Sequence-Structure Correlations in the MaSp1 Protein of *N. clavipes* Dragline Silk

by

Graham Hayden Bratzel
B.S. Mechanical Engineering
Seattle University, 2009

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of
Master of Science in Mechanical Engineering

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2011

Author .....
Department of Mechanical Engineering
May 6, 2011

Certified by .....
Markus J. Buehler
Associate Professor
Department of Civil and Environmental Engineering
Thesis Supervisor

Certified by .....
Roger D. Kamm
Professor of Biological and Mechanical Engineering
Department of Mechanical Engineering
Thesis Reader

Accepted by .....
David E. Hardt
Chairman, Department Committee on Graduate Students
Department of Mechanical Engineering

Sequence-Structure Correlations in the MaSp1 Protein of *N. clavipes* Dragline Silk

by

Graham Hayden Bratzel

Submitted to the Department of Mechanical Engineering
on May 6, 2011, in partial fulfillment of the
requirements for the degree of
Master of Science in Mechanical Engineering

Abstract

Silk is a hierarchically structured protein fiber with exceptional tensile strength and extensibility, making it one of the toughest and most versatile biocompatible materials. While experimental studies have shown that the molecular structure of silk has a direct influence on the stiffness, toughness, and failure strength of silk, few molecular-level analyses of the nanostructure of silk assemblies, in particular under variations of genetic sequences, have been published. Here, atomistic-level structures of wildtype as well as modified MaSp1 protein from the *N. clavipes* spider dragline silk sequences are reported, obtained using an *in silico* approach based on replica exchange molecular dynamics (REMD) and explicit water molecular dynamics. In particular, the atomistic simulations discussed in this parametric study explore the effects of the poly-alanine length of the *N. clavipes* MaSp1 peptide sequence, solvent conditions, and nanomechanical loading conditions on secondary and tertiary structure predictions as well as the nanomechanical behavior of a unit cell of 15 strands with 900–1000 total residues used to represent a cross-linking β-sheet crystal node in the network within a fibril of the dragline silk thread. Understanding the behavior of this node at the molecular scale is critical for potentially bypassing strength limits at this length scale and vastly improving silk for medical and textile purposes as well as synthetic elastomers and polymer or aramid fiber composites with a similar molecular structure and noncovalent bonding for aerospace, armor, and medical applications. The main hypothesis tested is that there exists a critical minimum length of the poly-alanine repeat that ensures the formation of a robust cross-linking the β-sheet crystal. Confirming earlier experimental and computational work, a structural analysis reveals that poly-alanine regions in silk predominantly form distinct and orderly β-sheet crystal domains while disorderly regions are formed by glycine-rich repeats that consist of 3_{10}-helix type structures and β-turns. These predictions are directly validated against experimental data based on dihedral angle pair calculations presented in Ramachandran plots combined with an analysis of the secondary structure content. The key results of this study are:

- A strong dependence of the resulting silk nanostructure on the poly-alanine length. The wildtype poly-alanine repeat length of six residues defines a critical minimum length that consistently results in clearly defined β-sheet nanocrystals allowing for misalignment. For poly-alanine lengths below six residues, the β-sheet nanocrystals are not well-defined or not visible at all, while for poly-alanine lengths above six the characteristic nanocomposite structure of silk emerges with no significant improvement of the quality of the β-sheet nanocrystal geometry.
- A simple biophysical model is presented that explains the minimum length scale based on the mechanistic insight gained from the molecular simulations. The efficient stacking of the β-sheets of a well-defined crystal reinforces local hydrophobicity and prevents water diffusion into a crystal above a critical size.
- Nanomechanical testing reveals that the combination of the 12-alanine length case and central pull-out loading conditions results in delayed failure by employing a hierarchy of strong β-sheets and soft, extensible semi-amorpous regions to overcome a predicted H-bond saturation.

This work constitutes the most comprehensive study to-date of the molecular structure prediction and nanomechanical behavior of dragline silk. Building upon previous computational studies that used similar methods for structure prediction and mechanical analysis, *e.g.* REMD and force-control loading, this work presents:

- the first results of the near-native structures determined by REMD after equilibration in TIP3P explicit solvent,
- the first parametric study of the effects of modifying the wildtype poly-alanine segment length to values outside the range naturally observed for MaSp on structure prediction and nanomechanical behavior, and
- the first comparison between previously published loading conditions, *i.e.* the Stretch test, and the novel Pull-out loading conditions that are hypothesized to be more appropriate for modeling of the *in situ* loading of the cross-linking β-sheet crystal.

Further parametric studies in peptide sequence to optimize bulk fiber properties must involve changes in simulated nanomechanical loading conditions to properly assess the effects of the changes in peptide sequence. These findings set the stage for understanding how variations in the spidroin sequence can be used to engineer the structure and thereby functional properties of this biological superfiber, and present a design strategy for the genetic optimization of spidroins for enhanced mechanical properties. The approach used here may also find application in the design of other self-assembled molecular structures and fibers and in particular biologically inspired or completely synthetic systems.

Thesis Supervisor: Markus J. Buehler
Title: Associate Professor
Department of Civil and Environmental Engineering

Acknowledgments

I express my profound gratitude for all the people and institutions that have supported, guided, and inspired the path of my academic career, making it an extraordinary experience of knowledge and curiosity.

I thank Prof. Frank Shih of Seattle University for his guidance and supervision of my academic and research career, for teaching me to learn from failure when the tricycle only works in reverse, and that activity is not always accomplishment. I also thank the rest of the faculty at Seattle University, who gave me a solid foundation of an education in mechanical engineering and made it possible for me to be admitted at the Massachusetts Institute of Technology.

I express my enormous gratitude to Prof. Markus Buehler, who introduced me to the field of atomistic, molecular, and multiscale modeling and to the fascinating world of spider silk and hierarchical materials. I greatly appreciate his excellent mentorship and constant guidance, and for getting me to write more journal articles, book chapters, and e-mails than I ever expected. I am very thankful to Steve Cranford, an inspired researcher, a great teacher, and a fast friend. I express my gratitude to him for his insightful suggestions and for having a Python code at the ready for almost everything. I also thank all my other peers at LAMM for guidance, suggestions, some excellent discussions, and the friendship we've shared: Sinan Keten, Dipanjan Sen, Andre Garcia, Raffaella Paparcone, Melis Arslan, Zhao Qin, and Alfonso Gautieri. I express gratitude and respect to MIT — a place I only dreamed of, and still often do — for paving the way of my future career.

I am endlessly thankful to my family for all their support and their love. Thanks to my wife Rachel and my dog Teddy for getting me out of the office and across the river once in a while. And finally, thanks to my parents, who always got me to try new things, never stopped being impressed by where those things got me, and who are ultimately responsible for me being at MIT.

