone hydrogen bonding sites, are poised to recruit additional peptides and form two, new daughter fibrils at such a putative branch point. For values less than *θ* = 180°, there is potential exposure of valines and hydrophobic surface as this angle decreases. Analysis of the exposed solvent accessible hydrophobic surface area (VMD), however, reveals only a weak negative correlation of this surface area with crossing angle for the small oligomers considered in the simulations (Pearson's correlation coefficient of −0.3). This is consistent with the notion that smaller crossing angles have greater exposed solvent accessible hydrophobic surface area. Simulations of larger assemblies are likely necessary, however, to better resolve the roles of exposed hydrophobic groups in fibril extension and defect formation.

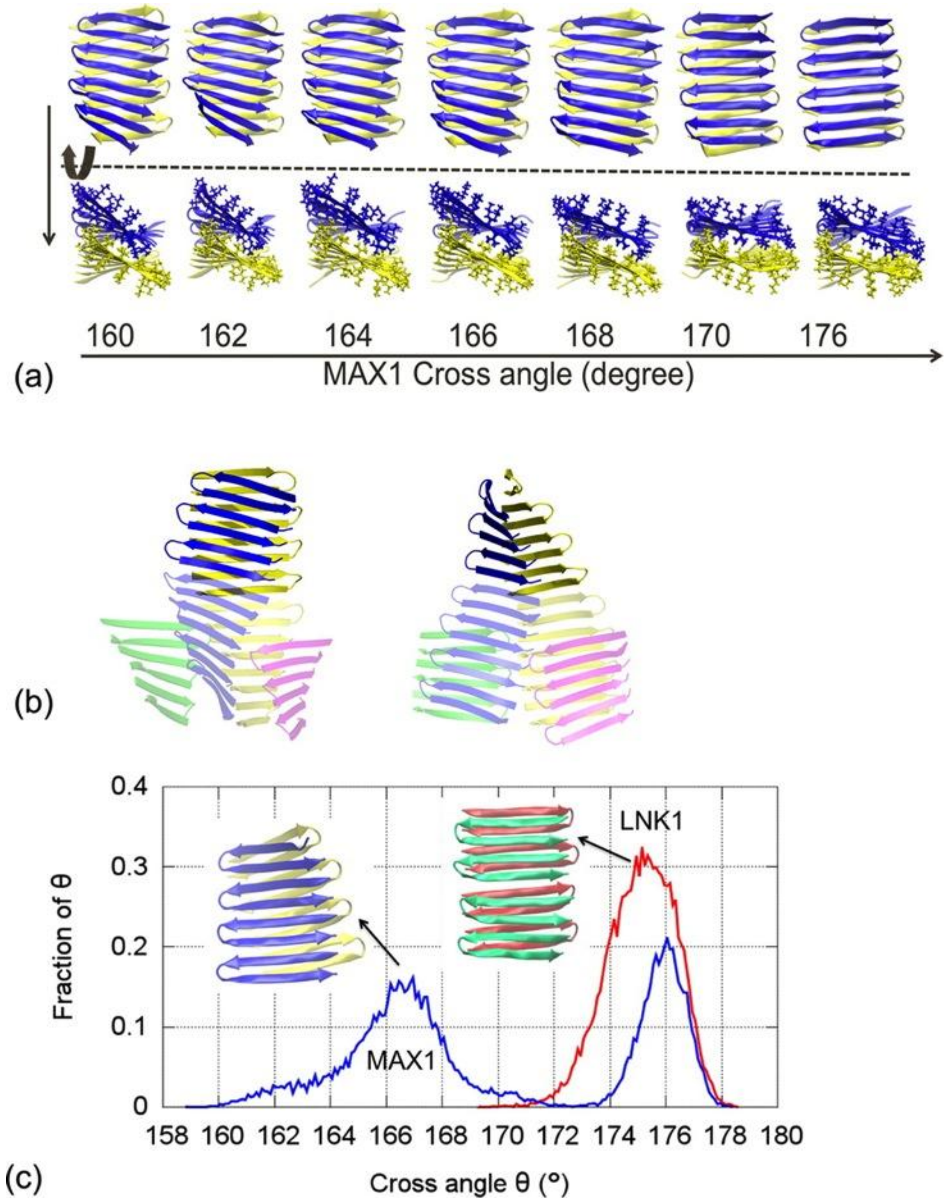


Figure 2.8 (a) Representative conformations of the MAX1 model at different crossing angles. The top series of structures show the extent of the crossing angle *θ*. The bottom series presents qualitative visualization of the twist and gradual exposure of hydrophobic surface as the crossing angle decreases. **(b)** Modeled MAX1 branch point from two perspectives. Transparent regions (not sampled from the simulation) are modeled based upon the MAX1 configuration at *θ* =160° and positioned along the fibril growth direction. **(c)** Distributions of crossing angle *θ* for MAX1 (blue) and LNK1 (red). Sampled configurations are obtained from five simulations of MAX1 and six simulations of LNK1. Representative structures of MAX1 (*θ* =167°) and LNK1 (*θ* =175°) from the simulations are rendered. Sampled configurations are collected every 10 ps after the first 20 ns for each MAX1 simulation and after the first 15 ns for the LNK1 simulations. Adapted with permission from *Biomacromolecules* (2014).

Although the short oligomers considered in the simulations are certainly not fibrils, they are likely suggestive of the fibril's interior molecular structure. The octamers used in the MD simulations, composed of 8 *β*-hairpins and 16 *β*-strands, provide a representation of the *β*-fibril likely present in the hydrogel. The 16 *β*-strands form a "cross-*β*"-like structure. Other recent MD simulations also show examples of the formation of the "cross-*β*" structure in small collections of peptide molecules(*156, 157*). For example, MD simulations performed by the activation-relaxation technique coupled with a coarse-grained energy model showed the formation of the "cross-*β*" structure with six strands of one of the shortest amyloid-forming peptides KFFE(*158*). Another example is the tetramer of the peptide Beta2m (NHVTLSQ) that has been shown to visit conformations of a double-layered *β*-sheet(*159*).

The twisted "cross-*β*" structure observed in the simulation is not unique to MAX1 and LNK1. In fact, a number of other amyloid-like fibrils show a twisted and bent *β*-sheet structure in simulations, such as the layered peptide self-assembly mimic (PSMA)(*160*) and the Alzheimer's A*β* fibril(*161*). Although fibril bending was not observed in the simulations of MAX1/LNK1 octamer units due to the limited fibril length of these model oligomers, the bending feature is confirmed by the TEM images of both fibril networks. In contrast, due to the approximately cylindrical cross-section of the *β*-hairpin fibrils, any twisting of the fibrils is not resolvable experimentally in TEM. If the *β*-hairpin fibril cross-section is more ribbon-like due to molecular design and has a larger aspect ratio, then the twists of the *β*-sheet fibril can be observed microscopically(*162*). In the simulations, MAX1 is more twisted than LNK1. This is because Nal residues align in the hydrophobic interface of LNK1 to stack and form an “aromatic ladder” pattern and the complementary steric interactions at the bilayer interface prevent sliding of the layers. Similar aromatic ladders can flatten and rigidify the “cross β” structure and stabilize the structure via increased numbers of hydrogen bonds and side-chain contacts between sheets(*160*). The LNK1 fibril would be expected to be less flexible and more resistant to structural defects, such as branching, during self-assembly.

2.6 Conclusion

The LNK1 peptide design employs a steric lock and key specificity in the hydrophobic core of the β-sheet fibrils formed by the peptide. Experimentally, stark differences in the network properties of the hydrogels formed by the LNK1 peptide were observed when compared to hydrogels formed from MAX1 peptide. A prominent difference was the lack of recovery of storage modulus G' (Pa) values from the LNK1 network after the brief application and cessation of steady-state shear. In addition to this, sequential syringe injections applied to the networks as a means of shear treatment caused complete elimination of network properties of the LNK1 networks whereas MAX1 samples retained hydrogel properties. These experimental differences are attributed to the lack of branching in LNK due to the lock and key hydrophobic packing in the constituent fibril cores. The experimental observations were supported by molecular dynamics simulations carried out on modeled structures of MAX1 and LNK1. The simulations reveal that in the bilayer fibril complementary hydrophobic interactions yield a “lock and key” hydrophobic core packing. Moreover, the LNK1 hairpins form an aromatic ladder-like pattern, which renders it less flexible. The LNK1 fibril is expected to be resistant to structural defects like branching during self-assembly, which is consistent with the experimentally observed unbranched homogeneous fibrillar structure of LNK1. MAX1 differs from the LNK1 primarily in terms of having a sterically nonspecific hydrophobic core amenable to fluctuations in spacing and packing between fibril hairpins, suggestive of a system susceptible to branching. Thus, molecular design, here applied to the creation of the LNK1 peptide, can be used to control and modulate the assembly, nanoscale structure, and rheological properties of a variety of peptide based hydrogel networks.

CHAPTER 3 Computationally Designed Peptides for Self-Assembly of Nanostructured Lattices

Adapted from the manuscript under review with *Science Advances.* Science (2016)

3.1 Abstract

Folded peptides present complex exterior surfaces specified by their amino acid sequences, and the control of such surfaces offers high-precision routes to self-assembling materials. The complexity of peptide structure and the subtlety of noncovalent interactions make the design of predetermined nanostructures difficult. Computational methods can facilitate such design and are used here to determine 29-residue peptides that form tetrahelical bundles that, in turn, serve as building blocks for lattice-forming materials. Four distinct assemblies were engineered. Peptide bundle exterior amino acids were designed in the context of three different interbundle lattices in addition to one design to produce bundles isolated in solution. Solution assembly produced three different types of lattice-forming materials that exhibited varying degrees of agreement with the chosen lattices used in the design of each sequence. Transmission electron microscopy revealed the nanostructure of the sheet-like nanomaterials. In contrast, the peptide sequence designed to form isolated, soluble, tetrameric bundles remained dispersed and did not form any higher-order assembled nanostructure. Small-angle neutron scattering confirmed the formation of soluble bundles with the designed size. In the lattice-forming nanostructures, the solution assembly process is robust with respect to variation of solution conditions (pH and temperature) and covalent modification of the computationally-designed peptides. Solution conditions can be used

to control micron-scale morphology of the assemblies. The findings illustrate that with careful control of molecular structure and solution conditions, a single peptide motif can be versatile enough to yield a wide range of self-assembled lattice morphologies across many length scales (1 nm – 10 μm).

3.2 Introduction

Self-assembly of designed molecules in solution provides striking potential for efficiently achieving complex, robust materials with nanometer precision. Traditional nanomaterial assembly strategies have employed small molecules(*163*, *164*) or polymeric(*165–167*) amphiphiles. Recently developed assembly methods can produce complexity in structure and composition through chemical variation of the assembling molecules(*168–171*) or the use of hierarchical solution assembly protocols(*20*, *172–174*). Biopolymers offer unique capabilities to encode both local molecular building block structure and long range material morphology via the design of specific sequences; such design has been applied to DNA(*23*, *175*), polypeptides(*68*, *77*, *176–180*), and polymer-biomolecule hybrids(*23*, *181*). Solution assembly of peptides can readily produce "one-dimensional" nanostructures such as fibrils(*182–184*) and tubes(*185–187*). New peptide nanostructure formation strategies have employed non-natural peptide sequences(*68*) as well as biomimetic strategies using modified natural proteins(*35*, *46*). Much of the peptide work involves the synthesis of new systems and subsequent characterization of the structures they form. The a priori design of proteins and peptides that form targeted assemblies is subtle, however, due to the complexities and subtleties of folding and protein-protein interactions. Moreover, such assemblies can be highly sensitive to sequence and mutation. These difficulties have motivated the experimental use of more easily programmed interactions at protein interfaces, such as metal coordination(*48*), to drive intermolecular assembly.

Theoretical and computational methods provide a way to approach the design of intermolecular, noncovalent interactions between self-assembling peptides or proteins in solution to produce materials with predetermined morphologies, including desired point and space group symmetries(*37*, *39*, *40*, *188*). Nearly all these efforts in assembly design have employed variants of natural proteins as building blocks, and different tertiary and quaternary structures are often employed for different local geometries in the assembly. Herein, we present the computational de novo design of peptides that are robust, easily synthesized, and versatile.

Our aim is to explore the extent to which the information required for folding and intermolecular long-range order can be designed de novo into short peptide sequences, as opposed to the redesign of large natural proteins. The effort is focused on (a) de novo design of a homotetrameric helix bundle motif that is robust with respect to variation of exterior residues, (b) design of the exterior residues to guide the solution assembly of variants of this motif into distinct lattices having rectangular, square, or hexagonal local symmetries, (c) experimental characterization and determination of the extent to which the nanostructures are robust with respect to solution phase conditions and, (d) exploration of how solution conditions can be used to control micron-scale morphology.

3.3 Results

3.3.1 Computational design

The designed building block motif consisted of a helical homotetramer of four 29-residue peptides arranged with D_2 symmetry (Figure 3.1). While many similar oligomeric helical proteins have been designed and investigated, we seek a structure that is robust with respect to variation of exterior residues, and we opt to design the structure and hydrophobic core de novo. Candidate bundles were generated via a multiparameter mathematical model of helical coiled coils(*78*) with the final bundle structure specified by a set of five defining geometric parameters (see details below). For each candidate bundle motif, a probabilistic approach was applied to calculate the sites pecific probabilities of the amino acids at variable residues(*44*). The calculations also yield an average energy, E, over sequence probabilities for a given bundle structure(*44*). Using E as an objective function in a Monte Carlo search over helix bundle parameters was performed. A helical peptide structure and 11 interior hydrophobic residues were specified (highlighted in gray in Table 3.1), providing the tetrameric helix bundle motif, or building block, for subsequent design of the material assemblies. The remaining 18 residues were designed in the context of four predetermined material nanostructures: an isolated, water-soluble helix bundle not expected to self-associate (Figure 3.1A) and three material assemblies derived from P622, P422, and P222 space group symmetries (Figure 3.1B-3.1D). These layered space group symmetries each contain D_2-symmetric positions on which the individual peptide bundles were positioned (Figure 3.1B-3.1D). Calculations were performed using only a single, isolated layer from the corresponding space group. For a given nanostructure symmetry, the variation of the unit cell parameters produced a set of possible lattice structures consistent with the chosen symmetry; the amino acid probabilities and E were calculated for each such assembly structure. From the resulting energy landscape for each type of assembly, energy minima were identified. Within these minima, sequences were identified, where the amino acid with the largest calculated probability was selected at each variable residue position. More detailed computational design approach is described below.

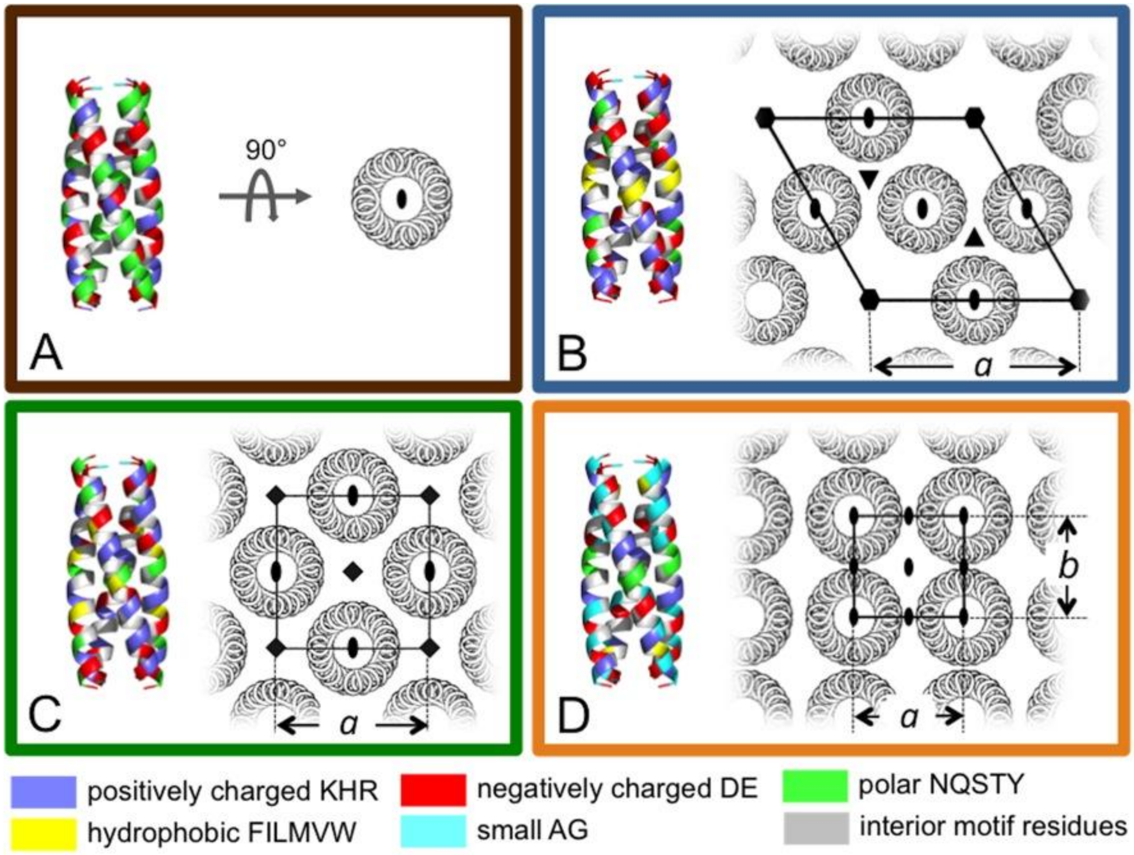
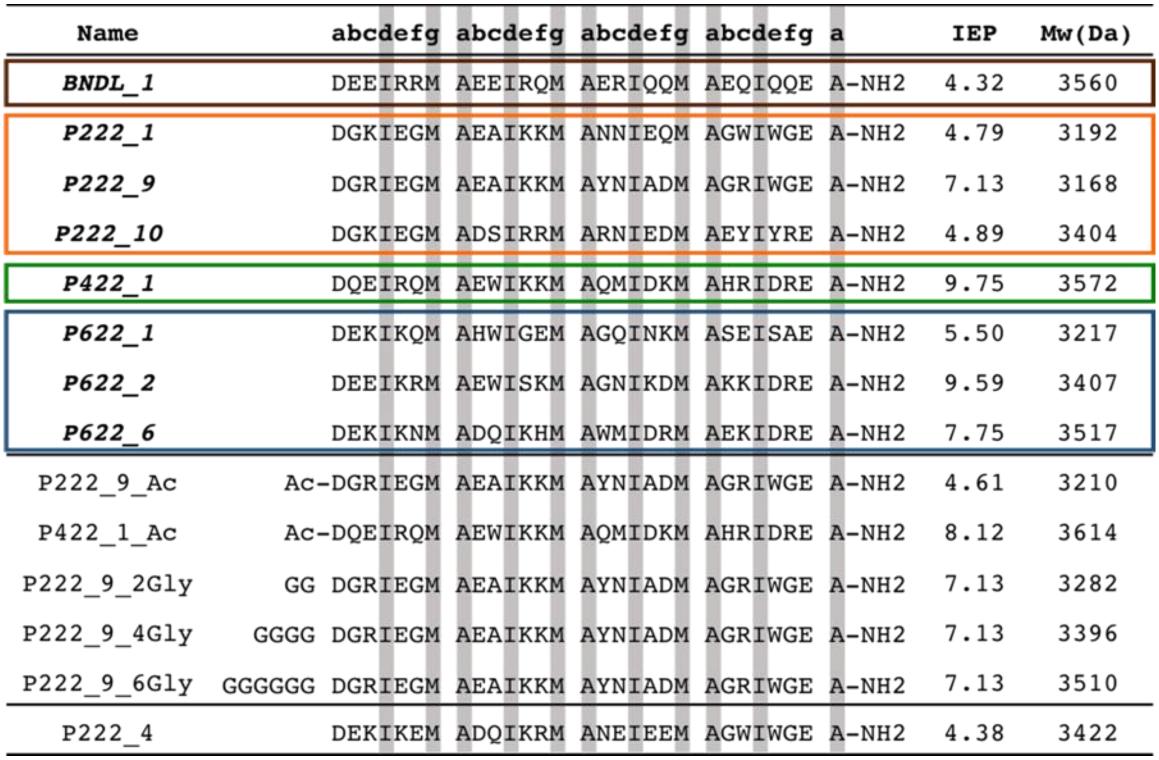


Figure 3.1: Computationally designed, helical, homotetramer assemblies. **A-D**: Models of peptides forming distinct nanostructures using a de novo designed helical homotetramer motif, which comprises both the backbone coordinates of the D_2 symmetric tetramer and interior hydrophobic residues. In the left of each panel, designed exterior residues are colored according to chemical properties: positively charged KHR (blue), negatively charged DE (red), polar NQSTY (green), hydrophobic FILMVW

(yellow), and small AG (cyan). Interior hydrophobic residues common to all the sequences are gray. On the right of each panel, the targeted assemblies are rendered along with symmetry axes (C_2 oval, C_3 triangle, C_4 square, C_6 hexagon) and the unique dimensions of the unit cell, *a* and *b*. **(A)** D_2 symmetric tetramer designed in isolation and targeted to remain not assembled in solution. The exterior residues of the remaining proteins were designed in the context of a single layer from the corresponding space groups: **(B)** P622; **(C)** P422; **(D)** P222.



Name		abcdefg	abcdefg	abcdefg	abcdefg	a	IEP	Mw(Da)
BNDL_1		DEEIRRM	AEEIRQM	AERIQQM	AEQIQQE	A-NH2	4.32	3560
P222_1		DGKIEGM	AEAIKKM	ANNIEQM	AGWIWGE	A-NH2	4.79	3192
P222_9		DGRIEGM	AEAIKKM	AYNIADM	AGRIWGE	A-NH2	7.13	3168
P222_10		DGKIEGM	ADSIRRM	ARNIEDM	AEYIYRE	A-NH2	4.89	3404
P422_1		DQEIRQM	AEWIKKM	AQMIDKM	AHRIDRE	A-NH2	9.75	3572
P622_1		DEKIKQM	AHWIGEM	AGQINKM	ASEISAE	A-NH2	5.50	3217
P622_2		DEEIKRM	AEWISKM	AGNIKDM	AKKIDRE	A-NH2	9.59	3407
P622_6		DEKIKNM	ADQIKHM	AWMIDRM	AEKIDRE	A-NH2	7.75	3517
P222_9_Ac		Ac-DGRIEGM	AEAIKKM	AYNIADM	AGRIWGE	A-NH2	4.61	3210
P422_1_Ac		Ac-DQEIRQM	AEWIKKM	AQMIDKM	AHRIDRE	A-NH2	8.12	3614
P222_9_2Gly	GG	DGRIEGM	AEAIKKM	AYNIADM	AGRIWGE	A-NH2	7.13	3282
P222_9_4Gly	GGGG	DGRIEGM	AEAIKKM	AYNIADM	AGRIWGE	A-NH2	7.13	3396
P222_9_6Gly	GGGGGG	DGRIEGM	AEAIKKM	AYNIADM	AGRIWGE	A-NH2	7.13	3510
P222_4		DEKIKEM	ADQIKRM	ANEIEEM	AGWIWGE	A-NH2	4.38	3422

Table 3.1: Table of computationally determined peptides for solution assembly. Colored rectangles contain eight candidate sequences that were experimentally characterized. Sequences were theoretically designed to produce tetrahelical bundles. BNDL_1 was designed in the absence of any lattice assembly and is expected to remain soluble (brown). The P222 (orange), P422 (green), and P622 (blue) sequences were designed in the presence of lattices of corresponding symmetry. The remaining P222_9 and P422_1 sequences contain covalently modified termini. P222_4 is the only sequence candidate that did not behave as predicted and could not be assembled into the desired nanostructure in the solution conditions used for the other peptides. The heptad repeat positions (abcdefg) of all peptides are shown in the table heading. The designed, hydrophobic interior residues of the motif shown are highlighted in grey.

3.3.1.1 Design of homotetrameric helix bundle motif

An antiparallel tetrameric coiled coil with D_2 point group symmetry was selected as the building block for the designed nanomaterials. Structures and sequences of the tetramer were *de novo* designed computationally to identify a single tetrahelical motif that would be robust with respect to variation of exterior residues, residues which could be subsequently designed for specific assemblies.

The construction of the homotetrameric helical (coiled coil) structures used a mathematical model that describes such structures with a small number of geometric parameters(*78*). Modifications were made to include rotation of peptide planes (comprising backbone N, C_α and C atoms)(*79*). Each sequence contained 29 residues to allow approximately four helical heptad repeats. In the model, the superhelix refers to the helical bundle structure formed by the coiling of four alpha helices. An ensemble of the four alpha helical peptides was created by varying a set of associated geometrical parameters: the super-helical radius R, the super-helical phase α (α=45° when the neighboring helices are equal-distant from each other), the relative displacement of the ends of the helices parallel to the super-helical axis Z, the minor-helical phase θ (rotation of the alpha helix about its axis), and the super-helical pitch P (Figure 3.2). Given a set of these geometric parameters that specify the coordinates of a single helix, the remaining three helices were constructed by performing symmetry operations consistent with the D_2 point group.

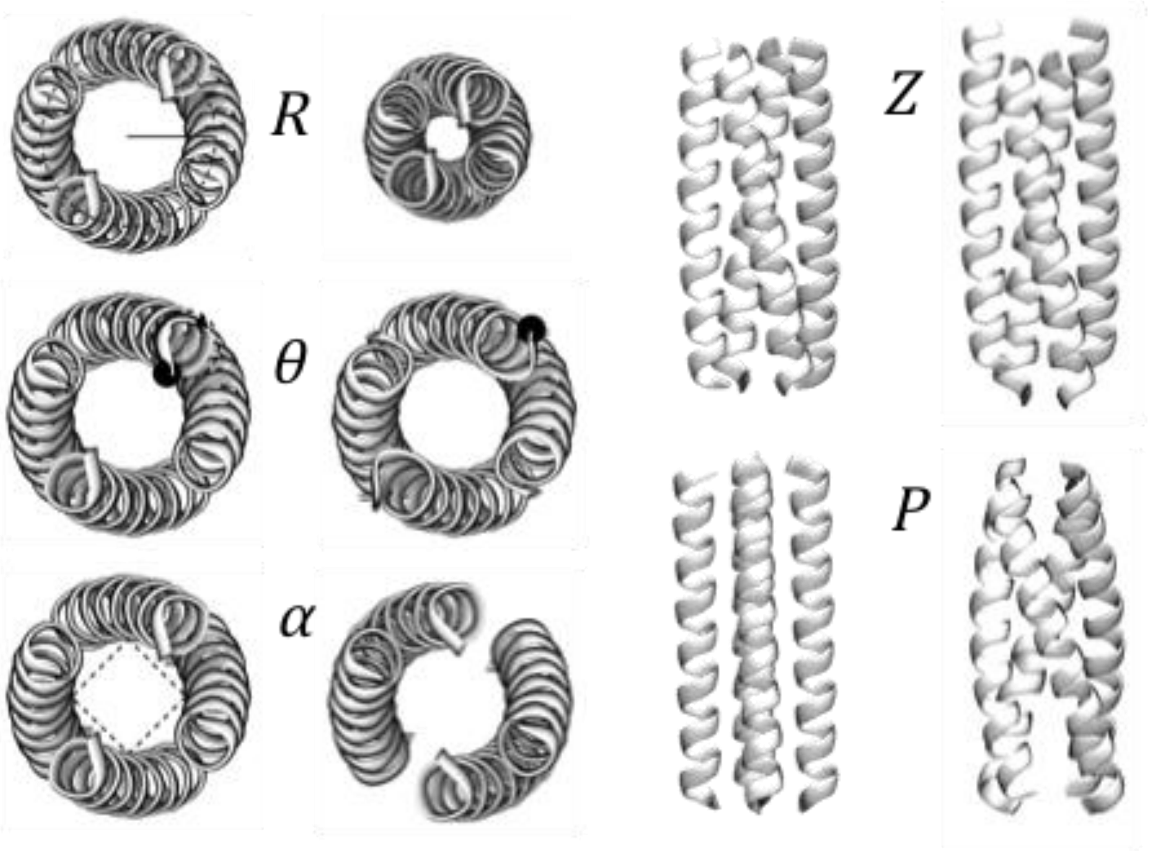


Figure 3.2: Representative backbone configurations of the helix bundle motif building block illustrating variation of the geometric parameters associated with the bundle.

Monte-Carlo simulated annealing (MCSA) was used to sample parameters and identify helical bundle sequences and structures as local minima on an energy surface. Let *{R, α, Z, P, θ}* denote the set of the parameters that characterize a particular backbone scaffold configuration. These parameters were confined within values associated with natural coiled-coil tetramers (*R*: 6.8Å -- 7.8Å, *α*: 35° -- 55°, *Z*: -1.5Å -- 3Å, *P: -118* -- ∞)(*78*). The values of *θ* were grouped into three categories of the bundle, based upon the location of heptad positions in the structure(*189*): *θ* = 10° -- 30° (interior heptad positions: a, d); *θ* = -10° -- 10° (interior heptad positions: a, d, e); *θ* = 30° -- 50° (interior heptad positions: a, d, f).

For each generated helical structure, a statistical sequence design methodology was used to calculate the probabilities of hydrophobic amino acids (A, V, I, L, M, F, Y and W) at interior residue positions. Sites 1 and 28 were located at the end of the helical motif and solvent exposed, thus defined as exterior positions. To estimate the probabilities, a statistical thermodynamic theory was applied, wherein an entropy or its Legendre transform are optimized subject to constraints on the sequences(*44*, *145*). Herein, the CHARMM22 force field(*93*) was used. As done in previous work, an average internal energy over the ensemble of sequences was calculated, and its conjugate temperature β^{-1} was specified such that β=0.5 mol/kcal. A helix propensity scoring function, E_h ,(*87*) was employed and constrained to values expected for helical peptides of the chosen length(*190*). For 29-residue helical segments in a database of natural protein structures, E_h takes on values in the range -12.1 *kcal/mol* to -7.5 *kcal/mol*; in the design calculations, values of E_h = -8.88 *kcal/mol*, -9.75 *kcal/mol*, and -10.18 *kcal/mol* were applied in separate Monte Carlo samplings. A symmetry assumption was applied to leverage the symmetry of the D_2 point group(*44*, *80*). For each sampled tetrahelical configuration, an average (internal) energy per peptide was calculated using the site-specific probabilities of the amino acids. This average energy was used as the objective function in 5000-step Monte Carlo searches with exponential annealing schedules.

Ten lowest energy structures were selected. The sequence considered in each case comprised the most probable amino acid at each residue position. For the resulting structures, void volumes were assessed by CastP(*191*). The final candidate was chosen because its total and maximum void volumes are comparable to those of 15 crystalized antiparallel homo-tetramers. This helix bundle structural motif, which comprised the tetrahelical structure and identities of 11 interior hydrophobic amino acids (Figure 3.3), was used in all the subsequent designs targeting distinct assemblies with predetermined nanostructures (Figure 3.1A, 3.1B, 3.1C, 3.1D).

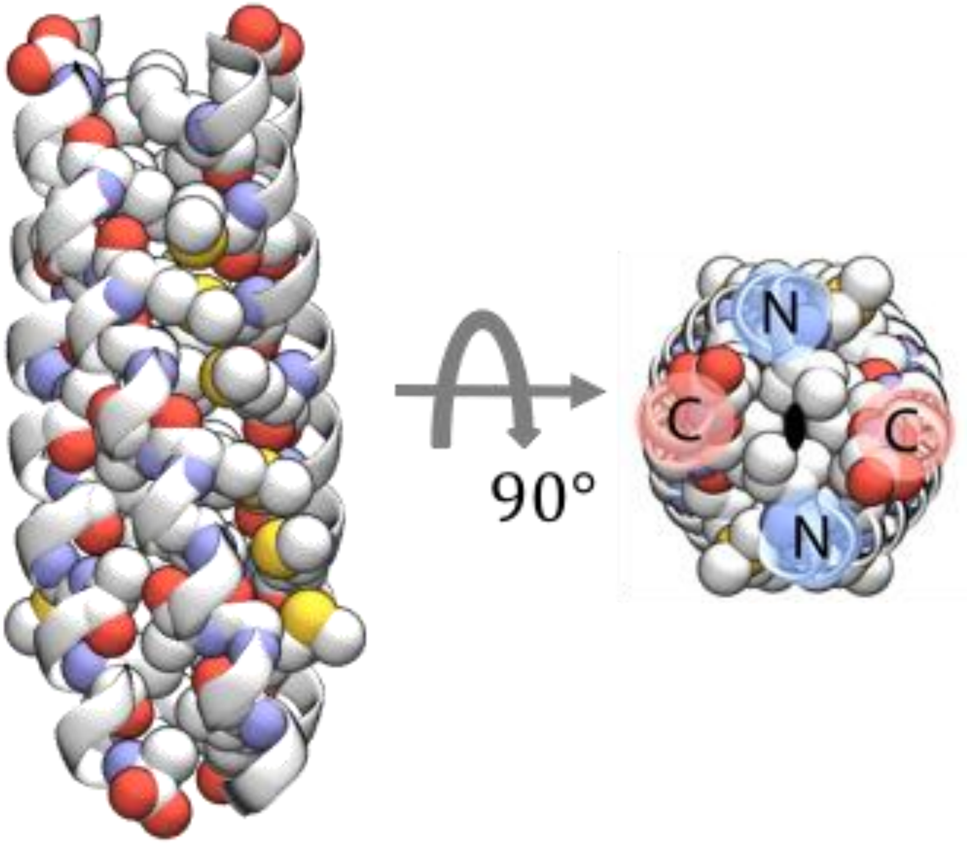


Figure 3.3: Side and top view of the selected low-energy helix bundle motif with the most probable amino acids at the interior sites shown in space-filling representations: N terminus is depicted in blue and C terminus in red circles.

3.3.1.2 Design of helix bundles targeting materials with predetermined nanostructures

The exterior residue positions of the helix bundles were determined so that self-assembly of the bundle motifs would produce assemblies with chosen, distinct nanostructures. The targeted assemblies were an isolated helix bundle (expected to remain soluble and not form higher order lattice assemblies in solution) and three distinct, two-dimensional lattices. Layered lattice symmetries were selected to satisfy the following criteria: 1) the lattice contains internal positions with D_2 point group symmetry on which the bundle motifs were positioned and 2) the lattice contains no axes of skew symmetry perpendicular to each layer. The space groups P222, P422, and P622 were selected. Each is a layer space group, and only a single, isolated layer was used in the design calculations.

Peptide sequences were identified that are compatible with the targeted nanostructure. The statistical mechanical approach mentioned previously was applied and symmetry assumptions were imposed to incorporate point and space group symmetries. For each targeted material, peptide sequence design was performed with the designed helix bundle building block motif, allowing 18 natural amino acids (all natural amino acids but Pro and Cys) at each of 18 variable exterior residue positions. The designed soluble, non-lattice forming helix bundle is labeled BNDL_1 (Figure 3.1A), and the quality of the structure was validated by Molprobity(*192*). For design in the context of one of the chosen lattice symmetries, unit cell parameters were modulated to vary the spacing between neighboring bundle motifs in the search for low-energy lattice structures. For P222, the unit cell dimensions *a* and *b* were varied. For P422 and P622, there is only one unique unit cell dimension *a*. For P622, an additional set of distinct lattice structures containing a 90° rotation about an interior C_2 symmetry axis was also considered. A grid search was used (at 0.1 Å increment) to construct the sequence-structure energy landscapes for each lattice nanostructure. For each lattice, the average energy over the sequence probabilities was calculated using the statistical design approach. Low energy candidates were selected from local minima on the landscape of each type of lattice. If Trp or Tyr were not among the most probable amino acid at any of the variable residues, sites where Trp or Tyr were probable were constrained as one of these two amino acids in subsequent calculations; these residues were introduced to facilitate determination of peptide concentrations. The numbers of lowest-energy sequences selected for further characterization were 4, 5, and 14 for the P222, P422, and P622 lattices respectively. The candidates' Molprobity scores fall within the range observed for 53 solved structures of coiled-coil proteins from the CC+ database. In the last step, PDBePISA(*193*) was used to assess the candidates' potential to form desired assemblies and the resultant protein-protein interactions at the bundle-bundle interfaces. Based upon these assessments, a total of 9 designed sequence candidates were chosen for experimental synthesis and characterization (Table 3.1).

3.3.2 Solution assembly and characterization

In solution, all but one of the computationally designed peptides exhibited the intended assembly properties: the peptide designed using the isolated tetramer template (Figure 3.1A) remained soluble and did not form higher order structure, while those designed in the context of a lattice (Figure 3.1B-D) formed some sort of regular array. Nine designed sequences were selected for experimental investigation (Table 3.1). The single sequence designed to form isolated bundles (BNDL_1) formed soluble homotetramers that did not undergo further interbundle assembly. Seven molecules computationally designed to form material assemblies with specific, interbundle packing did, in fact, produce sheet-like nanomaterials. No assembly was observed for the sequence P222_4 (Table 3.1).

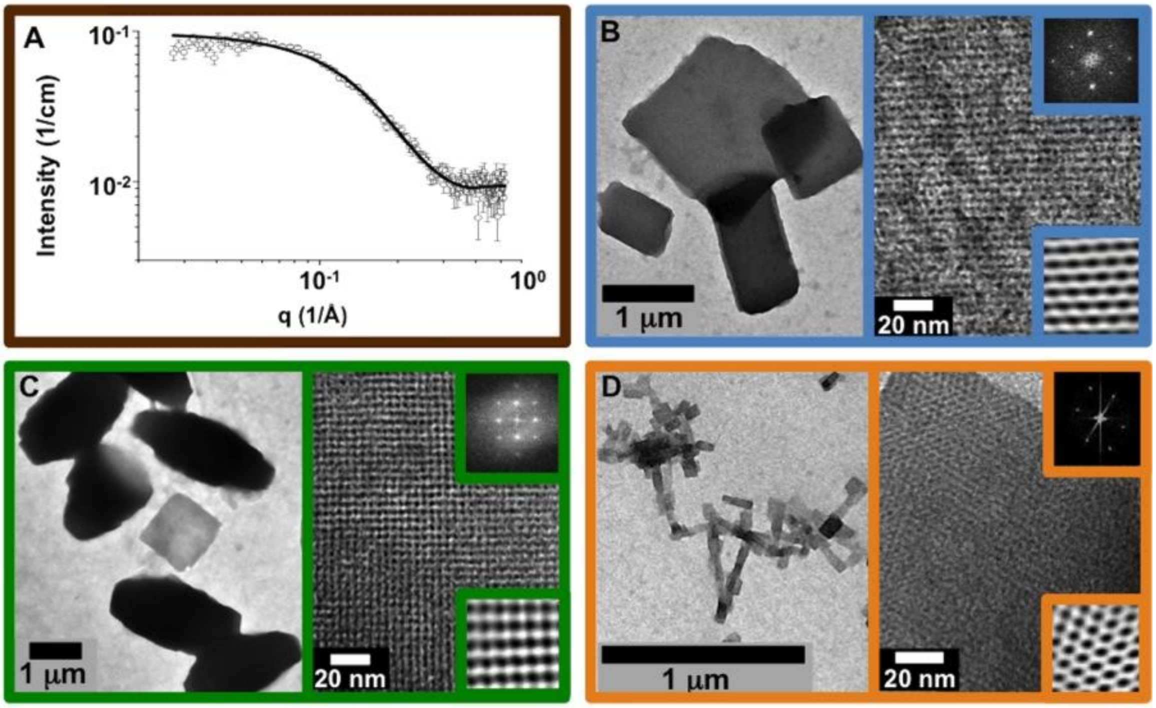


Figure 3.4: **A)** Small-angle neutron scattering data and nanocylinder fit (black curve) of BNDL_1 assembled from 5 mM peptide solution in borate buffer, pH10. Fit provides a cylinder length of ~3.5 nm and radius of ~ 1 nm consistent with tetrameric coiled coil soluble bundle design. Peptide solutions heated to 80°C to obviate inter or intramolecular structure and then allowed to cool to room temperature for **B-D)** intermolecular assembly. **B, left)** Low magnification cast film TEM image of P622_6 assembled from 1 mM peptide solution in phosphate buffer, pH7. **B, right)** High magnification of negatively stained lattice consistent with P622 symmetry. Upper inset is the FFT calculated from the high magnification TEM data while the lower inset is the inverse FFT (IFFT) calculated using the FFT maxima. **C, left)** Low magnification cast film TEM image of P422_1 assembled from 1 mM peptide solution in borate buffer, pH10. **C, right)** High magnification of negatively stained lattice. Upper and lower insets are the FFT and IFFT, respectively. D, left) Low magnification cast film image of P222_1 assembled from 1 mM peptide solution in phosphate buffer, pH7. **D, right)** High magnification of positively stained lattice. Upper and lower insets are the FFT and IFFT, respectively.

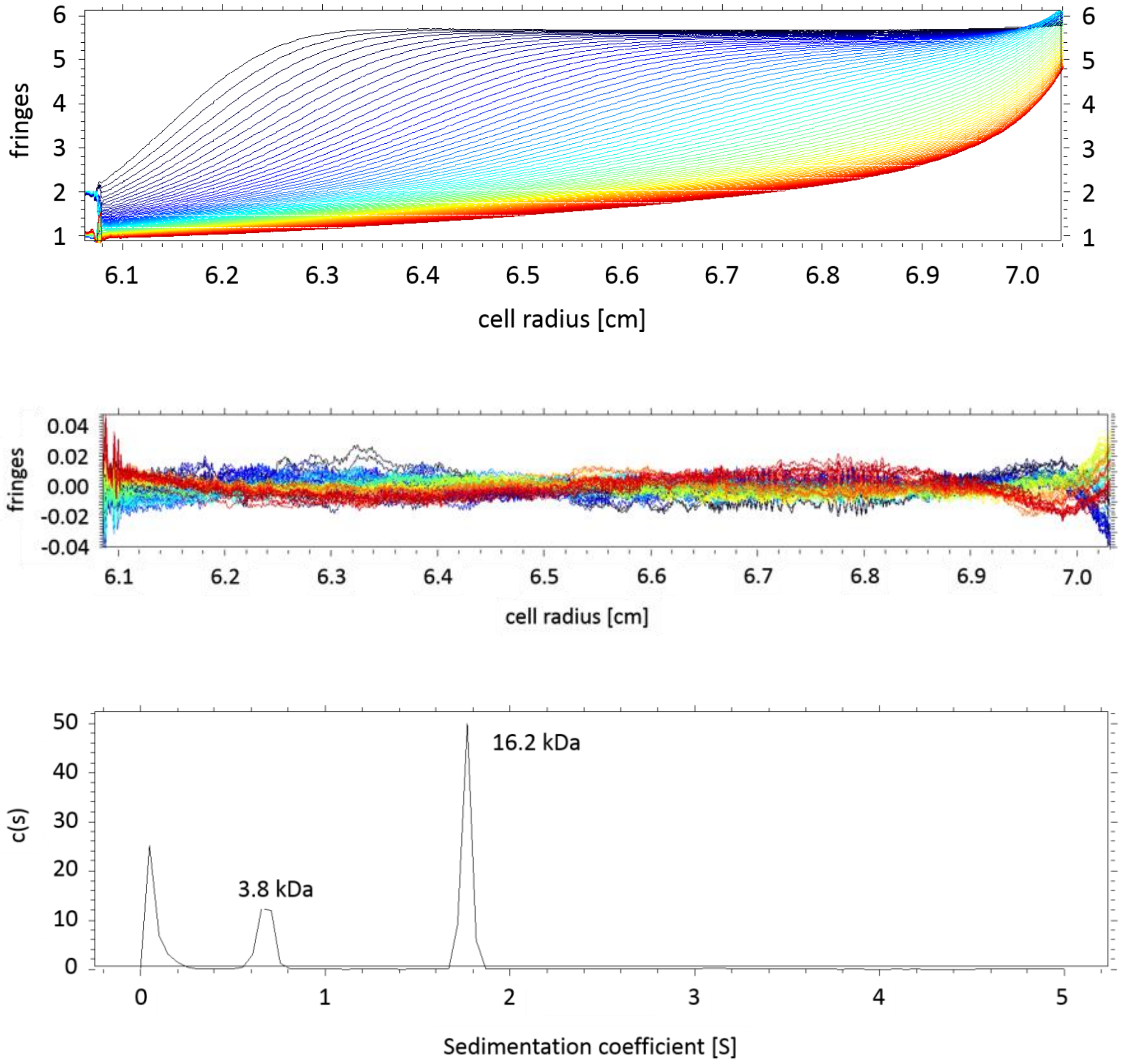


Figure 3.5: Analytical ultracentrifugation data and analysis of BNDL_1. Top: Experimental interference pattern including best fit of the sedimentation velocity data from 6.086 to 7.032 cm. Sample BNDL_1 at 0.5 mM in borate buffer (pH 10). Middle: Residual plot of the fitted sedimentation velocity data along the radius of the cell. Bottom: Distribution plot showing one solution population at ~ 3.8 kDa consistent with a peptide monomer and one at ~16.2 kDa consistent with a peptide tetramer or pentamer. The sedimentation coefficient, S, is directly proportional to M_f (buoyancy mass). The peak at very low S values is most likely due to residual TFA from the synthesis and purification of the peptide. Statistics of the data fit are: n =98116, root-mean-square deviation, rmsd =0.0061, sum of squares of the regression, SSR = 3.62, signal/rmsd > 1000.