List of Publications

- G. Bratzel & M.J. Buehler. Sequence-structure correlations and size effects in silk nanostructure: Poly-Ala repeat of N. clavipes MaSp1 is naturally optimized, *Journal of the Mechanical Behavior of Biomedical Materials* (in review), 2011.

- G. Bratzel & M.J. Buehler. Mechanical Properties of Hierarchical Protein Materials, *Encyclopedia of Nanotechnology*, Springer 2011.

- G. Bratzel, S. Cranford, M.J. Buehler. Bioinspired noncovalently crosslinked fuzzy carbon nanotube bundles with superior toughness and strength, *Journal of Materials Chemistry*, Vol. 20, pp. 10465-10474, 2010.

Contents

List of Figures

List of Tables

Chapter 1

Background and Motivation

1.1 Introduction to Silk

Spider silk is an extraordinary biomaterial that surpasses most synthetic fibers in terms of toughness through a balance of ultimate strength and extensibility [1, 2, 3, 4, 5]. The source of spider silk's remarkable properties has been attributed to the specific secondary, tertiary, and quartenary structures of proteins found in the repeating units of the polypeptide sequences that comprise spider silk [6], which self-assemble into a hierarchical structure. Experimental studies have primarily focused on mapping the repeating sequence units of spider silk and the basic structural building blocks and crystallinity of fibrils. The webs of higher spiders, including the Golden Orb-Weaver *Nephila clavipes*, are composed of multiple kinds of silk, each with distinct repeating sequence units and mechanical properties that are adapted for the purpose of that part of the web, as illustrated in Figure 1-1. Major ampullate silk, the strongest kind of silk, is used for the spokes, outer frame, and dragline [7]. Two distinct proteins are typically found in dragline silks with similar sequences across species [8]. The dragline silk of *Nephila clavipes*, one of the most studied spider silks, contains major ampullate spidroins MaSp1 and MaSp2 proteins with different repeat units and distinct mechanical functions [9, 6, 10, 11]. MaSp1 contains poly-alanine, *i.e.* poly-Ala or $(A)_n$, domains within glycine-rich $(GGX)_n$ and (GA) repeats, where X typically stands for alanine (A), tyrosine (Y), leucine (L), or glutamine (Q). Species surveys

have suggested that MaSp1 is more prevalent in the spider dragline silk than MaSp2, with a ratio of approximately 3:2 or higher, depending on the species [12, 13, 10, 14].

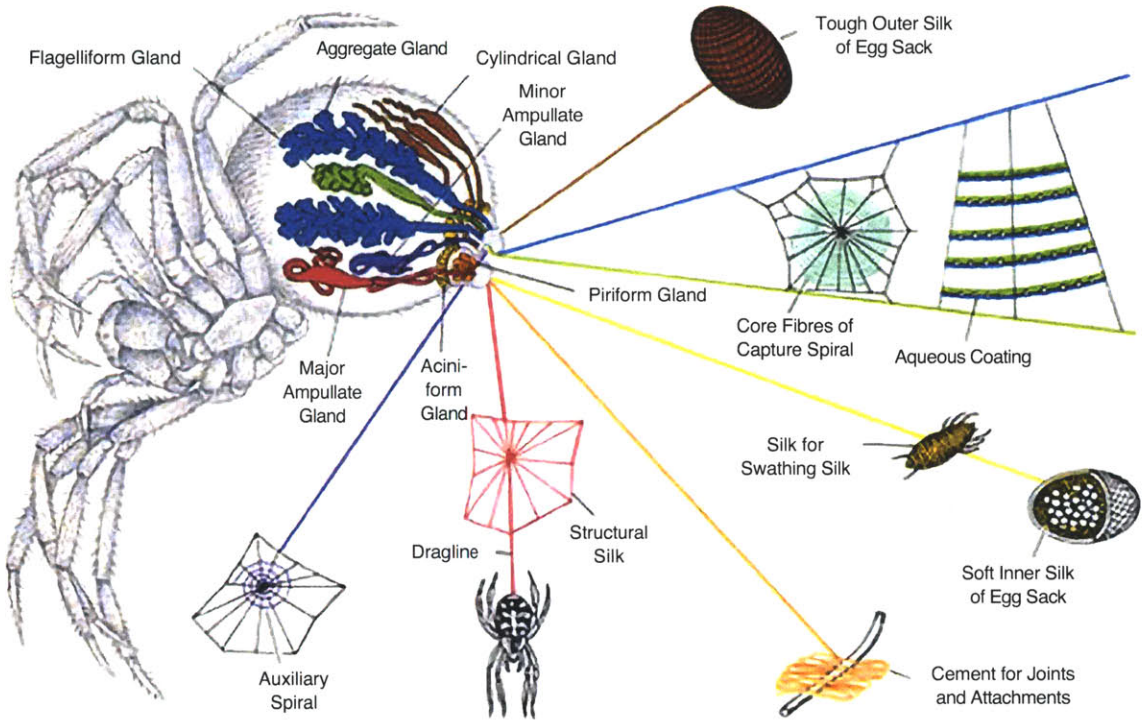

Figure 1-1: The types of glands and their associated silk threads and applications for higher spiders such as the Golden Orb-Weaver *Nephila clavipes*. This thesis investigates the Dragline silk (bottom, red) from the Major Ampullate Gland. Figure reprinted with permission from [15].

Recent investigations have revealed that antiparallel β-sheet crystals play a key role in defining the mechanical properties of silk by providing cross-linking domains embedded in a semi-amorphous glycine-rich matrix with extensible hidden-length [16, 17, 18, 19, 20]. Studies have also shown that the hydration level and solvent conditions (*e.g.* ion content and pH range) play a large role in the structure and mechanical properties of silk proteins [21, 22] and even the transition from concentrated dope to final silk in the spinning duct. The cross-linking β-sheet crystals employ a dense network of hydrogen bonds [19, 20], have dimensions of a few nanometers, and constitute at least 10-15% of the silk volume. The existence of 3_{10} helices and β-turn or β-spiral conformations has been suggested for the amorphous domains [23, 16, 17, 18]. Although atomistic structure predictions of the wildtype sequence

have been reported with implicit solvent methods [19, 20], no robust atomistic-level structural model with explicit solvent or a systematic analysis of sequence-structure correlations has yet been reported. It is anticipated that novel statistical mechanics approaches [24], experimental methods, such as X-ray diffraction and scattering [25, 26], solid-state nuclear magnetic resonance (NMR) [2, 27, 11, 28] and Raman spectroscopy [29, 18, 30], combined with multiscale atomistic modeling methods such as those based on Density Functional Theory (DFT) [31, 19] or molecular dynamics (MD) [32, 33, 34, 19] will provide more insight into the atomic resolution structure for spider silk and similarly complex materials. An earlier study of silk using replica exchange molecular dynamics (REMD) [35] has yielded the first results in comparison with experimental structure identification methods [36, 17, 11]. Other recent computational studies have characterized the mechanics of a protien similar to MaSp1 [37], but the authors used a constructed structure that is potentially far from the native state. Owing to the lack of current large-scale atomistic models, the links between peptide sequence of actual silk protein remains poorly understood. The availability of powerful new methods to synthesize varied protein materials from the bottom up and will full control over the genetic sequence opens exciting opportunities to engineer materials such as spider silk for specific mechanical and other functional purposes.