Figure 3.4 presents representative data from four different peptide molecules theoretically designed to produce the four chosen nanostructures in Figure 3.1. Figure 3.4A shows small-angle neutron scattering (SANS) results for BNDL_1 modeled as a short cylinder with dimensions consistent with the designed, tetrameric coiled coil. Analytical ultracentrifugation (AUC) supports the presence of tetrameric helix bundles coexisting with a minority of monomeric peptide (Figure 3.5). BNDL_1 solutions were experimentally monitored for three months, and no interbundle lattice structure was observed. Figure 3.4B shows the nanostructure formed from assembled P622_6 molecules that is consistent with hexagonal local symmetry and the targeted unit cell parameters (Table 3.2). The structure is stabilized by two unique bundle-bundle interfaces that arise from the design around the three-fold symmetry axis (Figure 1B). Both interfaces, one between two antiparallel helices and the other between two parallel helices, are populated by salt-bridges, hydrogen bonds and hydrophobic interactions in the computational design. Figure 3.4C shows materials formed from the self-assembly of P422_1 that have the targeted four-fold-like symmetry. While four-fold symmetry is clear in Figure 3.4B, the experimentally observed unit cell dimensions are larger than predicted (Table 3.2). Figure 3.4D shows materials formed from the self-assembly of P222_1. The observed morphology consists of regular nanosheets having the targeted rectangular structure, but differences are observed in symmetry and unit cell spacings from what was theoretically designed (Table 3.2). Although the targeted lattice parameters are not precisely recovered experimentally in each case, taken together, the data support the use of computational design to realize a set of homotetrameric helix-bundles having predetermined self-assembly properties (e.g. sheet-forming tetrameric peptide bundles vs.

soluble bundles) and distinct, local ordering that is determined by the exterior surface residues of the helical bundle.

	Design			TEM		
	a (nm)	b(nm)	γ(°)	a(nm)	b(nm)	γ(°)
P622_6	4.57	4.57	120	4.49 ± 0.25	4.54 ± 0.25	112.7 ± 0.4
P422_1	3.12	3.12	90	4.17 ± 0.17	3.90 ± 0.17	88.9 ± 0.9
P222_1	2.09	2.00	90	3.34 ± 0.25	3.21 ± 0.25	100.4 ± 0.9

Table 3.2: Lattice parameters of the self-assembling peptides from the design in comparison with those determined from analysis of Fourier transforms of the TEM images in Figure 3.4. *a* and *b* denote the dimensions of the two-dimensional unit cell, □ denotes the interior angle defined by sides *a* and *b*.

3.3.3 Covalent modifications and solution modulations

The robustness and versatility of the designed bundle-forming peptides were further explored by changing solvent conditions. Different solution conditions, selected on the basis of the physicochemical properties of the bundles, could be used to manipulate interbundle assembly and alter the resulting superstructural morphology of the nanomaterials. For example, Figure 3.6A and 13B reveal that micron-scale morphology of P622_2 can be manipulated simply by first melting secondary structure in solution at 80 °C and subsequently quenching to two different temperatures. Smaller particles were formed at the higher quenching temperature (50 °C) than at the lower quenching temperature (25 °C). The data suggest that a higher temperature results in a much slower assembly process. Figure 3.6C and 13D reveal the sensitivity of select bundle assemblies to changes in pH. Molecule P222_9 has a theoretical pI=7 and was assembled under two solution conditions that differed only in pH. Clearly, assembly at pH 7 resulted in two-dimensional, plate-like growth, while at pH 10, assembly yielded anisotropic growth and the formation of long, needle-like structures. The thickness of the needles prevented clear, high magnification lattice imaging, but the layer spacings of the underlying lattice (Figure 3.7) were consistent with the nanostructure observed for molecule P222_9 (Figures 8 and 11). The results suggest that at pH 10 there is a clear preference in growth direction during helix bundle solution assembly.

Covalent alteration of the original, designed peptides also can be employed to probe the robustness of the assembly as well as to modulate assembly at particular solution conditions. For example, acetylation of the N-terminus of P222_9 (denoted P222_9_Ac, Table 3.1) reduces its theoretical isoelectric point from pI=7 (P222_9) to pI=4 (P222_9_Ac), allowing for assembly at low pH. At pH 4.5, P222_9 remains dissolved whereas P222_9_Ac (Figure 3.6E) assembles into the same nanostructure as that observed for P222_9 at pH 7 (Figure 3.8). Similarly, P422_1_Ac forms plate nanostructures at pH 8 (Figure 3.6F), whereas P422_1 assembles at pH 10 (Figure 3.4B). Therefore, one can use covalent modifications of the designed peptide sequences to alter solution conditions in which nanostructures can be formed. Adding residues to the termini of the originally predicted sequences did not disrupt solution self-assembly. Specifically, sequences of two, four and six glycine residues were added to the N-terminus of P222_9 (Table 3.1), resulting in quite uniform assemblies (Figure 3.6G, H) with similar lattice structures to those observed for the unmodified P222_9 (Figure 3.9). These observations reveal the robust nature of these theoretically designed peptides and their material assemblies. Altogether, the data of Figure 3.6 exemplify that multiple modifications of the theoretically predicted sequences remain competent for forming the selected nanostructures.

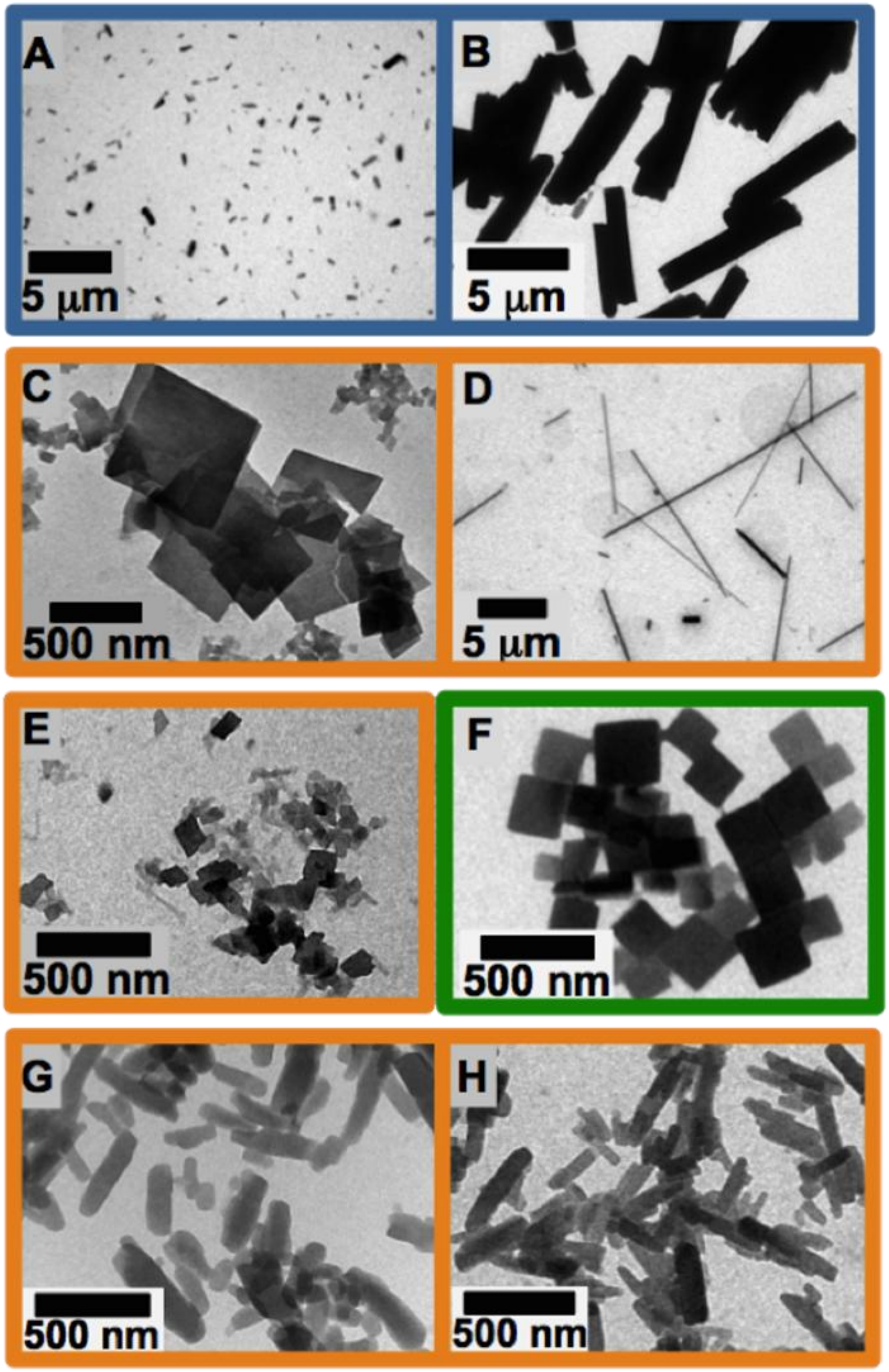


Figure 3.6: Cast film TEM examples of morphology control with manipulation of solution assembly conditions and peptide primary structure. All sample solutions heated to above 80 °C for 1 hour to obviate any assembled or secondary structure before respective cooling treatment. **A,B**: 0.5 mM P622_2 peptide at pH 7 (phosphate buffer) quenched to **A)** 50 °C vs. **B)** 25 °C , imaged after 1 day. **C,D**: 1.0 mM P222_9 ambiently cooled to room temperature with **C)** pH 7 (phosphate buffer) vs. **D)** pH 10 (borate buffer)

showing a clear difference in superstructure growth. **E,F**: Plates grown from 0.1 mM peptide solutions with peptide primary structure altered through acetylation of the Nterminus. **E)** Ambient cooling to room temperature allowing assembly P222_9_Ac at low pH of 4.5 in sodium acetate buffer and **F)** P422_1_Ac quenched to 50 °C at pH 8 in phosphate buffer. G,H: Plates grown from 1.0 mM peptide solutions at pH 7 (phosphate buffer) after ambient cooling to room temperature with P222_9 peptide primary structure altered through addition of **G)** 4 glycines vs. H) 6 glycines to the Nterminus of the P222_9 peptide molecule.

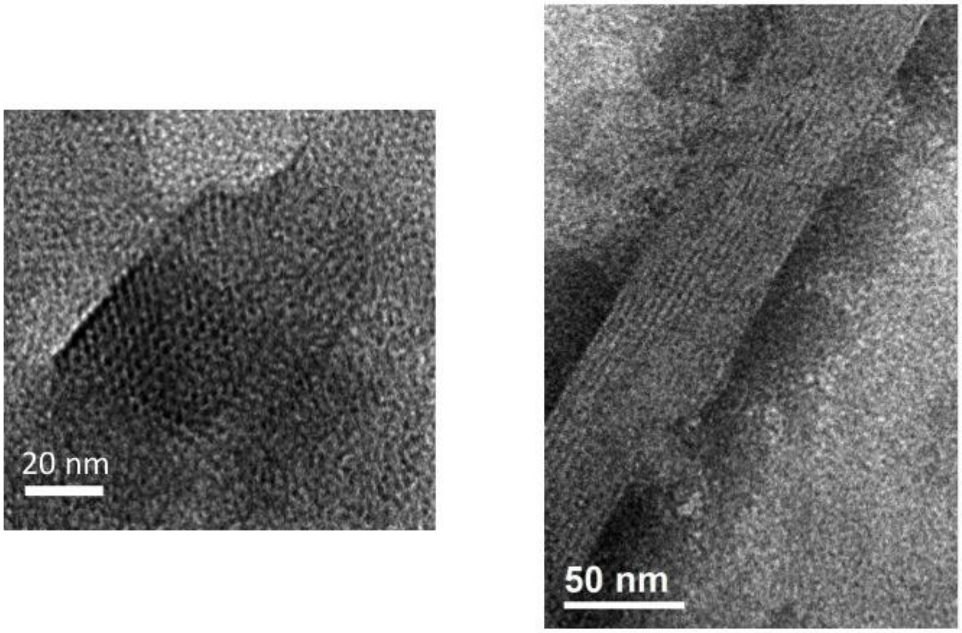


Figure 3.7: High magnification TEM of 1.0 mM P222_9 ambiently cooled to room temperature from 80 °C at pH 7 (left) and pH 10 (right).

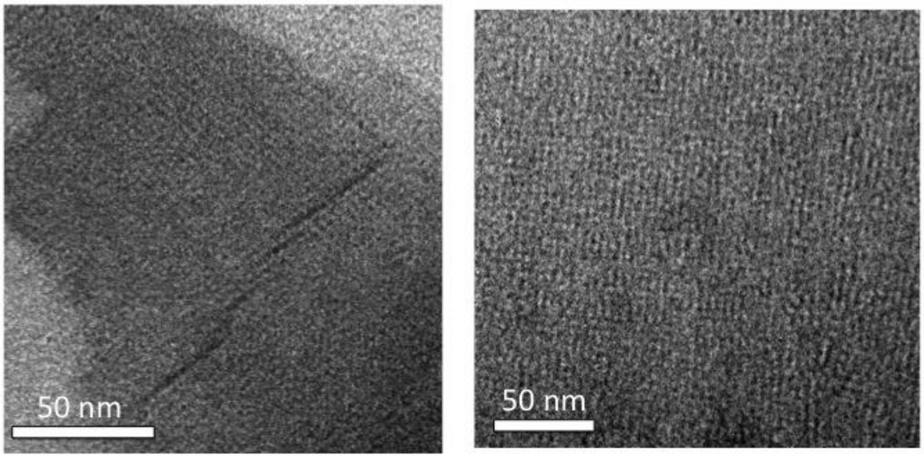


Figure 3.8: High-magification TEM of (left) lattice of P222_9_Ac at pH 7 ambiently cooled to room temperature from 80 °C and (right) P422_1_Ac at pH 8 quenched to 40 °C from 80 °C.

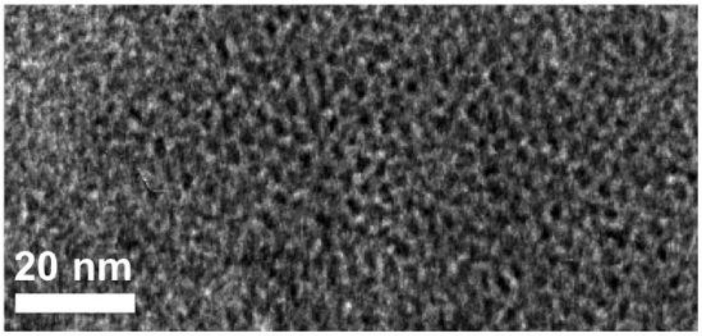


Figure 3.9: High-magnification TEM of lattice of P222_9_6Gly at pH 7 ambiently cooled to room temperature from 80 °C

3.4 Discussion

The data of Figures 3.4-3.9 support that theoretically designed sequences listed in Table 3.1 self-assemble to form predetermined sheet nanostructures or soluble bundles. Specifically, the sequence (BNDL_1), targeted to form soluble, non-associating bundles, formed homotetramers in solution but did not exhibit further lattice assembly. Similarly, the sequence (P622_6) was designed to form two-dimensional sheets with hexagonal local symmetry and formed a nanostructure consistent with that identified in the computational design of the sequences; this consistency is evidenced in the similarity of the unit cell parameters of theoretical model and that derived from analysis of TEM data.

3.4.1 Computational investigation of the lattices formed by P422_1 and P222_1

As shown in Figure 3.4, the molecules designed to produce two-dimensional plates with P422 symmetry do, in fact, assemble into a two–dimensional sheet nanomaterial with local four-fold bundle packing symmetry. However, the lattice spacing formed experimentally is larger than that expected from the computational design. Similarly, the molecules predicted to yield the two-dimensional materials with P222 symmetry assembled into sheet nanostructures. However, the lattice symmetry and unit cell parameters were different than those identified in the computational modeling. Additional calculations using the P422_1 molecule sequence within a related four-fold symmetry, P4, reveal a local energy minimum at the experimentally determined unit cell parameters (Figure 3.10). The distance between the C_2 symmetry axes of neighboring bundles within each model structure is essentially indistinguishable: 2.2 nm (P4) vs. 2.2 nm (P422). With the P4 structure, the unit cell contains 4 helical bundles (16 peptides) as opposed to 2 helical bundles (8 peptides) in the original model with P422 symmetry, which is consistent with the larger unit cell dimensions observed experimentally. Similar calculations involving P222_1 within a related two-fold symmetry, P2, reveal a local minimum that is consistent with the experimental observations (Figure 3.11).

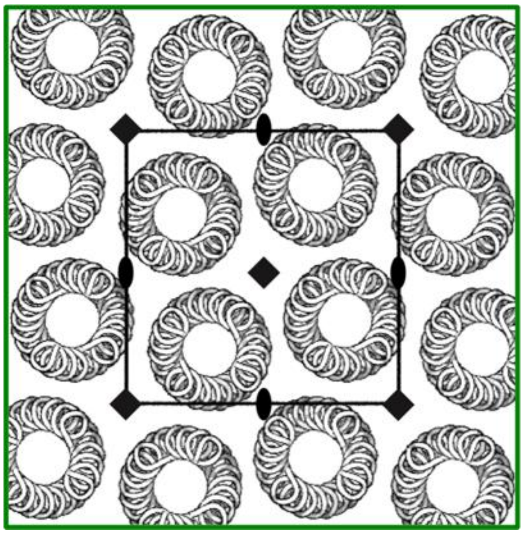


Figure 3.10: Putative structure of assembly of P422_1 helix bundles packed with P4 symmetry, which is a local minimum within the structure energy landscape with respect to variation of the unit cell parameter. The dimension of the unit cell is *a* = *b* = 4.25 nm, consistent with that observed experimentally.

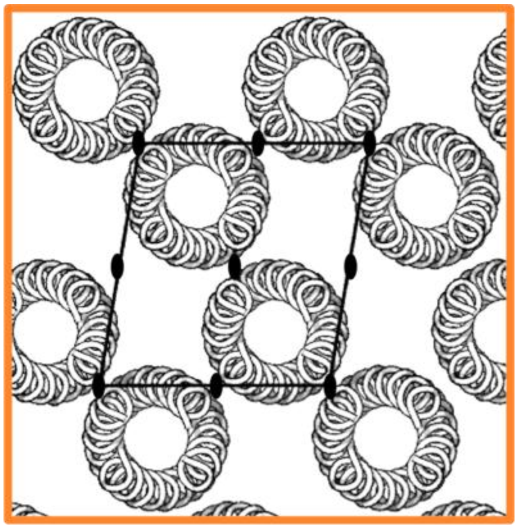


Figure 3.11. Putative assembly structures of P222_1 assemblies with P2 symmetry located at a local minimum of the structure energy landscape with respect to variation of the unit cell parameters. The dimensions of the unit cell are within the range of those observed experimentally: *a* = 3.24 nm, *b* = 3.03 nm and *γ*= 99.6°.

3.5 Conclusion

We have presented an approach for the de novo design of peptide assemblies that assemble into nanomaterials with predetermined local structures. A versatile, helical homotetrameric building block structure was computationally designed that was stabilized largely by the formation of a hydrophobic interior. Subsequent computational determination of the bundle exterior residues resulted in solution-assembled materials with predetermined morphologies (two-dimensional sheet vs. soluble bundle) and distinct, local nanostructures that varied with the bundle exterior sequence. Integrated theoretical and experimental studies will guide future refinements for predictably achieving specific, desired nanostructures and elucidating the principles underlying their formation. In addition to achieving symmetric assemblies of biopolymers, the display of a wide variety of chemical functional groups within designed assemblies offers abundant opportunities for hierarchical pathways for nanomaterial production, such as biopolymer templated growth and/or assembly of inorganic phases with nanoscale precision. Computational design combined with the experimental control of assembly pathways has the potential to provide exquisite control over new materials with desired nanostructures.

3.6 Materials and methods

Peptide synthesis: Peptides were prepared at a 0.25 mmol scale on Rink amide resin using an AAPPTec Focus XC synthesizer (AAPPTec, Louisville, KY). Standard Fmoc-based protocols were employed. Peptides were deprotected for 5 and then 10 minutes with 20% piperidine in dimethylformamide (DMF). The coupling reaction was conducted for 40 minutes with 4 eq. of the appropriate amino acid dissolved in N-methyl-2-

pyrrolidone (NMP) (5 mL), 3.8 eq. HCTU dissolved in DMF (2.5 ml), and 8 eq. diisopropylethylamine dissolved in NMP (1 ml). Five washes were performed in between steps with 50:50 (v:v) DMF:methylene chloride (12 mL) for the first two washes and DMF (10 mL) for the last three. Amino acids, resin and activator were purchased from ChemPep (Wellington, FL) and used as received. All solvents were analytical grade (Fisher Scientific). Peptide cleavage was achieved by shaking peptide solutions for 2 hours in a cleavage cocktail comprising (by volume) 95% trifluoroacetic acid (TFA), 2.5% tirisoproylsilane, and 2.5% Milli-Q water. The peptide was then precipitated by adding the cleavage cocktail and cleaved peptide to diethyl ether, and the mixture was then centrifuged and the supernatant discarded. The process of suspending in diethyl ether, centrifuging, and discarding the supernatant was repeated a total of three times. The resulting peptide was then dissolved in water and lyophilized.

Peptide purification: Purification was performed via reverse-phase HPLC using a BEH130 Prep C18 10 μm column (XBridge, Waters Corporation, Milford, MA). Crude peptides were dissolved in Milli-Q water containing 0.1%-vol TFA, and were filtered (0.20 μm filter, Corning, Inc., Corning, NY) before HPLC injection. Products were subjected to an elution gradient (Quaternary Gradient Module (Waters 2545), Waters Corporation) of 100% solvent A (0.1%-vol TFA) to 30% solvent A within 60 min; the composition of solvent B was acetonitrile with 0.1%-vol TFA. Fractions were detected using UV-Vis detection at 214 nm (Waters 2489, Waters Corporation) and collected (Waters Fraction Collector III, Waters Corporation). The collected fractions were examined by ESI-mass spectrometry (LCQ Advantage Mass Spectrometer System, Thermo Finnigan, San Jose, CA) with an auto sampler system (Surveyor Autosampler, Thermo Finnigan). Pure fractions were combined and lyophilized.

Analytical HPLC: Purity analysis was performed via reverse-phase Analytical scale HPLC using a BEH C4 3.5 μm column (XBridge, Waters Corporation, Milford, MA). Peptides were dissolved in 80/20-vol Milli-Q water/ACN containing 0.1%-vol TFA, and filtered with 0.20 μm filter (Corning, Inc., Corning, NY) before injection. Products were subjected to a linear elution gradient (Waters 600 Controller, Waters Corporation, Milford, MA) of 80% solvent A (Milli-Q water with 0.1%-vol TFA) to 5% solvent A in 70 min; the composition of solvent B was acetonitrile with 0.1%-vol TFA. Fractions were detected using photodiode array detector (Waters 2996, Waters Corporation, Milford, MA) tuned at 214 nm.

Circular dichroic spectroscopy (CD): Secondary structures and the temperature dependent behavior of the synthetic peptides were analyzed using circular dichroic spectroscopy on a Jasco J-820 spectropolarimeter (JASCO, Inc., Easton, MD). Sample solutions were prepared at 0.1 mM concentration in 10mM of buffer appropriate for the desired pH and were transferred into an absorption cuvette with 1-mm path length (110-QS, Hellma, Inc.). Pure buffer solutions were used for the background correction. Sample spectra were recorded from 190-250 nm at desired temperatures. Data points for the wavelength-dependent CD spectra were recorded at every nanometer with a 1nm bandwidth and a 4-second response time for each data point. The ellipticity at 222 nm was used to monitor the temperature-dependent unfolding and refolding of the peptides. Data points for the kinetics scans were recorded at 222 nm at 1-minute intervals. The mean residue ellipticity, $[\theta]_{MRE}$ (deg cm2 dmol^{-1}), was calculated using the peptide concentration, number of amino acid residues, and cell path length.

Transmission electron microscopy: Carbon-coated 200 mesh copper grids (CF200- Cu, Electron Microscopy Sciences, Inc.) were freshly treated by glow discharge using a plasma cleaner (PDC-32G, Harrica Plasma, Inc.) before sample preparation. 3 μL of sample suspension was applied onto the grids. After ~5 min, any remaining liquid was blotted from the edge of the grids using filter paper. Then 3 μL Milli-Q water was applied the grids and blotted immediately to remove excess unassembled peptides and buffer salts. The grids were incubated under ambient conditions for another 30 min before TEM observation (Tecnai 12, FEI) or staining. To negatively stain the grids, 3 μL of an aqueous solution of uranyl acetate (1% wt.) was applied to the cast-film grids and incubated for 20-30 seconds, then blotted with filter paper. The stained grids were allowed to sit for at least 10 min under ambient conditions before TEM observation. Positive staining of the samples was achieved by exposing the dried cast-film grids to ruthenium tetroxide aqueous solution vapor (0.5 % w/v) for 5-10 minutes. The TEM was operated at an acceleration voltage of 120 kV, and all images were recorded digitally using a Gatan multiscan 791 side-mounted CCD camera. Lattice plane spacings were determined by conducting a FFT of the area of interest of a TEM micrograph using DigitalMicrograph software v2.3 (Gatan Inc., Pleastanton, CA, USA) and the PASAD plug-in.31 FFT intensity was plotted relative to radial distance from the origin as well as relative to azimuthal angle. Corresponding real space values of interplanar distances and angles were calculated in order to estimate unit cell parameters. Uncertainties of the unit cell parameters were approximated by the size of the maxima in the FFT. Fourier filtered images (lower insets in Figure 3.4) were obtained by applying masks

Small angle neutron scattering (SANS): SANS measurements were conducted at the National Center for Neutron Research (NCNR), National Institute of Standards and Technology, Gaithersburg, MD on the NG-7 30m SANS beamline. Samples were dissolved in 10 mM pH 10 borate buffer prepared in D_2O and heated to 80 °C for 1 hour to melt any organized secondary structure/aggregated structure and subsequently cooled to room temperature for intermolecular assembly. A neutron beam with a mean wavelength of 6 Å was defined using a mechanical velocity selector. The wavelength spread ($\Delta\lambda/\lambda$) was 0.15 at full width half max. The 640 mm x 640 mm ^{3}He proportional counter used has a spatial resolution of 5.08 mm x 5.08 mm. Sample-to-detector distances of 1, 4, and 13 m were used to provide a q range of approximately 0.004 to 0.500 Å^{-1}, where q is the scattering wave vector defined by $q = (4/\lambda) \sin(\theta/2)$. Data obtained on this instrument was corrected for background noise and radiation, detector inhomogeneity, as well as empty cell scattering. Intensities were normalized to an absolute scale relative to the empty beam transmission. The uncertainties of individual data points were calculated statistically from the number of averaged detector counts.

SANS curves for BNDL_1 were fit using the SasView software. A cylinder fit was performed using the equations below for the form factor of a right circular cylinder with a uniform scattering length density(*194*, *195*). The form factor, $P(\boldsymbol{q})$, is represented by the following:

$$P(q) = \frac{scale}{V} \int_0^{\pi/2} f^2(q, \alpha) \sin\alpha \, d\alpha + bkg$$

$$f(q) = 2(\Delta\rho)V * \frac{\sin\,(q\,L\,cos\alpha/2)}{q\,L\,cos\alpha/2} * \frac{J_1(q\,r\sin\alpha)}{(q\,r\sin\alpha)}$$

$$V = \pi r^2 L$$

where $J_1(x)$ is the first order Bessel function. α is defined as the angle between the cylinder axis and the scattering vector, $\boldsymbol{q}$. The integral over alpha averages the form factor over all possible orientations of the cylinder with respect to $\boldsymbol{q}$. L is the cylinder length while r is the cylinder radius. The model fit produced a length = 34.4 +/- 1.1 Å and radius = 10.4 +/- 0.2 Å for BNDL_1 as shown in Figure 3.4D.

Analytical ultracentrifugation (AUC): AUC experiments were run on a Beckman Coulter ProteomeLab XL-I instrument. 400μL of 0.5mM peptide solution in buffer (10 mM borate buffer with extra 50 mM NaCl to screen long range coulombic interactions, pH10) was transferred into a two-channel cell equipped with sapphire windows. The cells were mounted into a 4-cell An-60 Ti analytical rotor and equilibrated at 20°C for 2 hours in the rotor chamber. Sedimentation velocity experiments were carried out at 50,000 rpm at 20°C using 800 scans with one scan per minute per sample. The changes of the concentration profiles along the cell radius were monitored using Rayleigh interference optics. All data were analyzed using the SEDFIT package v14.81.32.(*48*)

CHAPTER 4 Crystal Structure of a Computationally Designed Lattice-forming Peptide

Adapted from the manuscript in preparation (2016)

4.1 Abstract

Computational design of self-assembling peptides/proteins offers effective routes to engineer advanced novel biomaterials of various nanostructures with precise control over complex 3D structures via amino acid sequences. The complexity and subtlety of the noncovalent interactions at the assembly interfaces make the consistent success of such designs very challenging, and usually requires experimental screenings of a large number of candidates. Herein, we report the crystal structure and the solution phase characterizations of a computationally-designed peptides that robustly self-assemble into lattices. The study reveals a universal helical bundle motif applicable as the building block to design different peptide lattices. The critical role of specific hydrophobic interactions in stabilizing bundle-bundle interface in the crystal assembly is discussed. Additionally, further energetic analysis suggests fundamental ways to potentially improve the success rate of future computational designs of protein- and peptide- based lattices.

4.2 Introduction

Nature utilizes assemblies to achieve rich functionality through complex, yet highly-ordered structures made of biologically derived molecules, such as laminin(*4*), S-layer proteins(*196*, *197*), and nacre in shells(*6*). Such biological assembly processes are often dependent on the environment conditions or require external energy input, and thus are often hard to reproduce synthetically. Inspired by nature, designed peptide and protein self-assemblies and derivatives have achieved similar periodic spatial arrangements and have shown great promises in various biomaterial applications for catalysis(*47*, *198*, *199*), controlled drug delivery(*68*, *200*), light harvesting(*201–203*), and in-vivo imaging(*204*). Efforts towards enhanced nanomaterial applications include designed cage systems for encapsulation of nanoparticles, immobilization of metal complexes (*47*, *68*), and designed fibril, ribbon, and sheet systems to provide templating surfaces for chemical modifications and organic chromophores(*199*). Advantages of the designed peptide/protein assemblies include their spontaneous and usually robust self-assembly via careful selection of the amino acid primary sequences. Rational design methods are commonly used, where starting from self-assembling sequence patterns in natural proteins or well-studied motifs, such as peptide amphilphiles, sequences are modified, extended, and/or linked based on chemical properties of amino acids. Such strategy has been successful in generating one-dimensional (fibril-like)(*31*, *34*), two-dimensional (sheet-like)(*38*) and three dimensional self-assembling lattices(*35*, *41*). Fusion of natural proteins with known and matched rotational symmetry is another valid approach in rational design of self-assembling protein lattices(*36*, *77*).

Alternatively, computational design methods, where energetics of the assemblies are estimated by theoretical models and are used to guide the choice of sequences, allow for consideration of larger sequence variability, the possibility to engineer a wide variety of nanostructures, and the control of structures with increased precision. Computational design opens possible routes to the precise prediction of multicomponent and hierarchical assembly structures down to atomistic resolution with controlled geometry, orientation, and symmetry(*39*, *40*, *44*). This has allowed unprecedented chemical diversity for the designed protein/peptide assemblies and shows great potential for the discovery of advanced functional biomaterials.

Computational design usually relies on approximations of nonvalent interaction energetics, discrete amino-acid conformations, and a simplified solvent environment. However, the imperfection in these approximations poses challenges in achieving a high success rate of the design predictions among the selected candidate sequences(*37*, *39*, *64*). Yet the fundamental causes of low success rate are rarely explored, partially because the molecular structures of the designed sequences are difficult to characterize to the atomistic level without their self-assembling robustness to provide large 3D crystals of at least tens of microns for single-crystal x-ray diffraction studies. Neither are the potential improvements on the energetic approximations used in computational designs discussed in the context of designing protein- and peptide- based self-assemblies.

Most previous computational designs of 2D lattices and 3D crystals rely on natural proteins and only involve the redesign of protein-protein interfaces in the context of the targeted nanostructures(*37*, *38*, *44*–*46*). The design of protein lattices from sequences out of the natural realm remains extremely challenging due the large ensemble of complex intramolecular and intermolecular interactions that guide the folding of individual proteins and associations at protein-protein interfaces. The use of peptides, which usually are no larger than 50 amino acids, greatly reduces the ensemble size and complexity, and thus allows for the engineering of tertiary folds as well as assembly nanostructures from scratch at reasonable computational and synthetic cost. On the other hand, peptides, considering their shorter amino-acid sequence, might not be as well-structured nor provide as diverse chemical functionality as larger proteins, such as enzymes. Nonetheless, peptide-based lattices can be considered precursors and templates for further functionalized nanomaterials.

In this work, we report the crystal structure in addition to solution phase characterizations of a computationally designed lattice-forming peptide P422_1, and reveal the validity of a universal helical bundle motif that can be used as a building block for peptide-based lattice nanomaterials. We discuss the critical role of specific hydrophobic interactions for stabilizing peptide interfaces in the tightly-packed crystal. This study is also of great importance to provide feedback to the computational design methods and potentially improve the overall design success rate.

4.3 Materials and methods

Peptide synthesis: Peptides were prepared at a 0.25 mmol scale on Rink amide resin using an AAPPTec Focus XC synthesizer (AAPPTec, Louisville, KY). Standard Fmoc-based protocols were employed. Peptides were deprotected for 5 min and then 10 min with 20% piperidine in dimethylformamide (DMF). The coupling reaction was conducted for 40 min with 4 eq. of the appropriate amino acid dissolved in N-methyl-2-pyrrolidone (NMP) (5 mL), 3.8 eq. HCTU dissolved in DMF (2.5 ml), and 8 eq. diisopropylethylamine (DIEA) dissolved in NMP (1 ml). Five washes were performed in between steps with 50:50 (v:v) DMF:methylene chloride (12 mL) for the first two washes and DMF (10 mL) for the last three. Amino acids, resin and activator were purchased from ChemPep (Wellington, FL) and used as received. All solvents were analytical grade

(Fisher Scientific). After last step deprotection, peptide cleavage was achieved by shaking peptide solutions for 2 hours in a cleavage cocktail comprising (by volume) 95% trifluoroacetic acid (TFA), 2.5% tirisoproylsilane, and 2.5% Milli-Q water. The peptide was then precipitated by adding the cleavage cocktail and cleaved peptide to diethyl ether, and the mixture was then centrifuged and the supernatant discarded. The process of suspending in diethyl ether, centrifuging, and discarding the supernatant was repeated a total of three times. The resulting peptide was then dissolved in water and lyophilized.

Peptide purification: Purification was performed via reverse-phase HPLC using a BEH130 Prep C18 10 μm column (XBridge, Waters Corporation, Milford, MA). Crude peptides were dissolved in Milli-Q water containing 0.1%-vol TFA, and were filtered (0.20 μm filter, Corning, Inc., Corning, NY) before HPLC injection. Products were subjected to an elution gradient (Quaternary Gradient Module (Waters 2545), Waters Corporation) of 100% solvent A (0.1%-vol TFA) to 30% solvent A within 60 min; the composition of solvent B was acetonitrile with 0.1%-vol TFA. Fractions were detected using UV-Vis detection at 214nm (Waters 2489, Waters Corporation) and collected (Waters Fraction Collector III, Waters Corporation). The collected fractions were examined by ESI-mass spectrometry (LCQ Advantage Mass Spectrometer System, Thermo Finnigan, San Jose, CA) with an auto sampler system (Surveyor Autosampler, Thermo Finnigan). Pure fractions were combined and lyophilized.

Analytical HPLC: Purity analysis was performed via reverse-phase Analytical scale HPLC using a BEH C4 3.5 μm column (XBridge, Waters Corporation, Milford, MA). Peptides were dissolved in 80/20-vol Milli-Q water/ACN containing 0.1%-vol TFA, and filtered with 0.20 μm filter (Corning, Inc., Corning, NY) before injection. _Products were subjected to a linear elution gradient (Waters 600 Controller, Waters Corporation, Milford, MA) of 80% solvent A (Milli-Q water with 0.1%-vol TFA) to 5% solvent A in 70 min; the composition of solvent B was acetonitrile with 0.1%-vol TFA. Fractions were detected using photodiode array detector (Waters 2996, Waters Corporation, Milford, MA) tuned at 214 nm.

Circular dichroic spectroscopy (CD): Secondary structures and their temperature-dependent behavior of the synthetic peptides were analyzed using a CD spectropolarimeter (J-820, JASCO Corporation). Sample solutions were prepared at 0.1 mM concentration in pH 7 10mM phosphate buffer and transferred into an absorption cuvette with 1mm path length (110-QS, Hellma, Inc.). Pure pH 7 10mM phosphate buffer solutions were used for the background correction. Sample spectra were recorded from 190-250 nm at desired temperatures. Data points for the wavelength-dependent CD spectra were recorded at every nanometer with a 1nm bandwidth and a 4-second response time for each data point. The CD data was converted to mean residue ellipticity, [MRE] (deg cm^2 $dmol^{-1}$) using the formula: MRE in deg cm^2 $dmol^{-1}$ = mdegrees / (pathlength in mm × the molar concentration of peptide × the number of residues).The values of MRE at 222 nm were used to monitor the temperature-dependent unfolding process of peptides. Data points for the kinetics scans were recorded at 222 nm at 1-minute intervals.

Transmission electron microscopy (TEM): The carbon-coated 200 mesh copper grids (CF200-Cu, Electron Microscopy Sciences, Inc.) were freshly treated by glow discharge using a plasma cleaner (PDC-32G, Harrica Plasma, Inc.) before sample preparation. 5 μL of sample suspension was applied onto the grids. After ~5 min, any remaining liquid was wicked off. Then 5 μL Milli-Q water was applied to the grids and wicked. The grids were incubated under ambient condition for another 30 min before TEM observation (Tecnai 12, FEI or TALOS, FEI) or staining. To negatively stain the grids, 5 μL uranyl acetate aqueous solution (1% wt.) or phosphotungstic acid aqueous solution (2% wt.) was applied to the cast-film grids and retained for 20-30 seconds, then blotted with filter paper. The stained grids were left for at least 10 min before TEM observation.

Small angle x-ray scattering: SAXS measurement was conducted using synchrotron radiation source on beamline of Bio-CAT at APS Argonne National Laboratory. The solution was either loaded in capillary tube for stationary measurement or loaded into flow cell for flowing solution measurement. In stationary measurement, X-ray beam was aligned to shoot through the precipitation. The use of flow cell reduced the radiation damage, but has no effect on diffraction peaks position. The experimental settings in Bio-CAT are as follows: The storage energy was 12 keV, generating 1.033 Å wavelength X-ray. The calibrated sample distance was 3.5 m. Scattering pattern was collected using Pilatus 3 1M photon counter detector with 172x172 micron pixel size, conducted under multiframe mode with 100 or 200 ms exposure time and 5 ms readout time. Data was acquired using epics and initially viewed using ImageJ. Blank buffer scattering data was used for the background correction.

Crystallization: P422_1 crystals were prepared using hanging-drop vapor diffusion by combining 4 μL of protein solution [5 mg/mL P422_1, 10 mM sodium acetate (pH 4.5)] and 4 μL of precipitant solution [0.1 M HEPES (pH 7.0) and 30% (v/v) Jeffamine M-600 (pH 7.0)] on a 22mm square Hampton Research OptiClear cover slip. Solutions of 1.0 M sodium acetate trihydrate, 1.0 M N-(2-hydroxyethyl)piperazine-N′-(2-ethansulfonic acid)

(HEPES), and 50% (v/v) O-(2-Aminopropyl)-O'-(2-methoxyethyl)polypropylene glycol 500 (Jeffamine M-600) were obtained from Hampton Research. The drop was equilibrated against 100 μL of precipitant solution at 20 °C. Crystals formed in a single day and grew to full size after seven days. Crystals were flash cooled in liquid nitrogen with cryo protection provided by the mother liquor.

X-ray crystal structure determination: X-ray diffraction data from P422_1 crystals was collected at Stanford Synchrotron Radiation Lightsource (SSRL) on Beamline 14-1. Indexing, integrating, and scaling of diffraction data were performed using HKL-2000(*205*). See Table 1 for data collection statistics. P422_1 crystalized in the C121 space group and contained three peptide tetramers in the asymmetric unit (ASU). Molecular replacement was performed using PHASER(*206*) from the PHENIX software package(*207*). Molecular replacement rotation and translation functions were carried out using atomic coordinates from a truncated P422_1 design model, where all residues except Gly, His, Ile, Val, Trp(*208*) were changed to Ala, and the sequence shortened by two n-terminal and c-terminal residues. PHENIX(*207*) and COOT(*209*) were used to carry out sequential rounds of refinement and model building. Water molecules were added to the structure in the concluding steps of the refinement.

Table 4.1	
Values in parentheses are for the highest resolution shell	
PDB Entry	Name when listed
Wavelength (Å)	1.18076
Resolution limits (Å)	50.0-1.90
Total reflections measured	94489
Unique reflections measured	24843

Space group symmetry	C121
Unit cell dimensions: a, b, c (Å)	78.095, 72.116, 65.481
Unit cell angles: α, β, γ (degrees)	90.000, 116.188, 90.000
V_m ($Å^3$/Da) / Solvent Content (%)	1.93 / 36.27
$R_{p.i.m.}$ [a]	0.074 (0.804)
Redundancy	3.8 (3.6)
$CC_{1/2}$ [b]	Need overall (0.386)
I/σ(I)	36.458 (1.428)
Completeness (%)	96.6 (95.5)
Overall B factor from Wilson plot ($Å^2$)	37.74
Refinement	
No. of reflections	
Refinement	24826
Test set	1213
Rwork (%) [c]	17.08 (31.69)
Rfree (%) [c]	21.33 (38.71)
No. of non-H atoms per ASU	
Protein	2878
Solvent	21
R.m.s. deviations	
Bonds (Å)	0.012
Angles (°)	1.071
Average B factors ($Å^2$)	
Protein	56.05
Solvent	55.66
Ramachandran	
Favored (%)	100
Outliers (%)	0

[a] $R_{p.i.m.} = \sum[1/(n-1)]^{1/2}|I_h - \langle I_h\rangle|/\sum\langle I_h\rangle$; n is the number of observations (redundancy).