The atomistic simulations discussed in this thesis explore the effects of the poly-Ala length of the *N. clavipes* MaSp1 peptide sequence, solvent conditions, and nanomechanical loading conditions on the secondary and tertiary structure predictions as well as the nanomechanical behavior of a unit cell of 15 strands with 900–1000 total residues used to represent a cross-linking β-sheet crystal node in the network within a fibril of the *N. clavipes* dragline silk thread. The challenges of reaching native (that is, equilibrium) structures within the time-scales accessible to conventional molecular dynamics simulations require enhanced sampling methods such as replica exchange molecular dynamics (REMD) [38]. In this study, REMD is used to investigate the structures formed by assemblies of segments of MaSp1 protein oligomers and follow the REMD structure prediction step with a careful molecular dynamics equilibration in explicit water solvent to refine the geometry of predicted structures. Along with

other protein structure prediction approaches [39, 40], REMD is considered to be a quite effective tool for investigating folding and aggregation of proteins, as it reduces the likelihood of kinetic trapping at non-native states [41]. Through a fast search of the conformation space at high temperatures and more detailed investigations at low temperatures, REMD allows the system to overcome energy barriers and local minima corresponding to non-native structures of proteins [42, 43, 44, 45] and facilitates identification of native protein structures from the amino acid sequence with atomistic resolution. This is described in more detail in Chapter 2 Materials and Methods.

1.2 Hierarchical Protein Materials

With only 20 standard amino acids as universal building blocks, evolutionary adaptation has resulted in a multitude of polypeptides and protein structures with a wide range of properties and applications [46]. A fundamental application of protein materials is mechanical, providing static and dynamic support and defenses for the organisms they comprise. To achieve specialized mechanical properties, animal tissue shows a wide range of complex, structured composites composed of proteins (*e.g.* collagen, keratin, and chitin) and inorganic minerals (*e.g.* calcite, hydroxyapatite, and aragonite) [47, 48]. Instead of summarizing primarily single-molecule protein mechanics studies [49, 50, 51, 52, 53], the main goal of this chapter is to give an introduction to the current knowledge of the nanomechanical properties of important structural protein materials and to explain the generic role of the hierarchical structures that give rise to exceptional mechanical properties that are superior to many synthetic engineering materials. For example, shell, bone, and antler are hierarchically structured composites of collagen and minerals and have toughnesses that are about one order of magnitude greater than engineering ceramics [54]. Natural polymers and polymer-based composites like collagen, keratin, and silk show elastic moduli and tensile strengths larger than those of engineering polymers made from similar building blocks. For example, silks have elastic moduli between 2 and 20

GPa and strengths in the range from 0.3 to 2 GPa [54]. Of man-made polymers, only Kevlar has a higher stiffness at 200 GPa and a strength up to 4 GPa, which it achieves through a highly oriented molecular structure which is typically characteristic of biological materials. Natural elastomers — including skin, muscle, cartilage, abductin, resilin, and elastin — exhibit elastic moduli and densities similar to those of engineering elastomers. Muscle is unique in combining a low modulus on the order of an elastomer with a high strength on the order of steel [54]. Natural cellular materials like cancellous bone maintain high strength at low bulk densities due to the high volume fractions of voids they contain [55].

Whether natural composite, polymer, or elastomer, many biological materials achieve their notable mechanical properties through a balance of strong covalent bonds and weaker noncovalent bonds (*e.g.* hydrogen bonds) in a multi-level hierarchical structure that spans from nanoscopic to macroscopic length scales [56, 57]. Known as the Universality-Diversity Paradigm [58] and summarized in Figure 1-2, a limited number of universal protein structures or mineral components can provide a wide diversity of geometrical arrangements within the natural composites that give rise to exceptional bulk mechanical properties as compared to the material properties of the individual constituents. The hierarchical structure of protein materials is analogous to the construction of symphonic music. In much the same way that all music is built from the superposition of simple sound waves, a limited list of universal protein building blocks can be combined in various ways at multiple hierarchical levels to yield a diverse groups of protein-based biomaterials. Only a small subset of this diverse group exhibit exceptional mechanical properties, *e.g.* bone or an orb-weaver spider, akin to Bach or Mozart standing out from the rest. Certain specially-constructed geometrical arrangements can enhance bulk properties much more than the rule of mixtures can predict. This puts the concept of geometry, assembly, and, more generally, the architecture of hierarchical structures at the forefront of the science of protein materials.

This section will first discuss how the multi-level structures of collagen fibers and silk present examples of specialized geometric arrangements of disparate protein and