[b] $CC_{1/2} = \sigma_\tau^2/(\sigma_\tau^2 + \sigma_\varepsilon^2)$, where σ_τ^2 is the true measurement of error variance and σ_ε^2 is the independent measurement of error variance.

[c] $R_{work} = \sum // F_0| - |F_c // /\sum|F_0|$ for reflections contained in the working set. $R_{free} = \sum // F_0| - |F_c // /\sum|F_0|$ for reflections contained in the test set which are withheld during refinement (5% of total). $|F_0|$ and $|F_c|$ are the observed and calculated structure factor amplitudes.

4.4 Results

4.4.1 Computational design of P422_1 Peptide

P422_1 is a 29-residue peptide with sequence (motif residues in bold, described below): DQE**I**RQ**MA**EW**I**KK**MA**QM**I**DK**MA**HR**I**DRE**A**-NH$_2$. In the design of the oligomer of P422_1, four separate copies of this peptide compose an antiparallel homotetramer helical bundle of D$_2$ point group symmetry (Figure 4.1A and 4.1B). The computational design of the structure and sequence of P422_1 is described in detail in section 3.3.1 and briefly summarized here. We used a coiled-coil mathematical model to construct the backbone coordinates of the antiparallel helical bundle(*78*, *79*). The geometrical parameters of the bundle (super-helical phase, radius, offset, pitch and minor-helical phase) (*210*) were sampled using Monte Carlo simulated annealing to identify structures and compatible sequences of the 11 hydrophobic amino acids interior to the bundle; sequence-structure combinations (bolded in the sequence above) were identified at local minima of an average energy over sequences(*44*, *85*). The objective function of the Monte Carlo search is an average energy <E> calculated from statistical-mechanical molecular design theory(*44*, *85*, *145*, *211*). The design of the backbone structure and the interior sequence yields a helical bundle motif. Further design of the bundle exterior residues in a lattice configuration used a lattice symmetry approximation(*80*) and was performed by generating lattice configurations of incrementing unit-cell dimension a in a single layer of the P422 space group, where the point symmetry of the helical bundle was aligned with that of the lattice (Figure 4.1C). A statistical design theory was applied to evaluate <E> of each lattice configuration(*44*, *85*).

We selected the most probable sequence at the energy minimum as P422_1 for further experimental synthesis and characterizations.

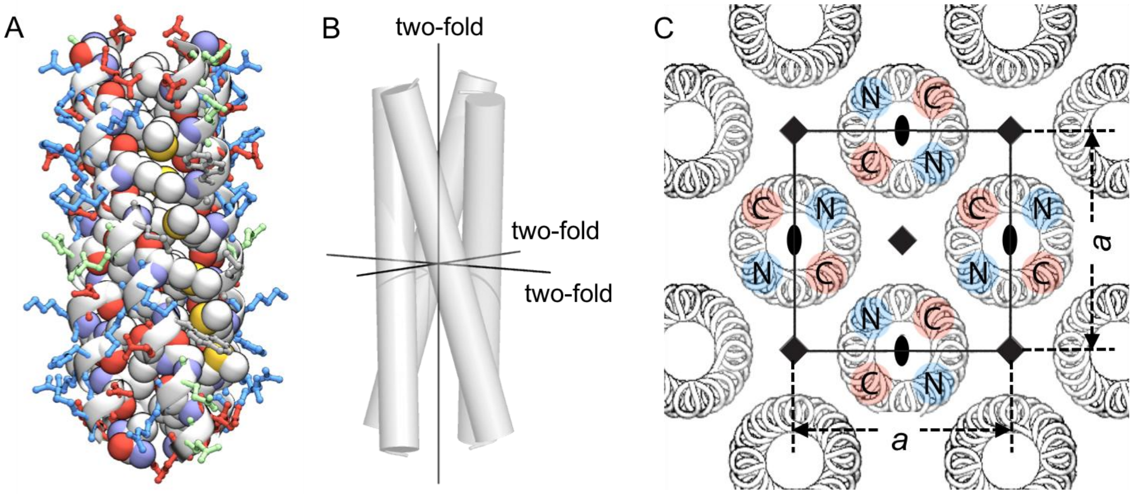


Figure 4.1 The designed model structure of P422_1. A) A single helical bundle, where the motif residues are rendered in spheres with carbon in white, nitrogen in violet, oxygen in red, and sulfur in yellow. The exterior residues are rendered in sticks and colored according to their chemical properties: positively charged KHR (blue), negatively charged DE (red), polar Q (green), hydrophobic WM (grey). **B)** Four helices in the bundle as rods (white) with three two-fold symmetry axes in the D_2 point group. **C)** 2D lattice where the bundles in A) assemble with P422 space group symmetry. The peptides in the 2D lattice model viewed from top are shown as helical tubes. The unit cell is positioned on the lattice (black box) with dimensionality labeled by a. The four-fold and two-fold symmetry axes are represented by square and oval shapes. Individual helical bundles are designed to be antiparallel with the relative position of N- and C-terminus labeled.

4.4.2 Solution phase assembly, crystallization and molecular structure

We have reported in Chapter 3 the successful solution phase assembly of P422_1 into square lattices of nanosheet morphology by TEM. Fast Fourier Transform analysis of the TEM images further indicates the expected four-fold like symmetry in the lattice (Section 3.3.2). However, the spacing observed in the TEM images is on average 4 nm compared to 3.1 nm in the designed lattice configuration. Additionally, Small Angle X-ray Scattering (SAXS) of the solution assembly reveals a series of sharp diffraction peaks and confirms the crystalline nature of the peptide assemblies (Figure 4.7). In order to understand the similarities and differences between the solution assemblies and the designed lattices on the molecular level, we conducted crystallography experiments and solved the structure of P422_1 by molecular replacement (Table 4.1). P422_1 is very prone to crystallization and relatively condition insensitive. We got single crystals in 6 different solution conditions of dimensions up to 180 μm (Figure 4.2). This is in agreement with the designed robustness of the assembly structure. P422_1 crystalized in the monoclinic C2 space group rather than intended tetragonal P422 space group, and the unit cell had dimensions of a=78.1 Å, b= 72.1 Å, c= 65.5Å, α=90.0°, β=116.2°, **γ**=90.0° (Figure 2A) rather than a=31.2Å, b=31.2Å, c=N/A, α=90.0°, β=90.0°, **γ**=90.0° from the design. The asymmetric unit (ASU) of the P422_1 crystal structure contains twelve helices assembled into a trimer of antiparallel homotertramers. The antiparallel homotetramers (helical bundles) are as predicted by the computational design described above. However, instead of the parallel alignment of the superhelical axes of the helical bundles as designed (Figure 4.1C), there is a twisting in the parallel packing between neighboring bundles in the ASU of the crystal structure (Figure 4.3A). Such twist is extended via a 2-fold rotational crystallographic symmetry to the neighboring ASU in the crystal, forming a short twisted ribbon of helical bundles (hexamer of tetramers) (Figure 4.3B, upperleft). The twisted ribbons further assemble through "knobs-into-holes" packing where the end of a ribbon sticks into the major groove of its neighboring copy, thus forming the C2 crystal (Figure 4.3B).

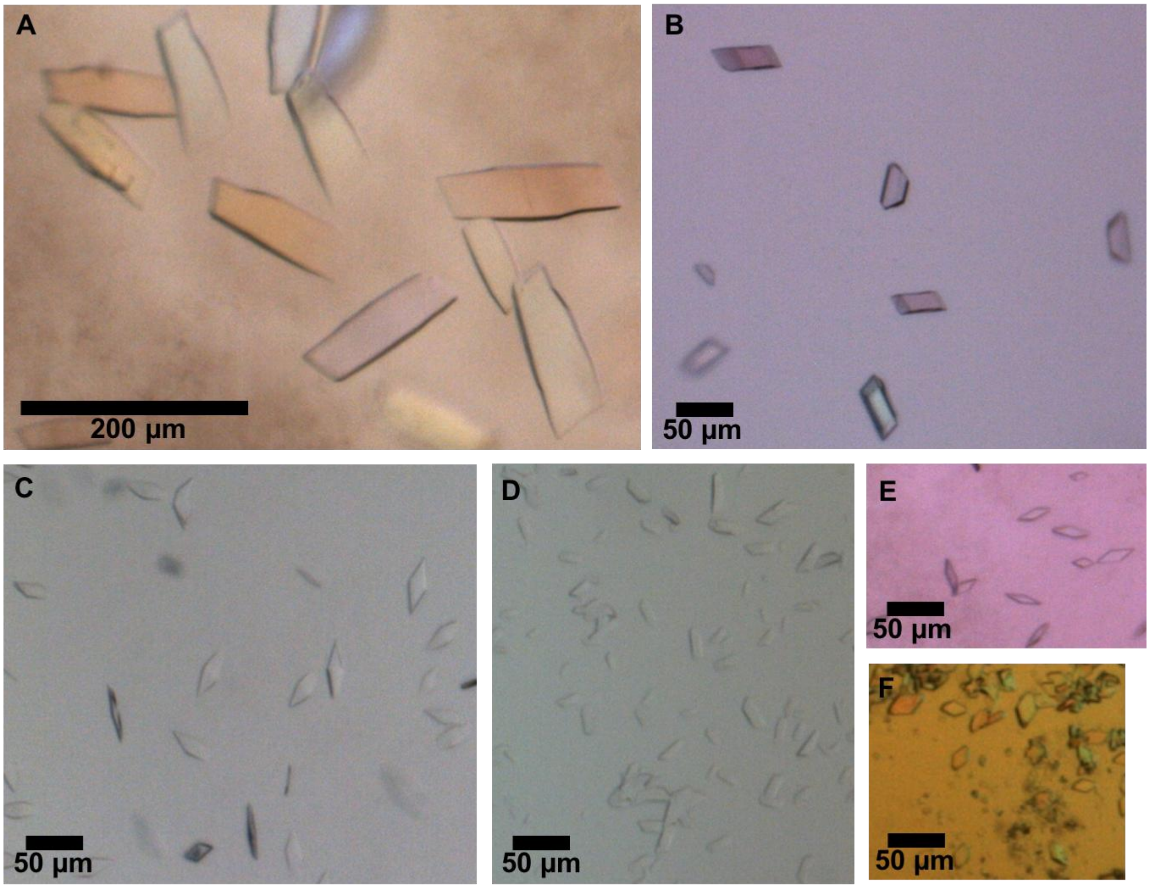


Figure 4.2 P422_1 Single Crystals A-E). P422_1 peptide in 10 mM NaAcetate pH 4.5, 100 µL reservoir volume, grown at 20 °C for 7 days. **A)** 6 µL drop of 5 mg/mL peptide. 0.1 M HEPES pH 7.0, 30% v/v Jeffamine M-600 reservoir. **B)** 1 µL drop of 3 mg/mL peptide. 0.1 M HEPES pH 7.5, 70% v/v (+/-) 2-Methyl-2,4-pentanediol reservoir. **C)** 1 µL drop of 3 mg/mL peptide. 0.05 M Cesium chloride, 0.1 M MES monohydrate pH 6.5, 30% v/v Jeffamine M-600 reservoir. **D)** 1 µL drop of 3 mg/mL peptide. 0.075 M Tris pH 8.5, 18.75% v/v tert-butanol, 25% v/v glycerol reservoir. **E)** 1 µL drop of 3 mg/mL peptide. 0.2 M Potassium chloride, 0.05 M HEPES pH 7.5, 35% v/v pentaerythritol propoxylate (5/4 PO/OH) reservoir. **F)** 1 µL drop of 3 mg/mL P422_1 peptide in 10 mM phosphate buffer pH 7.5 suspended over a 500 µL reservoir of 20 mM NaCl, grown at 50 °C for 12 days.

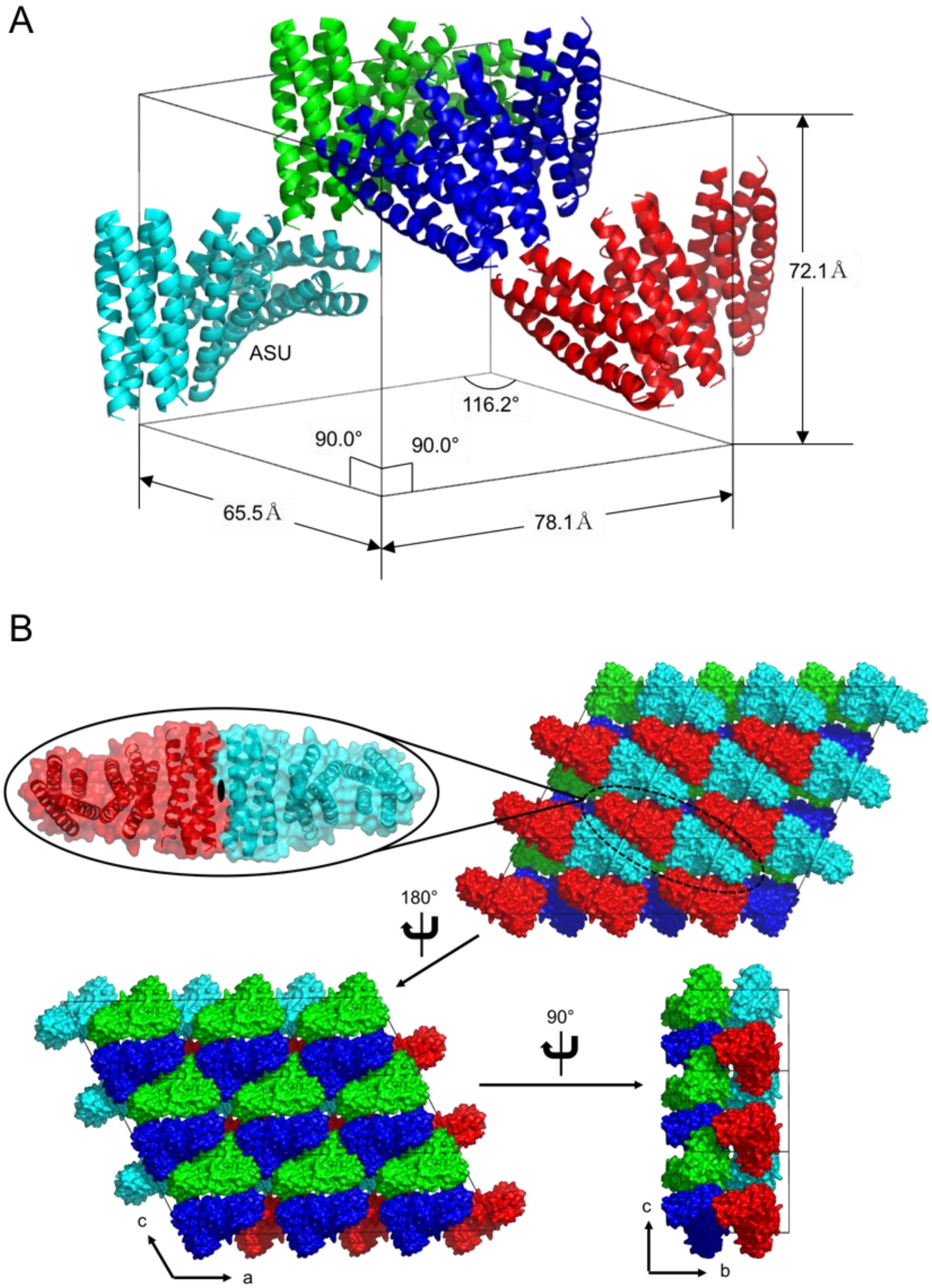


Figure 4.3 The analysis of P422_1 crystal structure. A) Four ASUs, colored differently, pack within a unit cell (black box). The unit cell dimensions are labeled. **B)** The packing alignment of ASUs in the crystal from different perspectives. A 3×1×3 block of unit cells is show. ASUs take the same color scheme as in A). In the inset (upper left), two neighboring ASUs pack to form a twisted ribbon. The oval shape depicts the location of a 2-fold rotational symmetry axis.

4.4.3 Comparison between the crystal structure and the design

4.4.3.1 Helical bundle structure

The successful design of the helical bundle motif is evident in the crystal structure. Figure 4.4A shows the superposition of each helix, twelve in total, in the ASU of the crystal structure to the designed helix. The average RMSD across twelve helices is 0.5Å over all Cα atoms and 1.0Å over all backbone atoms compared to the design (Figure 4.4B top). Furthermore, the superposition of each of the three unique helical bundles in the ASU of the crystal structure to the designed bundle yields an average RMSD of 0.8 Å over all Cα atoms and 1.2 Å over all backbone atoms (Figure 4.4B middle). Also worth notice is the atomistic agreement between the conformations of the motif residues (Ile, Met, Ala) in the crystal structure and in the design, with an RMSD of 1.3 Å over all sidechain heavy-atoms in the helical bundle superpositions (Figure 4.4B bottom).

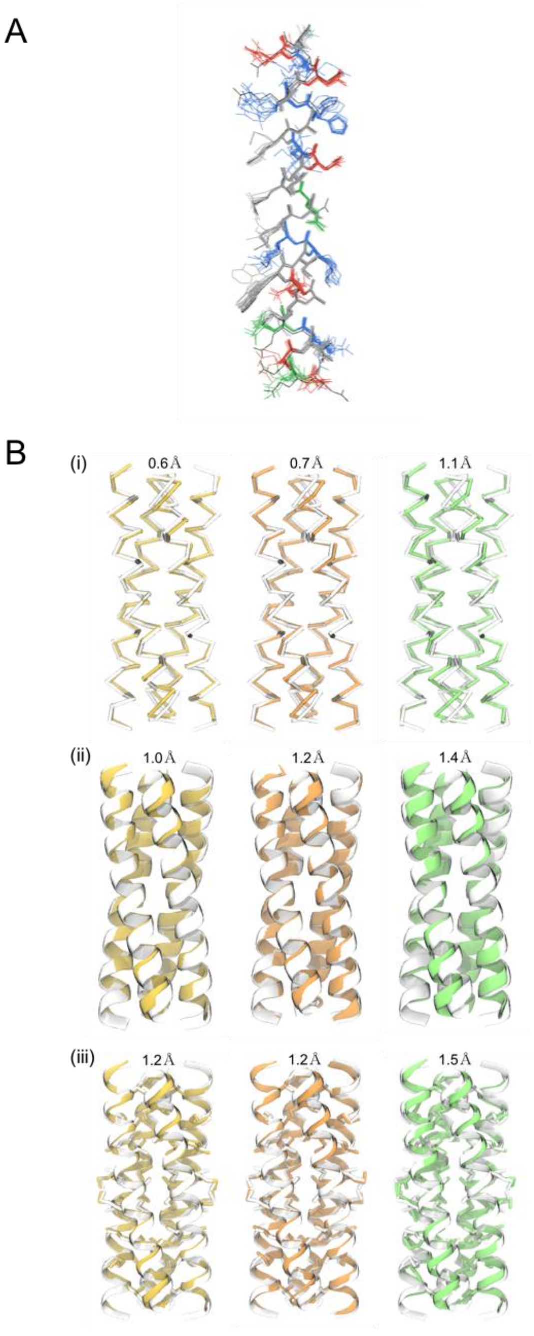


Figure 4.4 Comparison of the helical bundle between the crystal structure and the design. A) The superposition of twelve helices in the ASU of the crystal structure (amino acids colored according to their chemical properties as in Figure 4.1) and the designed peptide (single helix, black). **B)** Three helical bundles in the ASU, colored as the same scheme in Figure 4.5A top, are aligned to the design (white) respectively. (i) superposition of the Cα atom trace; (ii) superposition of all backbone atoms (N, C, Cα, O) rendered in cartoon; (iii) superposition of the hydrophobic motif sidechain conformations (Ile, Ala, Met). RMSDs of the alignment are shown for each superposition.

4.4.3.2 Interfaces between two neighboring bundles

Within each twisted ribbon constituting six helical bundles, there are three unique bundle-bundle interfaces related to each other by non-crystallographic pseudo-symmetry (Figure 4.5A). The interfaces between helical bundles are composed of mostly hydrophobic Met and Trp residues (Figure 4.5 and 4.6B) and three to four hydrogen bonds between Lys13 and Gln16 on neighboring antiparallel helices per interface (Figure 4.5C). The mutation of Gln16 to Leu16 has failed to produce the peptide lattice assemblies under the same solution condition and proves the critical role of this hydrogen bond to the lattice formation (data not shown). A repeating packing motif of Met-Trp-Met interactions is found between Met17-Trp10-Met21 of two antiparallel helices at each interface (Figure 4.5B). Met17 and Met21 locate on one helix and Trp10 on the other antiparallel to the former. There are four such motifs per bundle-bundle interface and the combination can greatly stabilize the bundle-bundle associations. A deeper examination of the motifs shows a distance of 4.9 ± 0.2Å between S in Met21 and the center of the benzene ring in Trp10 and an angle of 13.5 ± 4.9° between the S-aromatic and the ring normal vector. Such interaction geometry is among the most frequently observed Met-aromatic interactions in the PDB database(*212*). Additionally, the distance between S in Met17 and methylene CD1 in Trp10 is 3.9 ± 0.2Å as very close van der Waals contacts.