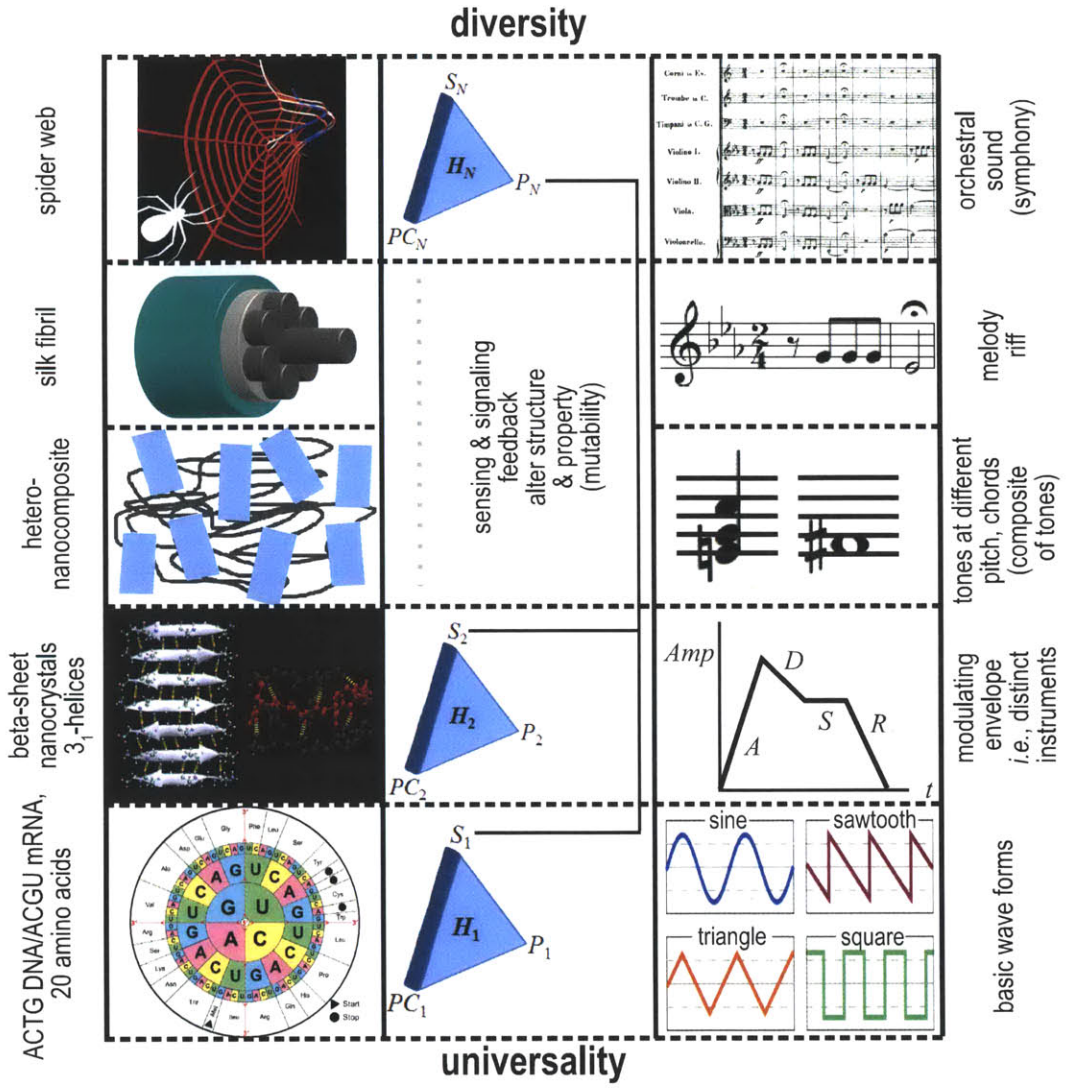

Figure 1-2: Overview of the Universality-Diversity-Paradigm (UDP) [56, 58]. The hierarchical structure of many protein materials is analogous to the construction of symphonic music by the superposition of simple sound waves. Figure reprinted from [56].

mineral components in a hierarchical structure that can overcome the strength limit at the base level to yield highly functional macroscale materials that rival engineering composite materials. Next is described an intrinsic strength limit at the nanoscale that exemplifies the constraints under which many biological systems operate, how these constraints are overcome, and how biomaterials exemplify a powerful paradigm turning molecular weakness into macroscopic strength.

1.2.1 Collagen and Silk: Exemplary Hierarchical Proteins

Collagen and silk are prime examples of proteins that organize a small set of amino acids and secondary structures, the universal building blocks of proteins, into a diverse class of materials. Collagen is present in various forms in almost all structural animal tissue and represents how the limitations of protein sub-units can be overcome by the complex way they are assembled (Figure 1-3). Skin, tendon, and cartilage are all largely collagenous composites [47]. In skin, collagen is sandwiched between a basement membrane and an overlying keratinized epidermis. In tendon, collagen fibers form rope-like structures which make up 70–80% of the volume, with the remainder being a combination of non-collagenous protein, polysaccharides, and inorganic salts. In cartilage, the collagen fibers are in a proteoglycan matrix with a small fraction of elastin fibers. Bone, shell, and antler are composites of calcite, hydroxyapatite or aragonite platelets dispersed in a helical matrix of collagen. A diverse class of materials result from the universal collagen protein. Figure 1-3 reveals that the hierarchical levels of collagen and silk show similar trends in structure across orders of magnitude of length scales. The specific sequence of amino acids, the universal building blocks, at the base level result in macroscale protein fibers that are optimized for tension and energy dissipation in their respective applications.

Silk is a hierarchically structured protein fiber with a high tensile strength and great extensibility, making it one of the toughest and most versatile materials known [9, 60, 61]. In contrast to synthetic polymers based on petrochemicals, silk is spun into strong and totally recyclable fibers at ambient temperatures, low pressures, and with water as the solvent. However, biomimetic reproductions of silk remain a challenge because of silk's unique microstructural features that can only be achieved by controlled self-assembly of protein oligomers with molecular precision [62, 63]. Unlike silkworms, some spiders can use different glands to create up to seven types of silk, from the strong structural dragline to the viscoeleastic capture silk and tough eggsack casing [60], as shown in Figure 1-1. Major ampullate silk, containing a high fraction of densely H-bonded domains, is used to provide the structural frame for the web

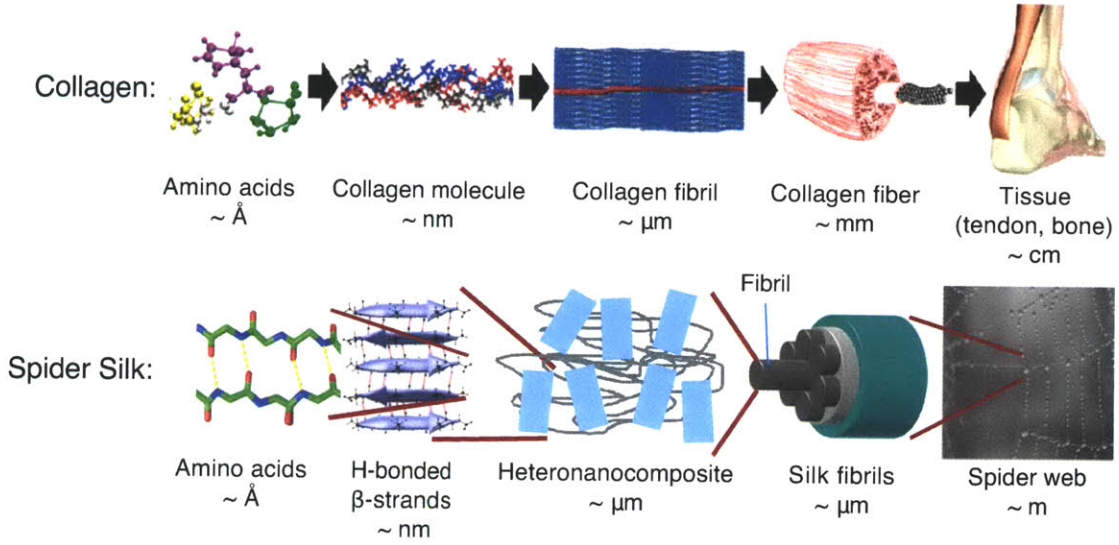

Figure 1-3: The hierarchical levels of collagen and silk reveal similar trends in structure across orders of magnitude of length scales. The specific sequence of amino acids, the universal building blocks, at the base level result in macroscale protein fibers that are optimized for tension and energy dissipation in their respective applications. Collagen figure adapted from [59]; silk figure adapted from [35].

and has an elastic modulus of around 10 GPa [61]. Capture silk, on the other hand, is a viscid biofilament containing cross-linked polymer networks and has an elastic modulus that is comparable to that of other elastomers [7].

The stress-strain behavior of silks are compared to Kevlar^{TM} and rubber in Figure 1-4. While rubber is extensible, and Kevlar^{TM} is stiff and strong, silks feature a combination of strength and toughness not typically found in synthetic materials. It is known that β-sheet crystals at the nanoscale play a key role in defining the mechanical properties of silk by providing stiff and orderly crosslinking domains embedded in a semi-amorphous matrix that consists predominantly of less orderly structures [65, 66]. These β-sheet nanocrystals, bonded by means of assemblies of H-bonds, have dimensions of a few nanometers and constitute roughly 10–15% of the silk volume. When silk fibers are stretched, the β-sheet nanocrystals reinforce the partially extended and oriented macromolecular chains by forming interlocking regions that transfer the load between chains under lateral loading, similar to their function in other mechanical proteins [67]. The hierarchical network of spider draglines, contrasting synthetic elastomers like rubber, enables quick energy absorption and efficiently

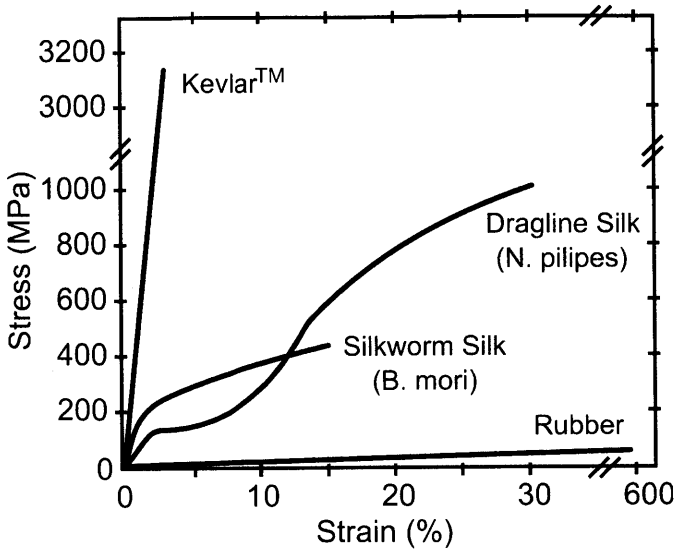

Figure 1-4: A summary of stress-strain plots showing superior strain-hardening behavior of dragline and silkworm silks. Rubber is extensible but not strong, while Kevlar^{TM} is stiff and strong but not tough. Silks, especially dragline silk, behave with a combination of strength and toughness not typically found in synthetic materials. Figure adapted from [64].

suppresses vibration during an impact [64]. Rubbers, composed of random polymer chains, display an elasticity primarily due to the change in conformational entropy of these chains. In contrast, the amorphous chains in silk filaments are extended and held in partial alignment with respect to the fiber axis in its natural state, resulting in remarkably different mechanical behaviors from rubbers [66, 67]. Mesoscale simulations reveal that the characteristic, cross-linking β-sheet nanocrystals govern the coupled effects of silk network realignment and strain-hardening behavior beyond the regime of unfolding amorphous chains [65].

Advances in nanotechnology may allow for an optimization of silk and silk-like synthetic fibers for specific applications by modifying the governing network at critical length scales. At the base of the structural hierarchy, β-sheet nanocrystal size may be tuned though genetic modification and control of the self-assembly process. Computational studies of spider silk proteins reveal that the nanoscale confinement of β-sheet nanocrystals in silks indeed has a fundamental role in achieving great stiffness, resilience, and fracture toughness at the molecular level of the structural hierarchy [65, 35], where H-bond cooperativity depends quite strongly on the size of

the crystals and breaks down once β-sheet nanocrystals exceed a critical size. Larger β-sheet nanocrystals are softer and fail catastrophically at much lower forces owing to crack-like flaw formation. This catastrophic breakdown leads to rapid disintegration of silk fibers, whereas reducing the nanocrystal dimensions below 3 nm increases the ultimate strength and the modulus many-fold [35]. Furthermore, noncovalent H-bonds are able to reform during stick-slip deformation, allowing smaller crystals a self-healing ability until complete rupture occurs.

It is also noted that at the macroscale, experiments have demonstrated that when the size of β-sheet nanocrystals is reduced by moderating the reeling speed or by metal infiltration, silk shows enhanced toughness and greater ultimate strength, exceeding that of steel and other engineered materials [67]. By understanding and controlling how H-bonded silk threads can reach great strength and toughness and overcome strength limitations at the molecular scale, the behavior of a broader class of β-sheet-rich protein materials and similar synthetic composite fibers can also be tuned and adapted.

Nature has much to offer for the next generation of both macroscale and nanoscale materials, in particular relating to the linking of vast and seemingly separated scales (from nano to macro). The expansion of the structural design space through the concept of merging structure and material (Figure 1-2) is a powerful concept that enables us to create highly functional materials through the use of relatively simple building block, a mechanism established in the Universality-Diversity-Paradigm of materials science [56, 58]. The transfer of concepts derived from biological materials, for example mimicking of the strain-hardening behavior of spider silk, can give rise to the design of advanced functional materials with exceptional kinetic energy buffering and absorption. Developing ordered nanocomposites that are based on the structures found in bone, antler, or nacre can dramatically increase the toughness of typically brittle structures and may enable us to use abundant materials such as silica in the design of tough, elastic and strong materials. Even the architecture of collagen's reinforced helices, fibrils, and fibers offer much inspiration in creating advanced tensile struts from seemingly inferior materials. The key issue is to better understand the

mechanisms of the origin of the mechanical properties of proteins from the bottom up, and the identification of systematic approaches to transfer the insight from biology to engineering design through an approach referred to as materiomics [56].

1.2.2 Overcoming Strength Limits

While covalent peptide bonds make up the backbone of a protein, noncovalent bonds, in particular weak bonds such as hydrogen bonds (H-bonds), are indispensable to biological function in defining the native folding and stability of protein structures. The amino acid sequence defines a protein's primary structure and guides the formation of higher-level structures via folding. Arrays of parallel adjacent H-bonds govern major universal secondary and tertiary structural motifs, chiefly α-helices and β-sheets, that organize the three-dimensional structure of polypeptides and thus determine their mechanical properties at a fundamental level. Refer to Figure 1-5 for definitions of the main secondary and tertia y structures of proteins.