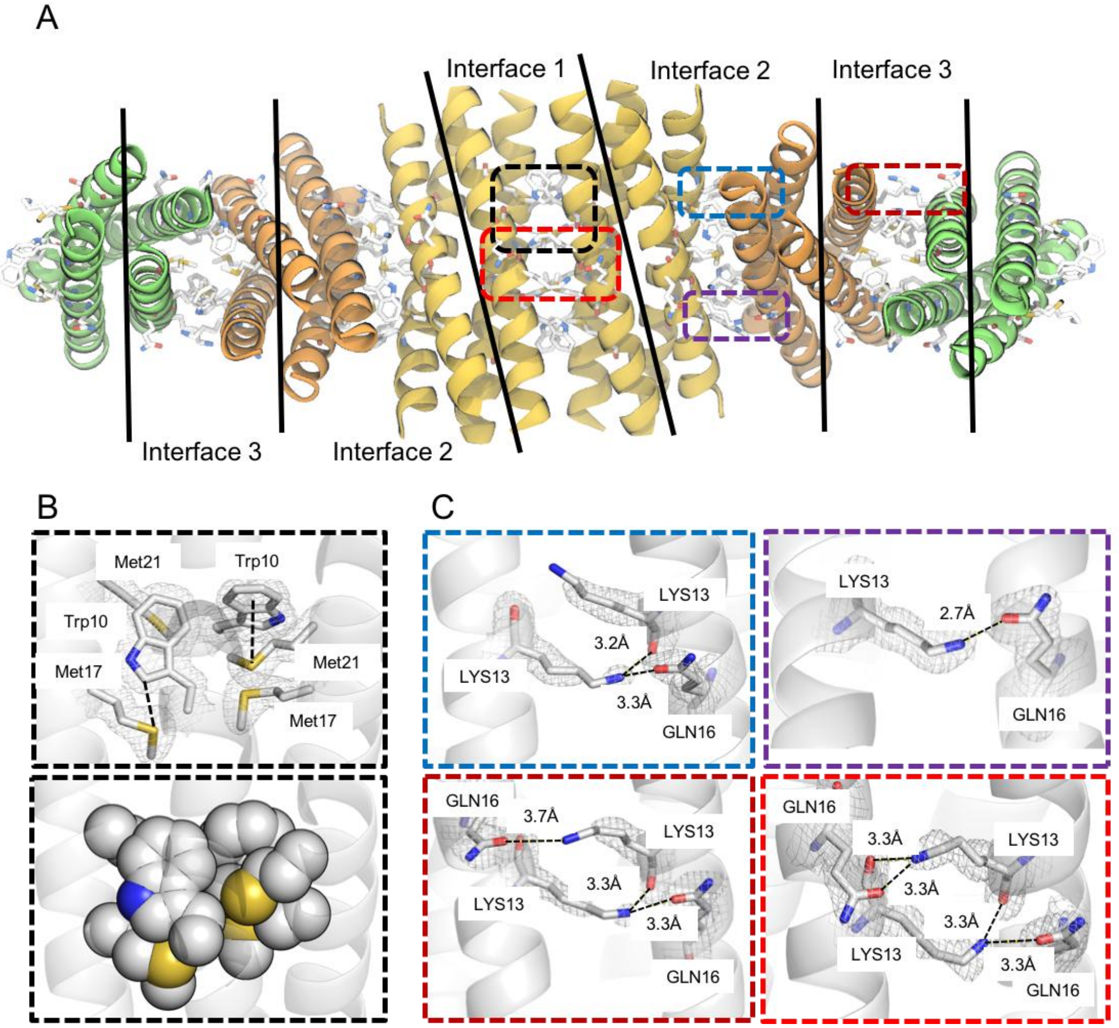


Figure 4.5 The interfaces between neighboring bundles in the crystal. A) The helical bundle-bundle interfaces are depicted in the context of the twisted ribbon (composed of two ASUs). The full-backbone is shown and colored yellow, orange, and lime respectively for each non-crystallographic symmetry related helical bundle. Within each bundle-bundle interface, sidechains of critical stabilizing interactions are rendered as sticks and colored by atom type as in Figure 4.1A. Colored dash boxes correspond to the enlarged and reoriented regions of the interfaces (B, C). **B)** Two example Met-Trp-Met motifs at the bundle-bundle interface are shown in sticks with electron density map ($2F_o{-}F_c$, contoured at 1.0σ). Dashed black lines between Trp10 and Met21, and Trp10 and Met17 label two consistent distances across twelve motifs in three bundle-bundle interfaces. The packing of the motifs (spheres) on the top is shown in the bottom box. **C)** The hydrogen bonds between Lys13 and Gln16 are shown (black dashed lines) with distance between donors and acceptors labeled. Outline colors correspond to the boxes on the twisted ribbon structure in A).

Despite the striking agreement of the backbone structure of individual bundles and the hydrophobic residue conformations to the design, the packing between bundles in the crystal structure deviates from the design (Figure 4.6A). Although in both cases, each neighboring bundle contributes a pair of antiparallel helices to the interface, the relative orientation of the interfacial helix pair is very different, with strict 4-fold rotational symmetry between pairs in the design (Figure 4.1C) compared to a 2-fold-like rotational symmetry in the crystal structure. As mentioned above, the bundle-bundle interfaces in the crystal structure are populated with specific hydrophobic interactions (Figure 4.6B). In comparison, the designed interfaces contain only two non-specific hydrophobic Met17-Met17 and Trp10-Trp10 interactions not packed with optimal geometry; rather, the stabilizing interactions are dominant by electrostatic interactions including eight hydrogen bonds and four additional salt bridges between sidechains of polar residues, such as Arg, Glu and Gln, revealed by PDBePISA(*193*, *213*) interface analysis (Figure 4.6C). These polar residues are mostly buried (buried surface area > 50%) at the bundle-bundle interfaces in the design. Additionally, interfacial Trp10 is partially solvent accessible in the designed lattice whereas it is about 3.6-fold more buried in the crystal structure when comparing the buried surface area. As a result, the calculated solvation energy gain by PDBePISA on interface formation doubles in the crystal structure compared to the designed lattice.

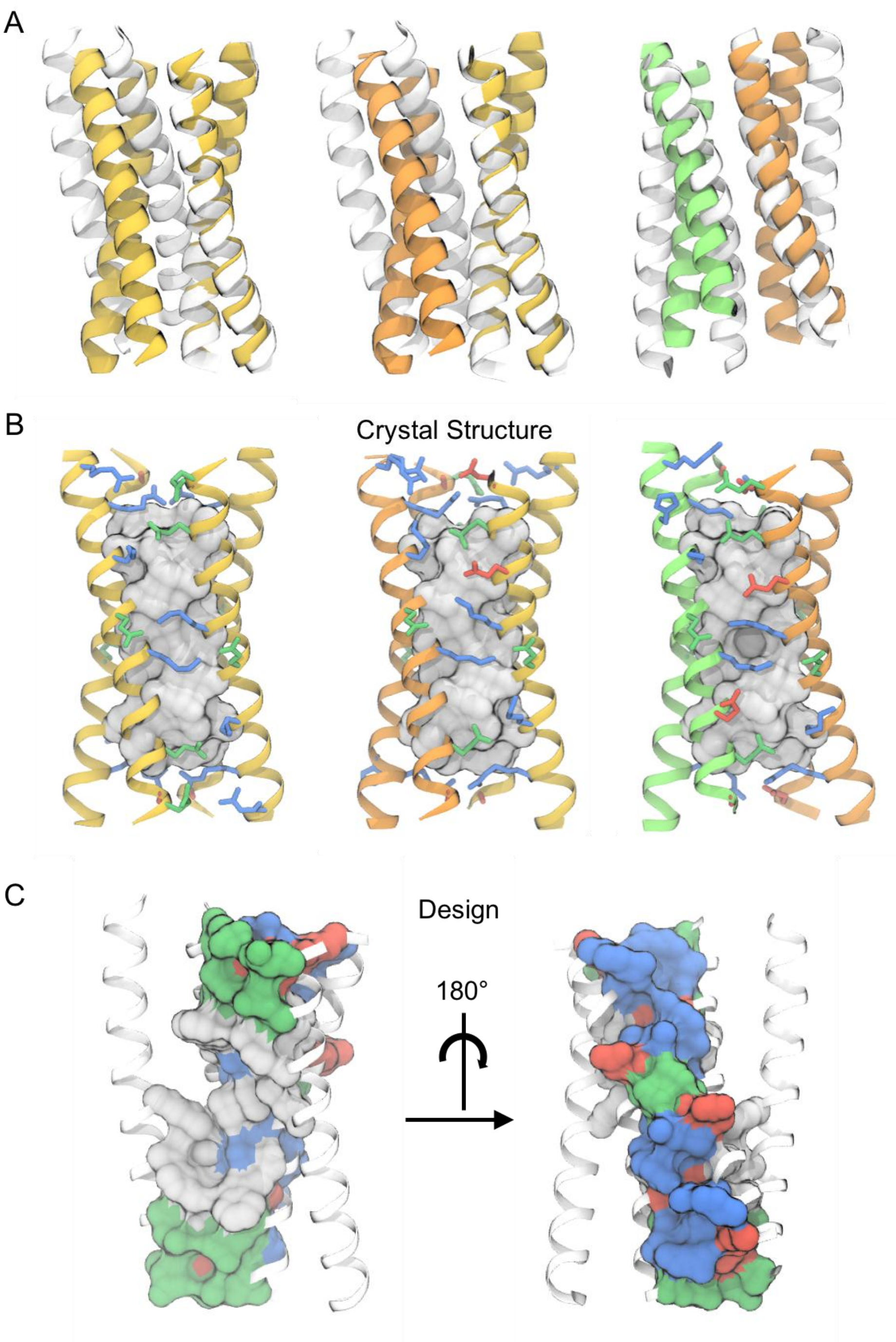


Figure 4.6 Comparison of the interfaces between the crystal structure and the design. A) Intended heavy-atom alignment between bundle-bundle interfaces in the crystal structure (same color scheme as in Figure 4.5A) and the design (white), rendered as cartoons. **B)** Hydrophobic interaction surfaces (grey) in the crystal-structure interfaces; surrounding interfacial polar and charged residues are shown as sticks and colored as in Figure 4.4A**. C)** Interaction surfaces in the designed interfaces (same color scheme in B).

4.4.4 Consistent lattice structure between the crystal and the solution assembly

Since the solution conditions for crystallization are different from the solution assembly conditions reported previously in Chapter 3, it is essential to establish a strong connection between the solved crystal structure lattice and the lattice structure observed by TEM and SAXS. Figure 4.7 shows the superposition between the SAXS pattern simulated from the crystal structure by Mercury(*214*) and the experimental SAXS pattern of the P422_1 solution assembly. The patterns agree strikingly well, where the largest difference in d spacing is 1.3Å between the solution assembly and the crystal structure. The slightly larger spacing overall in the solution assembly could be a result of hydration when the assemblies freely float in the buffer. Additionally, we observed the periodic surface holes of four-fold-like symmetry in the crystal structure that appeared in the high magnification TEM image previously reported (Figure 4.8). Lastly, we performed crystallization using the same condition as the solution assembly and obtained smaller crystals (Figure 4.2F) that were indexed to the same C2 space group with very similar crystal parameters to the solved structure. The indexed dimensions of the smaller crystal are a=80.7Å, b= 72.7 Å, c= 66.7Å, α=90.0°, β=118.0°, **γ**=90.0°. The evidence indicates a consistent lattice structure shared between the solution assembly and the crystal.

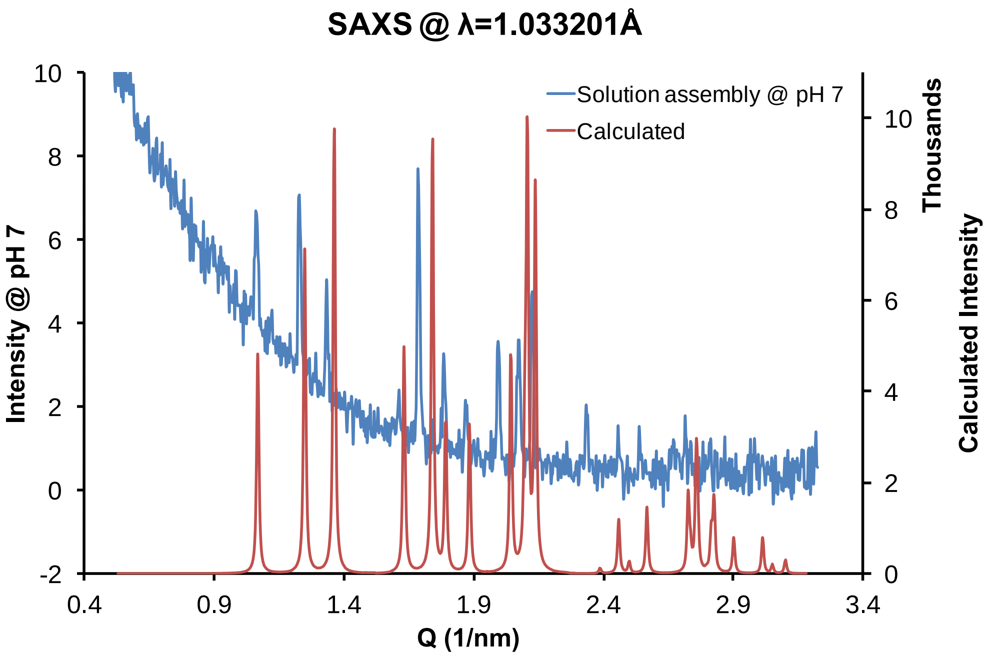


Figure 4.7: The superposition of the measured (blue) and the calculated (red) SAXS data of P422_1. The crystal structure was used as input model for the calculation of the corresponding SAXS pattern.

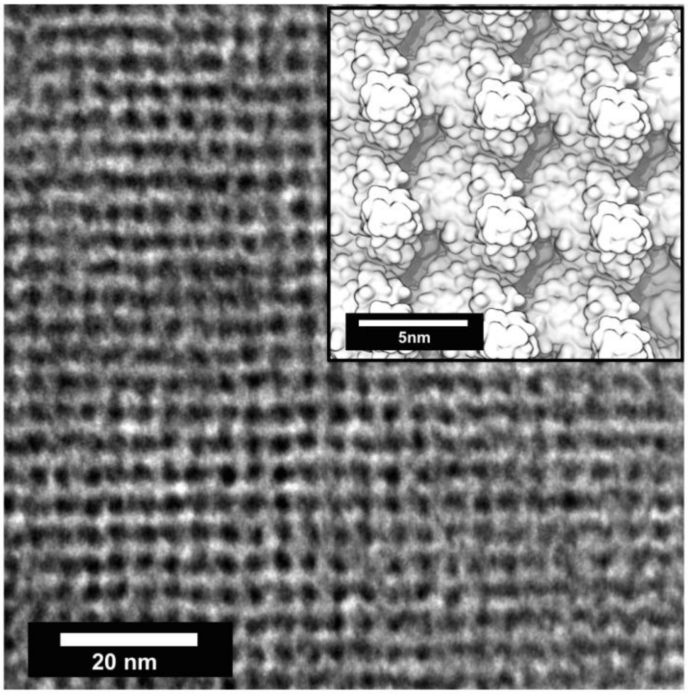


Figure 4.8: The four-fold-like symmetry related surface holes on the ab plane of the crystal (inset, grey shadows in the surface rendering) that is consistent with the TEM image of the solution assembly nanosheet.

4.4.5 Molecular modeling consistent with the crystal structure

To computationally characterize the energetics of the crystal structure in comparison with the designed lattice, we calculated the average conformational energy $\langle E\rangle_{conf}$ of the P422_1 helical bundle homotetramer using the same statistical mechanical design theory used to identify the sequence of P422_1. The calculation using P422_1 sequence yields $\langle E\rangle_{conf,lattice}^{C2}$ for the crystalized C2 lattice configuration, $\langle E\rangle_{conf,lattice}^{P422}$ for the designed single-layer P422 lattice configuration and $\langle E\rangle_{conf,bundle}$ for an isolated bundle of D_2 point symmetry respectively. Since there are three non-equivalent helical bundles in the C2 lattice that are not related by crystallographic symmetry, $\langle E\rangle_{conf,lattice}^{C2}$ is calculated as an average over the three bundles in the ASU. We define the lattice association energy as,

$$\langle E\rangle_a = \langle E\rangle_{conf,lattice} - \langle E\rangle_{conf,bundle}$$

Theoretically, a more negative $\langle E\rangle_a$ indicates a greater energetic gain upon lattice formation from isolated helical bundles. Interestingly, we find that $\langle E\rangle_a^{P422} = -226.67\, kcal/mol$ is significantly more negative than $\langle E\rangle_a^{C2} = -105.28\, kcal/mol$. The inconsistency in the energetic preference of lattice formation between the theory (C2) and the experiment (P422) suggests possible corrections to the energetic terms used in the computational design, such as the use of a linear, distance dependent dielectric constant to attenuate the Coulomb potentials between charges as a simplified solvent model. We will have more detailed discussion on this aspect below.

4.5 Discussion

4.5.1 The successful design of the helical bundle motif

The crystal structure of P422_1 demonstrates the successful design of the helical bundle motif with precision to the atomistic level. Although the antiparallel homotetramer of P422_1 follows the rule of coiled-coil packing, it is a completely novel peptide designed computationally that has no putative conserved domains found in the NCBI’s non-redundant protein database. The helical bundle motif has been proved previously a robust and versatile building block for four distinct peptide nanostructures, one non-assembling coiled-coil and three lattices, and can tolerate covalent modifications of additional two, four and six glycine residues respectively at the N-terminus without losing the lattice features (section 3.3.3). There is enormous potential of such motif to be universally applied to create new peptide-based biomaterials with defined nanostructures. However, we found that the successful prediction of the accurate packing between the helical bundles in a predefined lattice configuration is not a trivial process and might need subtle adjustment of the potentials used in the design.

4.5.2 Hydrophobic Met-Trp-Met motif

One interesting finding in the P422_1 crystal structure is the putative Met-Trp-Met motif prevalent at the bundle-bundle interfaces. The interaction between sulfur in Met and π-electron donors, such as aromatic rings, exists in about one-third of the all known protein structures, and are greatly energetically stabilizing(*212*, *215*). More specifically, the up conformation (methyl group in Met points towards the aromatic ring) found in all the motif copies in the P422_1 structure (Figure 4.5B) is reported to have a calculated interaction energy of -4.5 to -6.6 kcal/mol at a specific interaction distance of ~4.8 Å(*212*). In comparison, the measured and calculated interaction energy for a single hydrogen bond is -0.5 to -1.5 kcal/mol in protein-ligand complexes(*216*, *217*). Although surface salt bridges could stabilize proteins by 0.6 to 1.5 kcal/mol(*218–220*), it is found if buried, their stabilizing effect can be cancelled by the desolvation penalty(*221*). The desolvation penalty is a result of moving two oppositely charged residues exposed in a hydrophilic solution, such as buffer, to a hydrophobic region upon burial of the residues, such as upon protein-protein association to form a tightly-packed crystal. The other half of the Met-Trp-Met motif contains a less well studied sulfur-methylene interaction in the protein environment, yet exists in about 40% of all Met(sulfur) related interactions in protein-ligand complexes(*222*). Although the experimental energetic characterization is not found for such motif, it is known that hydrophobic interactions are less sensitive to the solution environment, such as pH and ionic strength, and Met and Trp residues suffer less desolvation penalty upon burial compared to the polar residues involved in hydrogen bonds and salt bridges. Therefore, the interactions in the Met-Trp-Met motif could be overall more energetically advantageous for helical bundle associations than the designed electrostatic interactions.

4.5.3 Lattice association energy discrepancy between the design and the crystal

We also want to understand the possible sources of energy discrepancy on the lattice association energy between the design and the crystal structure. One explanation for such discrepancy can be favorable desolvation energy for residues involved in the bundle-bundle interfaces in the crystal than in the design. A PDBePISA (*193*) scoring function analysis indicates the association of the bundle-bundle interface is on average 3.2 kcal/mol more favorable in the crystal structure than in the designed P422 lattice. More specifically, the average solvation energy gain upon bundle-bundle association is 15.3 kcal/mol more negative (less desolvation penalty) in the crystal structure. As a result, the large difference in desolvation penalty could play a critical role in offsetting the energetic stabilizing effect of electrostatic interactions, such as hydrogen bonds and salt bridges, in the designed lattice.

4.5.4 Feedback to improve the design program: solvent model for peptide lattice

The electrostatic interactions in the design are attenuated by a linear, distance-dependent dielectric screening. This is a simplified form to model solvent effects(*223*). It is a useful parameterization when combined with environmental energy model for the design of solvated globular proteins(*85*, *86*). However, in the context of a tightly-packed peptide lattice, such calculation of the electrostatic potential could lead to an overestimation of the energy-stabilizing contribution from hydrogen bonds and salt bridges. Therefore, a separate solvation energy model might be needed to take into consideration the desolvation penalty of the buried hydrophilic residues upon protein-protein association in the crystal. Such solvation energy approximation can be achieved by using empirical atomic solvation parameters(*193*) or atomic contact energies(*224*) used in the approximation of electrostatics in the case of designing macromolecule crystals. Alternatively, more rigorous analytical solvent models could be applied to estimate the solvation energy of buried charges and to compensate the energy model used in the design. For example, a mean-field treatment of solvents is applicable by using

Langevin dipoles for water in a grid(*225*) surrounding the protein-protein interface in the lattice approximation. The development and benchmark of adjustments to the calculation of the electrostatic potential and to include possible solvent models will be the focus of future improvements on the statistical-mechanical design theory in light of enhancing the success rate of computational design of peptide lattices.

4.6 Conclusion

We have presented the computational design of a novel and robust crystal-forming peptide P422_1 with solution phase characterizations and its crystal structure. The building block of P422_1 lattice, a helical bundle motif, has atomistic agreement between the design and crystal structure. Such motif is of broad interest for the design of different peptide lattices. The bundle-bundle interfaces in the lattice are stabilized by a Met-Trp-Met motif featuring hydrophobic interactions in a specific geometry, rather than the buried hydrogen bonds and salt bridges between polar residues with higher desolvation penalty that are designed. Such observation reveals the critical role of specific hydrophobic interactions in the protein-protein interfaces in a crystal. It also reveals the necessity of a separate solvent model for the design of peptide lattices other than a simplified representation used to calculate electrostatic energies. This work provides direct evidence of the successful and yet-to-be improved components of the computational design approach, and rare insights into potential ways to enhance the design success rate on a fundamental basis.

CHAPTER 5 Conclusion

5.1 Summary and Outlook

Peptide-based self-assemblies are a new category of nanomaterials that can be engineered with defined molecular structures with atomistic precision, and controllable nanostructures by design and solution assembly conditions. These peptide-based self-assemblies can also serve as hybrid material templates for chemical modifications and inorganic/organic doping with controlled geometry and spacing.

In this thesis, we focused on the design aspects of peptide-based self-assemblies, and the understanding of fundamental molecular principles that guides the assembly interactions. We presented rationally designed peptide self-assembly hydrogels, and computationally designed peptide self-assemblies targeted for distinct nanostructures: a single helical bundle and three nanosheets of different lattice structures. We further characterized the designed peptide self-assemblies by various experimental and modeling techniques to understand their material properties, nanostructures and molecular structures. In the case of the designed peptide hydrogel, we showed that through careful rational, molecular design of the specific hydrophobic interactions in the β-sheet fibrils LNK1, one can control and modulate the assembly network morphology, the underling molecular structures and the rheological properties of the peptide-based hydrogel nanomaterials. Next, we presented the computational design approach to pre-engineer the well-defined nanostructures and morphologies into peptide-based self-assembling nanomaterials. One of the most important discoveries of this research project is that we computationally designed a novel, versatile, α-helical peptide bundle motif, stabilized by shape complementary hydrophobic core, which can be applied as a universal building block to a set of hierarchical nanomaterials solution-assembled with distinct lattice structures. This is realizable through subsequent computational design of motif exterior residues. We reported the successful design of two-dimensional sheet materials based on the motif. More specifically, the motif is designed to be shared among four distinct nanostructures: a non-assembling, isolated helical bundle in solution and three two-dimensional lattices with distinct local structures. The TEM, SAXS and a crystal structure of one of the lattices proved the concept of such common motif for various lattices is successfully explored and captured. However, the subtle packing of the exterior residues of the motifs caused the lattices did not always match the prediction. From the crystal structure of P422_1, we found the crucial role of Met-Trp-Met hydrophobic interactions at peptide interfaces inside the tightly-packed lattice rather than the computationally identified, stabilizing electrostatic interactions. The specific hydrophobic interactions lock in the relative orientations of neighboring helical bundles in the crystal structure deviating from the design. Therefore, moving forward, it is necessary to include a more thorough energetic approximation in the computational design approach to take into account the desolvation effects at tightly packed protein-protein interfaces, in order to improve the design success rate for peptide nanomaterials with predefined lattice structures. Additionally, we are currently making efforts in solving the crystal structures of the rest of proposed peptide self-assembly designs to obtain the full molecular picture of the fundamental rules that determine the outcome of the designs.

We need to understand that the design methods, particularly the computational approach despite derived from first principles, are still approximations and do not guarantee the prespecified molecular- and nano-structures of peptide-based self-assembly nanomaterials. The feedback from the experimental characterization provides critical guidance for the improvement of the computational design methods, in terms of why the structures of the assemblies and the properties of the materials deviate from the predictions. In the meantime, existing peptide designs that are highly ordered can be used for advanced modifications and can evolve into hybrid materials, such as templating for metal nanoparticles, organic cofactors, and crosslinking sites.

Regarding future designs, one direction will target more advanced hybrid materials, to precisely control the display of different components in the assembled peptides, and to incorporate unnatural amino acids into the design methodology. Another direction is to expand the nanostructures and macrostructures of the peptide-based self-assemblies, particularly to include arbitrary symmetry outside of the natural crystal realm and to build superlattices similar to inorganic materials.

REFERENCES

1. T. EC, Ferritin: structure, gene regulation, and cellular function in animals, plants and microorganisms. *Annu Rev Biochem*. **56**, 289–315 (1987).

2. M. A. Edeling, C. Smith, D. Owen, Life of a clathrin coat: insights from clathrin and AP structures. *Nat. Rev. Mol. Cell Biol.* **7**, 32–44 (2006).

3. E. Baranova *et al.*, SbsB structure and lattice reconstruction unveil Ca2+ triggered S-layer assembly. *Nature*. **487**, 119–122 (2012).

4. H. Colognato, P. D. Yurchenco, Form and function: The laminin family of heterotrimers. *Dev. Dyn.* **218**, 213–234 (2000).

5. P. D. Yurchenco, Basement membranes: Cell scaffoldings and signaling platforms. *Cold Spring Harb. Perspect. Biol.* **3**, 1–27 (2011).

6. A. Finnemore *et al.*, Biomimetic layer-by-layer assembly of artificial nacre. *Nat. Commun.* **3**, 966 (2012).

7. R. M. Capito, H. S. Azevedo, Y. S. Velichko, A. Mata, S. I. Stupp, Self-assembly of large and small molecules into hierarchically ordered sacs and membranes. *Science*. **319**, 1812–6 (2008).

8. J.-F. Lutz, M. Ouchi, D. R. Liu, M. Sawamoto, Sequence-Controlled Polymers. *Science (80-.).* **341**, 1238149–1238149 (2013).

9. H.-J. Sun, S. Zhang, V. Percec, From structure to function via complex supramolecular dendrimer systems. *Chem. Soc. Rev.* **44**, 3900–3923 (2015).

10. E. V Shevchenko, D. V Talapin, N. A. Kotov, S. O’Brien, C. B. Murray, Structural diversity in binary nanoparticle superlattices. *Nature*. **439**, 55–59 (2006).

11. D. V Talapin *et al.*, Quasicrystalline order in self-assembled binary nanoparticle superlattices. *Nature*. **461**, 964–967 (2009).

12. A. Dong, J. Chen, P. M. Vora, J. M. Kikkawa, C. B. Murray, Binary nanocrystal superlattice membranes self-assembled at the liquid-air interface. *Nature*. **466**, 474–477 (2010).

13. M. Cargnello *et al.*, Substitutional doping in nanocrystal superlattices. *Nature*. **524**, 450–3 (2015).

14. T. Paik, B. T. Diroll, C. R. Kagan, C. B. Murray, Binary and Ternary Superlattices Self-Assembled from Colloidal Nanodisks and Nanorods. *J. Am. Chem. Soc.* **137**, 6662–6669 (2015).

15. M. A. Zhuravel, N. E. Davis, S. T. Nguyen, I. Koltover, Dendronized protein polymers: Synthesis and self-assembly of monodisperse cylindrical macromolecules. *J. Am. Chem. Soc.* **126**, 9882–9883 (2004).

16. V. Percec *et al.*, Visualizable cylindrical macromolecules with controlled stiffness from backbones containing libraries of self-assembling dendritic side groups. *J. Am. Chem. Soc.* **120**, 8619–8631 (1998).

17. J. G. Rudick, V. Percec, Induced helical backbone conformations of self-organizable dendronized polymers. *Acc. Chem. Res.* **41**, 1641–1652 (2008).

18. H. Qiu, Z. M. Hudson, M. A. Winnik, I. Manners, Multidimensional hierarchical self-assembly of amphiphilic cylindrical block comicelles. *Science (80-.).* **347**, 1329–1332 (2015).

19. K.-V. Peinemann, V. Abetz, P. F. W. Simon, Asymmetric superstructure formed in a block copolymer via phase separation. *Nat. Mater.* **6**, 992–996 (2007).

20. P. A. Rupar, L. Chabanne, M. A. Winnik, I. Manners, Non-centrosymmetric cylindrical micelles by unidirectional growth. *Science*. **337**, 559–62 (2012).

21. H. Cui, Z. Chen, S. Zhong, K. L. Wooley, D. J. Pochan, Block Copolymer Assembly via Kinetic Control. *Science (80-.).* **317**, 647–650 (2007).

22. A. H. Gröschel *et al.*, Precise hierarchical self-assembly of multicompartment micelles. *Nat. Commun.* **3**, 710 (2012).

23. M. R. Jones, N. C. Seeman, C. A. Mirkin, Nanomaterials. Programmable materials and the nature of the DNA bond. *Science*. **347**, 1260901 (2015).

24. B. Wei, M. Dai, P. Yin, Complex shapes self-assembled from single-stranded DNA tiles. *Nature*. **485**, 623–626 (2012).

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

26. J. Zheng *et al.*, From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature*. **461**, 74–77 (2009).

27. M. R. Jones *et al.*, DNA-nanoparticle superlattices formed from anisotropic building blocks. *Nat. Mater.* **9**, 913–917 (2010).

28. R. J. Macfarlane *et al.*, Nanoparticle Superlattice Engineering with DNA. *Science (80-.).* **334**, 204–208 (2011).

29. E. Auyeung *et al.*, DNA-mediated nanoparticle crystallization into Wulff polyhedra. *Nature*. **505**, 73–7 (2014).

30. M. B. Ross, J. C. Ku, V. M. Vaccarezza, G. C. Schatz, C. A. Mirkin, Nanoscale form dictates mesoscale function in plasmonic DNA–nanoparticle superlattices. *Nat Nano*. **10**, 453–458 (2015).

31. L. Haines-Butterick *et al.*, Controlling hydrogelation kinetics by peptide design for three-dimensional encapsulation and injectable delivery of cells. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7791–6 (2007).

32. G. Yang *et al.*, Precise and Reversible Protein-Microtubule-Like Structure with Helicity Driven by Dual Supramolecular Interactions. *J. Am. Chem. Soc.* **138**, 1932–1937 (2016).

33. R. A. Miller, A. D. Presley, M. B. Francis, Self-assembling light-harvesting systems from synthetically modified tobacco mosaic virus coat proteins. *J. Am. Chem. Soc.* **129**, 3104–3109 (2007).

34. D. J. Glover, L. Giger, S. S. Kim, R. R. Naik, D. S. Clark, Geometrical assembly of ultrastable protein templates for nanomaterials. *Nat. Commun.* **7**, 11771 (2016).

35. J. D. Brodin *et al.*, Metal-directed, chemically tunable assembly of one-, two- and three-

dimensional crystalline protein arrays. *Nat. Chem.* **4**, 375–82 (2012).

36. J. C. Sinclair, K. M. Davies, C. Vénien-Bryan, M. E. M. Noble, Generation of protein lattices by fusing proteins with matching rotational symmetry. *Nat. Nanotechnol.* **6**, 558–62 (2011).

37. S. Gonen, F. DiMaio, T. Gonen, D. Baker, Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. *Science (80-.).* **348**, 1365–1368 (2015).

38. Y. Suzuki *et al.*, Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals. *Nature*. **533**, 369–373 (2016).

39. N. P. King *et al.*, Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science*. **336**, 1171–4 (2012).

40. N. P. King *et al.*, Accurate design of co-assembling multi-component protein nanomaterials. *Nature*. **510**, 103–108 (2014).

41. D. J. E. Huard, K. M. Kane, F. A. Tezcan, Re-engineering protein interfaces yields copper-inducible ferritin cage assembly. *Nat. Chem. Biol.* **9**, 169–76 (2013).

42. Y.-T. Lai, K.-L. Tsai, M. R. Sawaya, F. J. Asturias, T. O. Yeates, Structure and Flexibility of Nanoscale Protein Cages Designed by Symmetric Self-Assembly. *J. Am. Chem. Soc.* **135**, 7738–7743 (2013).

43. J. B. Bale *et al.*, Accurate design of megadalton-scale two-component icosahedral protein complexes. *Science (80-.).* **353**, 389–394 (2016).

44. C. J. Lanci *et al.*, Computational design of a protein crystal. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7304–9 (2012).

45. K.-H. Kim *et al.*, Protein-directed self-assembly of a fullerene crystal. *Nat. Commun.* **7**, 11429 (2016).

46. Y.-T. Lai *et al.*, Structure of a designed protein cage that self-assembles into a highly porous cube. *Nat. Chem.* **6**, 1065–1071 (2014).

47. P. C. Jordan *et al.*, Self-assembling biomolecular catalysts for hydrogen production. *Nat. Chem.*, 1–7 (2015).

48. P. a Sontz, W. J. Song, F. A. Tezcan, Interfacial metal coordination in engineered protein and peptide assemblies. *Curr. Opin. Chem. Biol.* **19**, 42–9 (2014).

49. Y. Mou, J.-Y. Yu, T. M. Wannier, C.-L. Guo, S. L. Mayo, Computational design of co-assembling protein–DNA nanowires. *Nature*. **525**, 230–233 (2015).

50. J. D. Brodin, E. Auyeung, C. A. Mirkin, DNA-mediated engineering of multicomponent enzyme crystals. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 4564–9 (2015).

51. T. Sawada, M. Tsuchiya, T. Takahashi, H. Tsutsumi, H. Mihara, Cell-adhesive hydrogels composed of peptide nanofibers responsive to biological ions. *Polym. J.* **44**, 651–657 (2012).

52. M. T. McClendon, S. I. Stupp, Tubular hydrogels of circumferentially aligned nanofibers to encapsulate and orient vascular cells. *Biomaterials*. **33**, 5713–5722 (2012).

53. A. Altunbas, S. J. Lee, S. A. Rajasekaran, J. P. Schneider, D. J. Pochan, Encapsulation of curcumin in self-assembling peptide hydrogels as injectable drug delivery vehicles. *Biomaterials*. **32**, 5906–14 (2011).

54. S. Bulut *et al.*, Slow release and delivery of antisense oligonucleotide drug by self-assembled peptide amphiphile nanofibers. *Biomacromolecules*. **12**, 3007–14 (2011).

55. H. Guo *et al.*, Sustained delivery of VEGF from designer self-assembling peptides improves cardiac function after myocardial infarction. *Biochem. Biophys. Res. Commun.* **424**, 105–11 (2012).

56. A. S. Veiga *et al.*, Arginine-rich self-assembling peptides as potent antibacterial gels. *Biomaterials*. **33**, 8907–8916 (2012).

57. L. Liu *et al.*, Self-assembled cationic peptide nanoparticles as an efficient antimicrobial agent. *Nat. Nanotechnol.* **4**, 457–63 (2009).

58. A. Tanaka *et al.*, Cancer cell death induced by the intracellular self-assembly of an enzyme-responsive supramolecular gelator. *J. Am. Chem. Soc.* **137**, 770–775 (2015).

59. J. Li *et al.*, Enzyme-Instructed Intracellular Molecular Self-Assembly to Boost Activity of Cisplatin against Drug-Resistant Ovarian Cancer Cells. *Angew. Chemie - Int. Ed.* **54**, 13307–13311 (2015).

60. J. Zhou, X. Du, N. Yamagata, B. Xu, Enzyme-Instructed Self-Assembly of Small D-Peptides as a Multiple-Step Process for Selectively Killing Cancer Cells. *J. Am. Chem. Soc.* **138**, 3813–3823 (2016).

61. Y. Maeda *et al.*, Molecular Self-Assembly Strategy for Generating Catalytic Hybrid Polypeptides. *PLoS One*. **11**, e0153700 (2016).

62. B. K. Shanbhag, B. Liu, J. Fu, V. S. Haritos, L. He, Self-Assembled Enzyme Nanoparticles for Carbon Dioxide Capture. *Nano Lett.* **16**, 3379–3384 (2016).

63. H. C. Fry, Y. Liu, N. M. Dimitrijevic, T. Rajh, Photoinitated charge separation in a hybrid titanium dioxide metalloporphyrin peptide material. *Nat. Commun.* **5**, 4606 (2014).

64. Y. Hsia *et al.*, Design of a hyperstable 60-subunit protein icosahedron. *Nature*. **535**, 136–139 (2016).

65. L. Doyle *et al.*, Rational design of α-helical tandem repeat proteins with closed architectures. *Nature*. **528**, 585–588 (2015).

66. P.-S. Huang *et al.*, De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2015).

67. A. R. Thomson *et al.*, Computational design of water-soluble α-helical barrels. *Science (80-.).* **346**, 485–488 (2014).

68. J. M. Fletcher *et al.*, Self-assembling cages from coiled-coil peptide modules. *Science*. **340**, 595–9 (2013).

69. M. Balbirnie, R. Grothe, D. S. Eisenberg, An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated beta-sheet structure for amyloid. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2375–2380 (2001).

70. T. R. Jahn *et al.*, The common architecture of cross-beta amyloid. *J. Mol. Biol.* **395**, 717–27 (2010).

71. M. G. Ryadnov, D. N. Woolfson, Engineering the morphology of a self-assembling protein fibre. *Nat. Mater.* **2**, 329–32 (2003).

72. S. E. Fischer, L. Mi, H.-Q. Mao, J. L. Harden, Biofunctional coatings via targeted covalent cross-linking of associating triblock proteins. *Biomacromolecules*. **10**, 2408–17 (2009).

73. E. Genové, C. Shen, S. Zhang, C. E. Semino, The effect of functionalized self-assembling peptide scaffolds on human aortic endothelial cell function. *Biomaterials*. **26**, 3341–51 (2005).

74. T.-Y. Cheng, M.-H. Chen, W.-H. Chang, M.-Y. Huang, T.-W. Wang, Neural stem cells encapsulated in a functionalized self-assembling peptide hydrogel for brain tissue engineering. *Biomaterials*. **34**, 2005–16 (2013).

75. A. P. Nowak *et al.*, Rapidly recovering hydrogel scaffolds from self-assembling diblock copolypeptide amphiphiles. *Nature*. **417**, 424–8 (2002).

76. C. J. Bowerman, B. L. Nilsson, Self-assembly of amphipathic β-sheet peptides: insights and applications. *Biopolymers*. **98**, 169–84 (2012).

77. J. E. Padilla, C. Colovos, T. O. Yeates, Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 2217–21 (2001).

78. G. Grigoryan, W. F. Degrado, Probing designability via a generalized model of helical bundle geometry. *J. Mol. Biol.* **405**, 1079–100 (2011).

79. J. R. Quine, Helix parameters and protein structure using quaternions. *J. Mol. Struct. THEOCHEM*. **460**, 53–66 (1999).

80. X. Fu, H. Kono, J. G. Saven, Probabilistic approach to the design of symmetric protein quaternary structures. *Protein Eng.* **16**, 971–7 (2003).

81. N. Blomberg, E. Baraldi, M. Nilges, M. Saraste, The PH superfold: a structural scaffold for multiple functions. *Trends Biochem. Sci.* **24**, 441–445 (1999).

82. F. J. Kull, E. P. Sablin, R. Lau, R. J. Fletterick, R. D. Vale, Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature*. **380**, 550–5 (1996).

83. T. Lazaridis, M. Karplus, Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* **10**, 139–145 (2000).

84. J. Mendes, R. Guerois, L. Serrano, Energy estimation in protein design. *Curr. Opin. Struct. Biol.* **12**, 441–446 (2002).

85. H. Kono, J. G. Saven, Statistical theory for protein combinatorial libraries. Packing interactions, backbone flexibility, and the sequence variability of a main-chain structure. *J. Mol. Biol.* **306**, 607–28 (2001).

86. A. M. Slovic, H. Kono, J. D. Lear, J. G. Saven, W. F. DeGrado, Computational design of water-soluble analogues of the potassium channel KcsA. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 1828–33 (2004).

87. K. T. O'Neil, W. F. DeGrado, A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*. **250**, 646–51 (1990).

88. C. Yanover, O. Schueler-Furman, Y. Weiss, Minimizing and learning energy functions for side-chain prediction. *J. Comput. Biol.* **15**, 899–911 (2008).

89. O. Sharabi, C. Yanover, A. Dekel, J. M. Shifman, Optimizing energy functions for protein-protein interface design. *J. Comput. Chem.* **32**, 23–32 (2011).

90. C. S. Poultney *et al.*, Rational Design of Temperature-Sensitive Alleles Using Computational Structure Prediction. *PLoS One*. **6**, e23947 (2011).

91. C.-M. Yu *et al.*, Rationalization and Design of the Complementarity Determining Region Sequences in an Antibody-Antigen Recognition Interface. *PLoS One*. **7**, e33340 (2012).

92. C. King *et al.*, Removing T-cell epitopes with computational protein design. *Proc. Natl. Acad. Sci.* **111**, 8577–8582 (2014).

93. A. D. MacKerell, *et al.*, All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. *J. Phys. Chem. B*. **102**, 3586–3616 (1998).

94. W. D. Cornell *et al.*, A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).

95. J. W. Ponder, D. A. Case, Force Fields for Protein Simulations. *Adv. Protein Chem.* **66**, 27–85 (2003).

96. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **21**, 1087 (1953).

97. S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, Optimization by simulated annealing. *Science*. **220**, 671–80 (1983).

98. J. R. Desjarlais, T. M. Handel, De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–18 (1995).

99. J. A. Fallas, J. D. Hartgerink, Computational design of self-assembling register-specific collagen heterotrimers. *Nat. Commun.* **3**, 1087 (2012).

100. J. Desmet, M. De Maeyer, B. Hazes, I. Lasters, The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*. **356**, 539–542 (1992).

101. R. F. Goldstein, Efficient rotamer elimination applied to protein side-chains and related spin glasses. *Biophys. J.* **66**, 1335–1340 (1994).

102. L. L. Looger, H. W. Hellinga, Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J. Mol. Biol.* **307**, 429–45 (2001).

103. P. Biswas, J. Zou, J. G. Saven, Statistical theory for protein ensembles with designed energy landscapes. *J. Chem. Phys.* **123**, 154908 (2005).

104. J. G. Saven, Computational protein design: Advances in the design and redesign of biomolecular nanostructures. *Curr. Opin. Colloid Interface Sci.* **15**, 13–17 (2010).

105. J. P. Schneider *et al.*, Responsive hydrogels from the intramolecular folding and self-

assembly of a designed peptide. *J. Am. Chem. Soc.* **124**, 15030–7 (2002).