The rupture of parallel interstrand H-bonds during β-sheet failure defines the true ultimate strength of β-sheet-rich proteins. Atomistic simulation combined with single-molecule force microscopy studies [52] have shown that proteins rich in β-sheets exhibit higher rupture forces since they employ parallel strands with numerous H-bonds that act as strong mechanical clamps under shear loading. At experimental pulling rates, most mechanical proteins exhibit a rupture force of a few hundred piconewtons (pN) [68], as seen in Figure 1-6b. An asymptotic limit at very low pulling rates can be inferred from a limit analysis for deformation through equilibrium, which suggests an upper strength limit for the unfolding β-sheet-rich proteins [69, 70, 71]. Based on this analysis for typical single-molecule length scales at a temperature of 300 K and a vanishing pulling rate, the asymptotic force of rupture is found to be on the order of 100–300 pN [72] (Figure 1-6). This asymptotic force limit depends strongly on the H-bond dissociation energy, which can vary with solvent conditions and which explains the range of strengths obtained in experiment.

For a given H-bond dissociation energy, the force limit is independent of the number of H-bonds being sheared simultaneously, given that the number of H-bonds

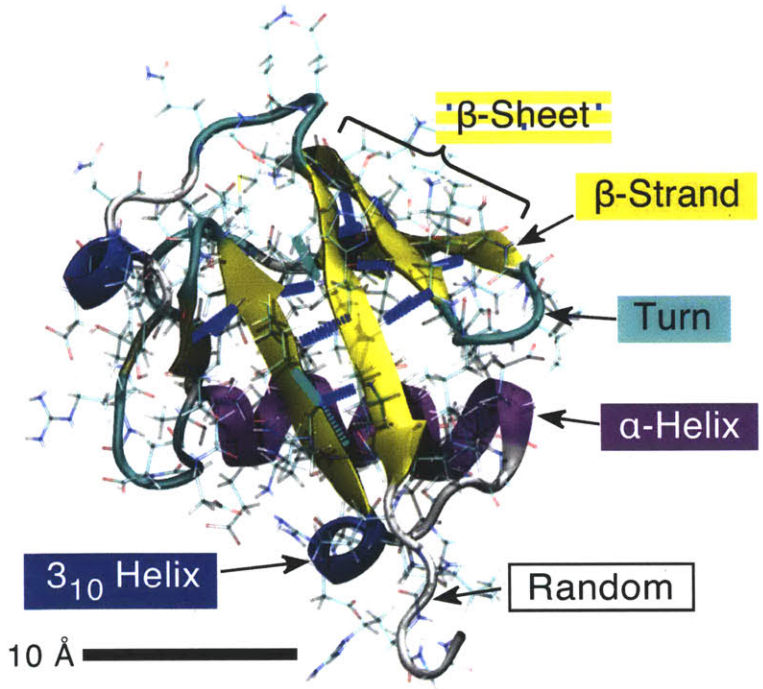

Figure 1-5: Secondary and tertiary structures of proteins (here, ubiquitin) are defined by folding. The color-coded "cartoon" representation is a graphical way to interpret secondary structure and is often more convenient than displaying the chemical bonds (shown as transparent sticks). H-bonds (dashed blue) connect adjacent β-strands into the pleated β-sheet tertiary structure.

in a deformed cluster is larger than a critical H-bond cluster size N_{cr}, which can be calculated for each case[72, 73, 74]. This is because it is found that no more than N_{cr}=3–4 H-bonds can rupture concurrently, thus defining a critical H-bond cluster size beyond which the strength can no longer increase and at which scale the asymptotic strength limit is reached. The structure of H-bond clusters in α-helices are in this range, with 3.6 residues per turn; and similarly, most β-sheet protein domains feature 3–8 H-bonds in each cluster.

Therefore, while the shear strength of β-sheets with less than N_{cr}=3–4 H-bonds benefits from the addition of H-bonds by employing longer β-strands, the shear strength (the maximum force divided by the sheared area) decays for β-sheets of 4 H-bonds or longer, as seen in Figure 1-7b, as the H-bonds along the full length of the β-sheets do not rupture concurrently since they can only rupture in groups of

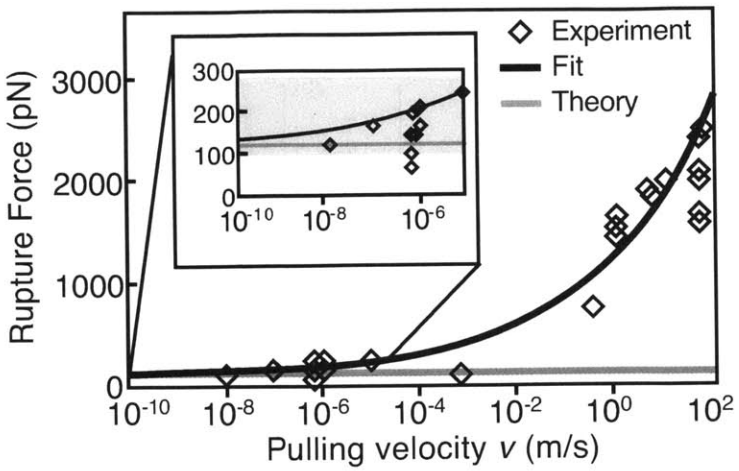

Figure 1-6: Strength limit of H-bond clusters under shear loading. At vanishing pulling rates (see inset for a detailed view of the data), clusters of H-bonds resist a force limit of 100–300 pN, according to experimental and computational results from fibronectin, immunoglobin, and titin deformation [72]. This data reflects the rate-dependence of failure seen in many biomaterials.

N_{cr}=3–4 H-bonds (see Figure 1-7c for an illustration of this concept). Notably, following this scaling law, most protein materials are indeed found to intersperse many small β-sheets close in size to N_{cr}, in order to turn the weakness of the building block (H-bond) into strength by using geometry as a paradigm to achieve this goal. For example, β-sheets show a typical H-bond cluster size of 3–8 H-bonds across thousands of proteins in the Protein Data Bank, and α-helices have 3.6 H-bonds per turn [72, 73, 74]. It is interesting to note that the scaling law of the shear strength, increasing for small values of H-bonds to a maximum value followed by a decay represents a biological material analogue of the Hall-Petch relationship known from polycrystalline engineering materials (*e.g.* metals) [75]. The intrinsic strength limit can only be overcome and scaled up by using structural hierarchies that circumvent β-sheet saturation and other limiting factors at critical length scales.