106. D. Pochan, J. Schneider, Responsive materials via designed peptides. *Biopolymers*. **71**, 300–301 (2003).

107. D. J. Pochan *et al.*, Thermally reversible hydrogels via intramolecular folding and consequent self-assembly of a de novo designed peptide. *J. Am. Chem. Soc.* **125**, 11802–3 (2003).

108. B. Ozbas, J. Kretsinger, K. Rajagopal, J. P. Schneider, D. J. Pochan, Salt-Triggered Peptide Folding and Consequent Self-Assembly into Hydrogels with Tunable Modulus. *Macromolecules*. **37**, 7331–7337 (2004).

109. B. Ozbas, K. Rajagopal, J. Schneider, D. Pochan, Semiflexible Chain Networks Formed via Self-Assembly of β-Hairpin Molecules. *Phys. Rev. Lett.* **93**, 268106 (2004).

110. K. Rajagopal, B. Ozbas, D. J. Pochan, J. P. Schneider, Probing the importance of lateral hydrophobic association in self-assembling peptide hydrogelators. *Eur. Biophys. J.* **35**, 162–9 (2006).

111. B. Ozbas, K. Rajagopal, L. Haines-Butterick, J. P. Schneider, D. J. Pochan, Reversible stiffening transition in beta-hairpin hydrogels induced by ion complexation. *J. Phys. Chem. B*. **111**, 13901–8 (2007).

112. T. Yucel, C. M. Micklitsch, J. P. Schneider, D. J. Pochan, Direct Observation of Early-Time Hydrogelation in beta-Hairpin Peptide Self-Assembly. *Macromolecules*. **41**, 5763–5772 (2008).

113. E. F. Banwell *et al.*, Rational design and application of responsive alpha-helical peptide hydrogels. *Nat. Mater.* **8**, 596–600 (2009).

114. W. Shen, R. G. H. Lammertink, J. K. Sakata, J. A. Kornfield, D. A. Tirrell, Assembly of an artificial protein hydrogel through leucine zipper aggregation and bisulfide bond formation. *Macromolecules*. **38**, 3909–3916 (2005).

115. W. A. Petka *et al.*, Reversible Hydrogels from Self-Assembling Artificial Proteins. *Science (80-.).* **281**, 522–392 (1998).

116. P. Jing, J. S. Rudra, A. B. Herr, J. H. Collier, Self-assembling peptide-polymer hydrogels designed from the coiled coil region of fibrin. *Biomacromolecules*. **9**, 2438–2446 (2008).

117. D. Papapostolou *et al.*, Engineering nanoscale order into a designed protein fiber. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10853–8 (2007).

118. C. T. S. Wong Po Foo, J. S. Lee, W. Mulyasasmita, A. Parisi-Amon, S. C. Heilshorn, Two-component protein-engineered physical hydrogels for cell encapsulation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22067–72 (2009).

119. S. Sacanna, W. T. M. Irvine, P. M. Chaikin, D. J. Pine, Lock and key colloids. *Nature*. **464**, 575–578 (2010).

120. T. Sprules, N. Green, M. Featherstone, K. Gehring, Lock and Key Binding of the HOX YPWM Peptide to the PBX Homeodomain. *J. Biol. Chem.* **278**, 1053–1058 (2003).

121. R. Schwyzer, 100 years lock-and-key concept: are peptide keys shaped and guided to their receptors by the target cell membrane? *Biopolymers*. **37**, 5–16 (1995).

122. C. A. Helm, W. Knoll, J. N. Israelachvili, Measurement of ligand-receptor interactions. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 8169–73 (1991).

123. F. W. Lichtenthaler, 100 Years“Schlüssel-Schloss-Prinzip”: What Made Emil Fischer Use this Analogy? *Angew. Chemie Int. Ed. English*. **33**, 2364–2374 (1995).

124. M. Wilchek, E. A. Bayer, The avidin-biotin complex in bioanalytical applications. *Anal. Biochem.* **171**, 1–32 (1988).

125. M. Holzinger, M. Singh, S. Cosnier, Biotin-β-cyclodextrin: a new host-guest system for the immobilization of biomolecules. *Langmuir*. **28**, 12569–74 (2012).

126. R. Rohs *et al.*, Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–69 (2010).

127. I. W. Hamley, Peptide fibrillization. *Angew. Chem. Int. Ed. Engl.* **46**, 8128–47 (2007).

128. H. Cui, M. J. Webber, S. I. Stupp, Self-assembly of peptide amphiphiles: from molecules to nanostructures to biomaterials. *Biopolymers*. **94**, 1–18 (2010).

129. E. van der Linden, P. Venema, Self-assembly and aggregation of proteins. *Curr. Opin. Colloid Interface Sci.* **12**, 158–165 (2007).

130. J. D. Hartgerink, E. R. Zubarev, S. I. Stupp, Supramolecular one-dimensional objects. *Curr. Opin. Solid State Mater. Sci.* **5**, 355–361 (2001).

131. C. M. Dobson, Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **15**, 3–16 (2004).

132. A. Aggeli *et al.*, Hierarchical self-assembly of chiral rod-like molecules as a model for peptide beta -sheet tapes, ribbons, fibrils, and fibers. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11857–62 (2001).

133. W. S. Gosal, A. H. Clark, S. B. Ross-Murphy, Fibrillar beta-lactoglobulin gels: Part 1. Fibril formation and structure. *Biomacromolecules*. **5**, 2408–19.

134. W. S. Gosal, A. H. Clark, S. B. Ross-Murphy, Fibrillar beta-lactoglobulin gels: Part 2. Dynamic mechanical characterization of heat-set systems. *Biomacromolecules*. **5**, 2420–9.

135. W. S. Gosal, A. H. Clark, S. B. Ross-Murphy, Fibrillar beta-lactoglobulin gels: Part 3. Dynamic mechanical characterization of solvent-induced systems. *Biomacromolecules*. **5**, 2430–8.

136. C. Veerman *et al.*, Gelation kinetics of β-hairpin peptide hydrogel networks. *Macromolecules*. **39**, 6608–6614 (2006).

137. K. Rajagopal, M. S. Lamm, L. A. Haines-Butterick, D. J. Pochan, J. P. Schneider, Tuning the pH responsiveness of beta-hairpin peptide folding, self-assembly, and hydrogel material formation. *Biomacromolecules*. **10**, 2619–25 (2009).

138. R. V. Rughani, M. C. Branco, D. J. Pochan, J. P. Schneider, De novo design of a shear-thin recoverable peptide-based hydrogel capable of intrafibrillar photopolymerization. *Macromolecules*. **43**, 7924–7930 (2010).

139. D. A. Salick, J. K. Kretsinger, D. J. Pochan, J. P. Schneider, Inherent antibacterial activity

of a peptide-based beta-hairpin hydrogel. *J. Am. Chem. Soc.* **129**, 14793–9 (2007).

140. R. P. Nagarkar, R. A. Hule, D. J. Pochan, J. P. Schneider, De novo design of strand-swapped beta-hairpin hydrogels. *J. Am. Chem. Soc.* **130**, 4466–74 (2008).

141. A. A. Canutescu, R. L. Dunbrack, Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci.* **12**, 963–72 (2003).

142. D. Ting *et al.*, Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Comput. Biol.* **6**, e1000763 (2010).

143. H. C. Fry *et al.*, Computational de novo design and characterization of a protein that selectively binds a highly hyperpolarizable abiological chromophore. *J. Am. Chem. Soc.* **135**, 13914–26 (2013).

144. P. Chowdhury *et al.*, Fluorescence Correlation Spectroscopic Study of Serpin Depolymerization by Computationally Designed Peptides. *J. Mol. Biol.* **369**, 462–473 (2007).

145. J. R. Calhoun *et al.*, Computational Design and Characterization of a Monomeric Helical Dinuclear Metalloprotein. *J. Mol. Biol.* **334**, 1101–1115 (2003).

146. J. C. Phillips *et al.*, Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

147. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).

148. W. Humphrey, A. Dalke, K. Schulten, {VMD} -- {V}isual {M}olecular {D}ynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

149. P. S. Nerenberg, T. Head-Gordon, Optimizing Protein−Solvent Force Fields to Reproduce Intrinsic Conformational Preferences of Model Peptides. *J. Chem. Theory Comput.* **7**, 1220–1230 (2011).

150. R. Shevchuk, D. Prada-Gracia, F. Rao, Water structure-forming capabilities are temperature shifted for different models. *J. Phys. Chem. B*. **116**, 7538–43 (2012).

151. J.-P. Ryckaert, G. Ciccotti, H. J. . Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).

152. C. Yan *et al.*, Injectable solid hydrogel: mechanism of shear-thinning and immediate recovery of injectable β-hairpin peptide hydrogels. *Soft Matter*. **6**, 5143 (2010).

153. S. R. Leonard *et al.*, Solid-state NMR evidence for β-hairpin structure within MAX8 designer peptide nanofibers. *Biophys. J.* **105**, 222–30 (2013).

154. K. Nagy-Smith, E. Moore, J. Schneider, R. Tycko, Molecular structure of monomorphic peptide fibrils within a kinetically trapped hydrogel network. *Proc. Natl. Acad. Sci.* **112**, 9816–9821 (2015).

155. A. Kahler, H. Sticht, A. H. C. Horn, Conformational stability of fibrillar amyloid-beta oligomers via protofilament pair formation - a systematic computational study. *PLoS One*. **8**, e70521 (2013).

156. A. Morriss-Andrews, G. Bellesia, J.-E. Shea, β-sheet propensity controls the kinetic pathways and morphologies of seeded peptide aggregation. *J. Chem. Phys.* **137**, 145104 (2012).

157. L. Larini *et al.*, Role of β-hairpin formation in aggregation: the self-assembly of the amyloid-β(25-35) peptide. *Biophys. J.* **103**, 576–86 (2012).

158. A. Melquiond, G. Boucher, N. Mousseau, P. Derreumaux, Following the aggregation of amyloid-forming peptides by computer simulations. *J. Chem. Phys.* **122**, 174904/1–174904/8 (2005).

159. G. Wei, W. Song, P. Derreumaux, N. Mousseau, S. Wei, Self-assembly of amyloid-forming peptides by molecular dynamics simulations. *Front. Biosci.* **13**, 5681–5692 (2008).

160. X. Yu *et al.*, Atomic-scale simulations confirm that soluble beta-sheet-rich peptide self-assemblies provide amyloid mimics presenting similar conformational properties. *Biophys. J.* **98**, 27–36 (2010).

161. J. Zheng, H. Jang, B. Ma, C.-J. Tsai, R. Nussinov, Modeling the Alzheimer Abeta17-42 fibril architecture: tight intermolecular sheet-sheet association and intramolecular hydrated cavities. *Biophys. J.* **93**, 3046–57 (2007).

162. R. A. Hule, R. P. Nagarkar, B. Hammouda, J. P. Schneider, D. J. Pochan, Dependence of Self-Assembled Peptide Hydrogel Network Structure on Local Fibril Nanostructure. *Macromolecules*. **42**, 7137–7145 (2009).

163. D. Danino, Y. Talmon, R. Zana, Vesicle-to-Micelle Transformation in Systems Containing Dimeric Surfactants. *J. Colloid Interface Sci.* **185**, 84–93 (1997).

164. P. K. Vinson, J. R. Bellare, H. T. Davis, W. G. Miller, L. E. Scriven, Direct imaging of surfactant micelles, vesicles, discs, and ripple phase structures by cryo-transmission electron microscopy. *J. Colloid Interface Sci.* **142**, 74–91 (1991).

165. D. E. Discher *et al.*, Emerging applications of polymersomes in delivery: From molecular dynamics to shrinkage of tumors. *Prog. Polym. Sci.* **32** (2007), pp. 838–857.

166. C. LoPresti, H. Lomas, M. Massignani, T. Smart, G. Battaglia, Polymersomes: nature inspired nanometer sized compartments. *J. Mater. Chem.* **19**, 3576 (2009).

167. R. P. Brinkhuis, F. P. J. T. Rutjes, J. C. M. van Hest, Polymeric vesicles in biomedical applications. *Polym. Chem.* **2**, 1449 (2011).

168. A. M. Rosales, R. a. Segalman, R. N. Zuckermann, Polypeptoids: A Model System to Study the Effect of Monomer Sequence on Polymer Properties and Self-Assembly. *Soft Matter*, 8400–8414 (2013).

169. Z. M. Hudson *et al.*, Tailored hierarchical micelle architectures using living crystallization-driven self-assembly in two dimensions. *Nat Chem*. **6**, 893–898 (2014).

170. A. O. Moughton, M. A. Hillmyer, T. P. Lodge, Multicompartment block polymer micelles. *Macromolecules*. **45** (2012), pp. 2–19.

171. M. Huang *et al.*, Selective assemblies of giant tetrahedra via precisely controlled positional interactions. *Science (80-.).* **348**, 424–428 (2015).

172. A. H. Gröschel *et al.*, Guided hierarchical co-assembly of soft patchy nanoparticles.

Nature. **503**, 247–251 (2013).

173. C. Liu, M. A. Hillmyer, T. P. Lodge, Evolution of multicompartment micelles to mixed corona micelles using solvent mixtures. *Langmuir*. **24**, 12001–12009 (2008).

174. D. J. Pochan *et al.*, Multicompartment and multigeometry nanoparticle assembly. *Soft Matter*. **7**, 2500 (2011).

175. T. Wang, D. Schiffels, S. Martinez Cuesta, D. Kuchnir Fygenson, N. C. Seeman, Design and characterization of 1D nanotubes and 2D periodic arrays self-assembled from DNA multi-helix bundles. *J. Am. Chem. Soc.* **134**, 1606–1616 (2012).

176. M. T. Krejchi *et al.*, Chemical sequence control of beta-sheet assembly in macromolecular crystals of periodic polypeptides. *Science*. **265**, 1427–32 (1994).

177. S. M. Yu *et al.*, Smectic ordering in solutions and films of a rod-like polymer owing to monodispersity of chain length. *Nature*. **389**, 167–170 (1997).

178. K. B. Vargo, R. Parthasarathy, D. a. Hammer, Self-assembly of tunable protein suprastructures from recombinant oleosin. *Proc. Natl. Acad. Sci.* **109**, 11657–11662 (2012).

179. E. P. Holowka, V. Z. Sun, D. T. Kamei, T. J. Deming, Polyarginine segments in block copolypeptides drive both vesicular assembly and intracellular delivery. *Nat. Mater.* **6**, 52–57 (2007).

180. J. D. Brodin, S. J. Smith, J. R. Carr, F. A. Tezcan, Designed, Helical Protein Nanotubes with Variable Diameters from a Single Building Block. *J. Am. Chem. Soc.* **137**, 10468–10471 (2015).

181. J. Y. Shu, B. Panganiban, T. Xu, *Peptide-Polymer Conjugates: From Fundamental Science to Application* (2013; http://www.annualreviews.org/doi/abs/10.1146/annurev-physchem-040412-110108), vol. 64.

182. J. Hume *et al.*, Engineered coiled-coil protein microfibers. *Biomacromolecules*. **15**, 3503–3510 (2014).

183. Y. Loo, S. Zhang, C. A. E. Hauser, From short peptides to nanofibers to macromolecular assemblies in biomedicine. *Biotechnol. Adv.* **30** (2012), pp. 593–603.

184. L. E. R. O'Leary, J. A. Fallas, E. L. Bakota, M. K. Kang, J. D. Hartgerink, Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. *Nat. Chem.* **3**, 821–828 (2011).

185. C. Xu *et al.*, Rational design of helical nanotubes from self-assembly of coiled-coil lock washers. *J. Am. Chem. Soc.* **135**, 15565–15578 (2013).

186. N. C. Burgess *et al.*, Modular Design of Self-Assembling Peptide-Based Nanotubes. *J. Am. Chem. Soc.* **137**, 10554–10562 (2015).

187. E. H. Egelman *et al.*, Structural plasticity of helical nanotubes based on coiled-coil assemblies. *Structure*. **23**, 280–289 (2015).

188. Y.-T. T. Lai *et al.*, Principles for designing ordered protein assemblies. *Trends Cell Biol.* **22**, 653–61 (2012).

189. Y. Deng *et al.*, Antiparallel four-stranded coiled coil specified by a 3-3-1 hydrophobic heptad repeat. *Structure*. **14**, 247–55 (2006).

190. O. D. Testa, E. Moutevelis, D. N. Woolfson, CC+: a relational database of coiled-coil structures. *Nucleic Acids Res.* **37**, D315–22 (2009).

191. J. Dundas *et al.*, CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116–8 (2006).

192. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 12–21 (2010).

193. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–97 (2007).

194. A. Guinier, G. Fournet, Small angle scattering of X-rays. *J. Polym. Sci.* **1**, 268 (1955).

195. S. R. Kline, Reduction and analysis of SANS and USANS data using IGOR Pro. *J. Appl. Crystallogr.* **39**, 895–900 (2006).

196. S. Howorka, Rationally engineering natural protein assemblies in nanobiotechnology. *Curr. Opin. Biotechnol.* **22**, 485–491 (2011).

197. F. Baneyx, J. F. Matthaei, Self-assembled two-dimensional protein arrays in bionanotechnology: from S-layers to designed lattices. *Curr. Opin. Biotechnol.* **28C**, 39–45 (2013).

198. H. Inaba, S. Kitagawa, T. Ueno, Protein Needles as Molecular Templates for Artificial Metalloenzymes. *Isr. J. Chem.* **55**, 40–50 (2015).

199. C. Zhang *et al.*, Self-Assembled Peptide Nanofibers Designed as Biological Enzymes for Catalyzing Ester Hydrolysis. *ACS Nano*. **8**, 11715–11723 (2014).

200. T. Luo, L. He, P. Theato, K. L. Kiick, Thermoresponsive Self-Assembly of Nanostructures from a Collagen-Like Peptide-Containing Diblock Copolymer. *Macromol. Biosci.* **15**, 111–123 (2015).

201. L. Miao *et al.*, Quantum-Dot-Induced Self-Assembly of Cricoid Protein for Light Harvesting. *ACS Nano*. **8**, 3743–3751 (2014).

202. S. Sengupta, F. Würthner, Chlorophyll J-Aggregates: From Bioinspired Dye Stacks to Nanotubes, Liquid Crystals, and Biosupramolecular Electronics. *Acc. Chem. Res.* **46**, 2498–2512 (2013).

203. H. Sun *et al.*, Micelle-Induced Self-Assembling Protein Nanowires: Versatile Supramolecular Scaffolds for Designing the Light-Harvesting System. *ACS Nano*. **10**, 421–428 (2016).

204. A. T. Preslar *et al.*, Gd(III)-Labeled Peptide Nanofibers for Reporting on Biomaterial Localization *in Vivo*. *ACS Nano*. **8**, 7325–7332 (2014).

205. Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276** (1997), pp. 307–326.

206. A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, Likelihood-enhanced fast

translation functions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **61**, 458–464 (2005).

207. P. D. Adams *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).

208. S. Rämisch, R. Lizatović, I. André, Automated de novo phasing and model building of coiled-coil proteins. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 606–614 (2015).

209. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

210. B. North, C. M. Summa, G. Ghirlanda, W. F. DeGrado, D(n)-symmetrical tertiary templates for the design of tubular proteins. *J. Mol. Biol.* **311**, 1081–1090 (2001).

211. J. Zou, J. G. Saven, Statistical theory of combinatorial libraries of folding proteins: energetic discrimination of a target structure. *J. Mol. Biol.* **296**, 281–94 (2000).

212. C. C. Valley *et al.*, The Methionine-aromatic Motif Plays a Unique Role in Stabilizing Protein Structure. *J. Biol. Chem.* **287**, 34979–34991 (2012).

213. E. Krissinel, Crystal contacts as nature’s docking solutions. *J. Comput. Chem.* **31**, 133–43 (2010).

214. C. F. Macrae *et al.*, Mercury CSD 2.0 - New features for the visualization and investigation of crystal structures. *J. Appl. Crystallogr.* **41** (2008), pp. 466–470.

215. C. Bissantz, B. Kuhn, M. Stahl, A medicinal chemist’s guide to molecular interactions. *J. Med. Chem.* **53**, 5061–84 (2010).

216. A. R. Fersht *et al.*, Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*. **314**, 235–8.

217. D. H. Williams, M. S. Searle, J. P. Mackay, U. Gerhard, R. A. Maplestone, Toward an estimation of binding constants in aqueous solution: studies of associations of vancomycin group antibiotics. *Proc. Natl. Acad. Sci.* **90**, 1172–1178 (1993).

218. J. Tsai, M. Levitt, Evidence of turn and salt bridge contributions to β-hairpin stability: MD simulations of C-terminal fragment from the B1 domain of protein G. *Biophys. Chem.* **101**, 187–201 (2002).

219. T. M. Iqbalsyah, A. J. Doig, Anticooperativity in a Glu-Lys-Glu salt bridge triplet in an isolated alpha-helical peptide. *Biochemistry*. **44**, 10449–10456 (2005).

220. A. Horovitz, L. Serrano, B. Avron, M. Bycroft, A. R. Fersht, Strength and co-operativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* **216**, 1031–1044 (1990).

221. U. C. Singh, Probing the salt bridge in the dihydrofolate reductase-methotrexate complex by using the coordinate-coupled free-energy perturbation method. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 4280–4 (1988).

222. Y. N. Imai, Y. Inoue, Y. Yamamoto, Propensities of Polar and Aromatic Amino Acids in Noncanonical Interactions: Nonbonded Contacts Analysis of Protein−Ligand Complexes in Crystal Structures. *J. Med. Chem.* **50**, 1189–1196 (2007).

223. S. Grimme, Molecular Electrostatic Potentials: Concepts and Applications. *Zeitschrift für*

Phys. Chemie. **205**, 136–137 (1998).

224. C. Zhang, G. Vasmatzis, J. L. Cornette, C. DeLisi, Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.* **267**, 707–726 (1997).

225. R. M. Jackson, H. A. Gabb, M. J. E. Sternberg, Rapid refinement of protein interfaces incorporating solvation: application to the docking problem. *J. Mol. Biol.* **276**, 265–285 (1998).