Silk is a prime example of how H-bond cooperativity within β-sheets of controlled lengths and a hierarchical structure of networks and fibrils can overcome molecular strength limits and create unusually strong and tough protein fibers at the macroscale. Understanding the behavior of the cross-linking nodes within the fibril network at the molecular scale is critical for potentially bypassing strength limits at this length scale

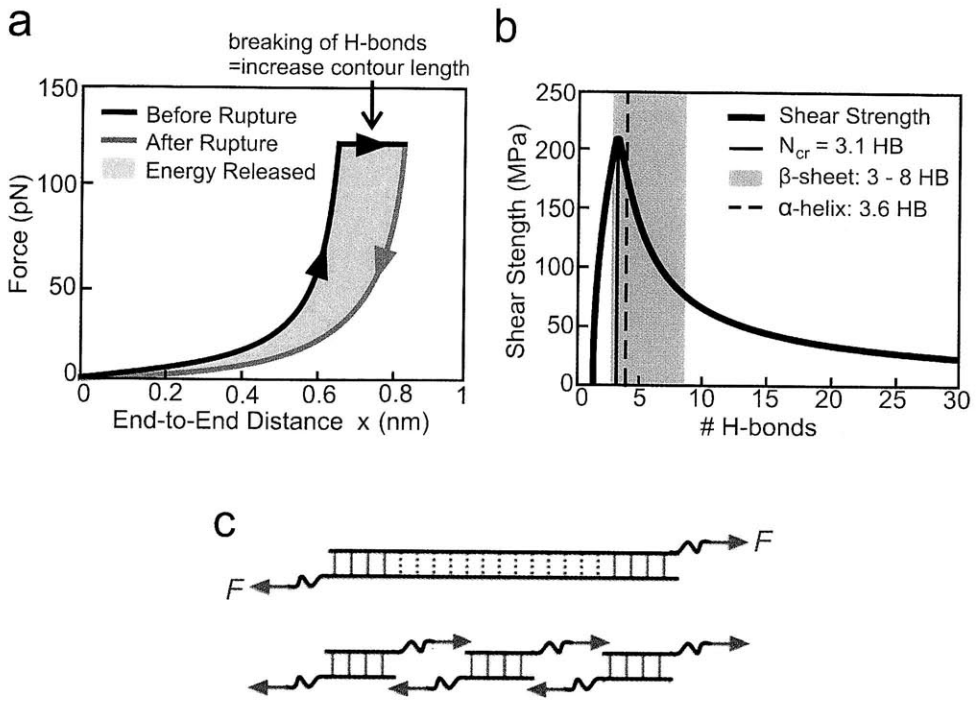

Figure 1-7: Mechanistic view into the H-bond rupture process and scaling of shear strength over size of H-bond cluster. (a) The area enclosed between loading and unloading curves is equivalent to the change in free energy, *i.e.* the adhesion energy per unit length. (b) A β-sheet of 3–4 residues in length displays the theoretical maximum shear strength, meaning that up to 3–4 H-bonds can rupture simultaneously. (c), Illustration of the physical concept behind the scaling law shown in panel (b), where the breaking of H-bond clusters is confined to a critical cluster size of N_{cr}=3–4. Figures adapted from [74].

and vastly improving silk for medical and textile purposes as well as synthetic elastomers and polymer or aramid fiber composites with a similar molecular structure and noncovalent bonding for aerospace, armor, and medical applications. This work constitutes the most comprehensive study to-date of the molecular structure prediction and nanomechanical behavior of dragline silk. While other computational studies have used similar methods for structure prediction and mechanical analysis, *e.g.* REMD and force-control loading [35, 20] or more course methods [37], this work presents:

- the first results of the near-native structures determined by REMD after equilibration in TIP3P explicit solvent,

- the first parametric study of the effects of modifying the wildtype poly-Ala segment length to values outside the range naturally observed for MaSp on structure prediction and nanomechanical behavior, and
- the first comparison between previously published loading conditions, *i.e.* the Stretch test, and the novel Pull-out loading conditions that are hypothesized to be more appropriate for modeling of the *in situ* loading of the cross-linking β-sheet crystal.

Chapter 2

Materials and Methods

This chapter describes the *in silico* setup and analysis of the test cases considered in this study. First, Section 2.1 describes the derivation of the initial lattice structure, and Section 2.2 explains the conditions of the replica exchange molecular dynamics (REMD) simulations used to create a test case ensemble from the initial lattice. Next, Section 2.3 describes the conditions of equilibration of the test cases in explicit solvent, and Section 2.4 reports the methods of secondary structure analysis used to compare solvent conditions. Finally, the loading conditions of the nanomechanical testing simulations are described in Section 2.5. A summary of the methods used in this study is presented in Figure 2-1.

2.1 Initial structure

Spider silk filaments are composed of folded and cross-linked polypeptide oligomers with distinct repeating sequence units: the poly-Ala $(A)_n$ repeat unit and the Gly-rich $(GGX)_n$ repeat unit. Previous computational studies have shown that the poly-Ala unit in the wildtype MaSp1 repeat unit forms a β-sheet nanocrystal [19]. The test cases treated in this study focus on a partial oligomer composed of a single poly-Ala repeat with one instance of the Gly-rich semi-amorphous repeat unit on each side. The wildtype *N. clavipes* MaSp1 partial peptide sequence used in this study, in one-letter amino acid code, is:

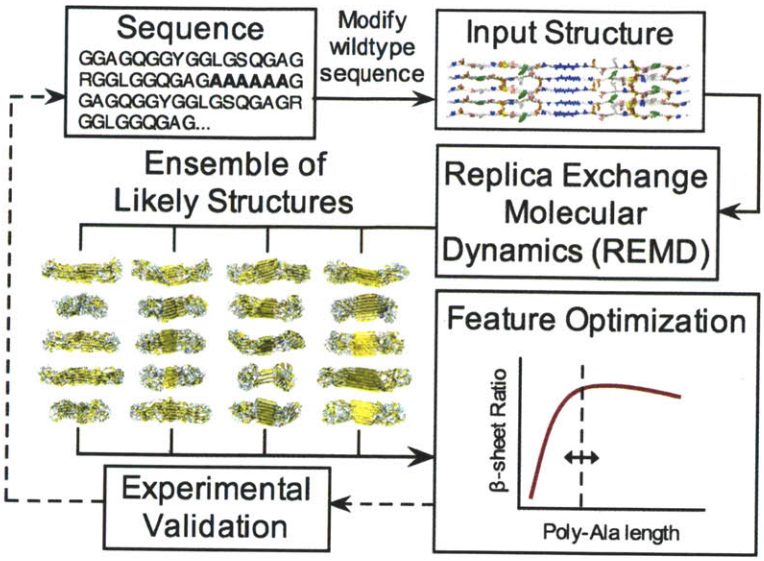

Figure 2-1: Feature optimization through the creation and analysis of a test ensemble. Replica exchange simulations [35] are performed on a lattice of aligned strands of the *N. clavipes* MaSp1 peptide sequence with multiple cases of poly-Ala repeat length, including wildtype, and sufficient 10 Å spacing between strands to avoid initial side-chain interactions. The method of k-means clustering of the 300 K replica timeline determines the most probable native structures for each case. Analysis of the secondary structure and dihedral angles determines the cases of poly-Ala length that result in the most defined β-sheet nanocrystals. Once the structure of nanocrystals is optimized, further modifications to the sequence can be explored to optimize other features using this general approach.

Residue IDs	1–27	28–33	34–60
Amino Acid	GAGQGGYGGLGSQGAGRGGLGGQGAG	AAAAAA	GGAGQGGYGGLGSQGAGRGGLGGQGAG

The wildtype poly-Ala length of six residues, 28–33 above, is systematically varied among the test cases in order to study the effect of the unit's length on β-sheet distribution. A chart of amino acid abbreviations, classifications, and side-chains are shown in Appendix A.5 for reference. REMD is then used to create an ensemble of near-native, energy-minimized test case structures from an initial aligned lattice of partial MaSp1 strands. Experimental studies of recombinant silk [22] suggest that mechanical shear within a narrowing elongational flow extends and aligns MaSp oligomers during spinning, and that this encourages alanine amyloidization into extended nanoscale β-sheet crystals that cross-link a semi-amorphous filament network (Figure 2-2).

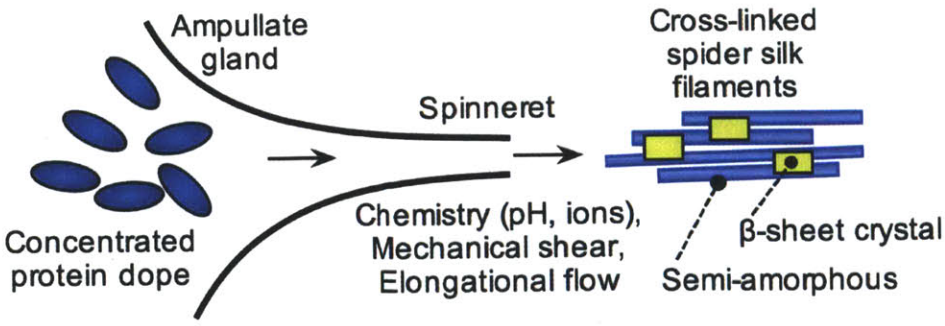

Figure 2-2: The REMD input structure is inspired by the natural silk spinning process. A combination of chemistry and shear flow transform the concentrated protein dope into a filament network cross-linked by β-sheet crystals. Figure based on insight gained from experimental work [22].

Mimicking this process, an extended conformation is used when creating a single strand of the MaSp1 partial sequence unit using the TINKER Molecular Modeling Package with CHARMM-19 topology. A rectangular lattice structure (Figure 2-3) is created by arranging three parallel layers, where each layer is made of five MaSp1 strands in an anti-parallel arrangement in the side-chain direction. In this initial lattice structure, the backbones of the strands are at a minimum distance of 10 Å to avoid side-chain interference and limit interaction. While this simulation protocol is somewhat limited, as it does not account for all processes and reactions occurring during silk spinning [76], the extended starting configuration mimics the elongational flow of the concentrated dope in the spinning duct and encourages alanine aggregation leading to crystal formation rather than the folding of each strand [22].

2.2 Replica exchange molecular dynamics (REMD)

To create test structures from the initial lattice structure, we simulate the lattice with Langevin dynamics in the CHARMM molecular modeling program using the CHARMM-19 force field and the EEF1.1 implicit solvation force field. The implicit solvation model allows a simulation time step of 2 fs by employing the SHAKE algorithm for hydrogen atoms. Solvent friction is added via a Langevin friction term that allows for high mobility and conformational sampling. While the EEF1.1 model has

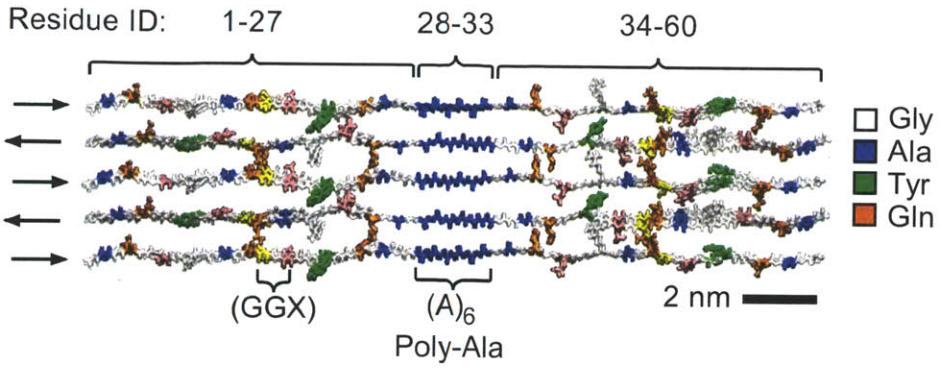

Figure 2-3: The extended lattice of repeat unit strands input into the REMD simulations mimic the elongational flow conditions within the spinneret. An initial 1 nm minimum distance between strands prevents side-chain interference and noncovalent bonding.

particular modifications and simplifications of solvent, side-chain, and hydrogen bond interactions, it is orders of magnitude faster than other implicit or explicit solvent models, making it ideal for preliminary simulations of the large-scale silk assembly processes. Since force fields are generally parameterized for room temperature calculations, only final ensemble structures from the lowest temperature replica (*i.e.* 300 K) are picked, while higher temperature replicas are used for fast conformational search and overcoming kinetic trapping in the REMD scheme. The REMD protocol is set up and performed using the MMTSB Toolset. Long initialization runs of the extended lattice structure are performed to obtain multiple distinct starting configurations to enhance sampling in the production run. This is followed by a production run starting from the final configurations of the replicas from the initialization run using an exchange time step of 2 ps to allow for relaxation of the system. The production run simulates 64 replicas distributed evenly over a temperature range of 300 to 650 K. Refer to Appendix A.1 for an example of arguments within the REMD submission script and to Figure 2-5 for a schematic of the REMD protocol. Each replica is simulated for a total of 10 ns, corresponding to a total effective simulation time of 640 ns for the peptide sequence. To simulate a protein structure 10–14 nanometers in length for 10–640 ns, non-reactive molecular dynamics force fields such as CHARMM are necessary for a convenient user timeframe, as shown in Figure 2-4. While non-reactive force fields cannot capture covalent bond chemistry, they can model noncovalent bonding,

electrostatics, and solvent diffusion sufficiently for protein simulation.

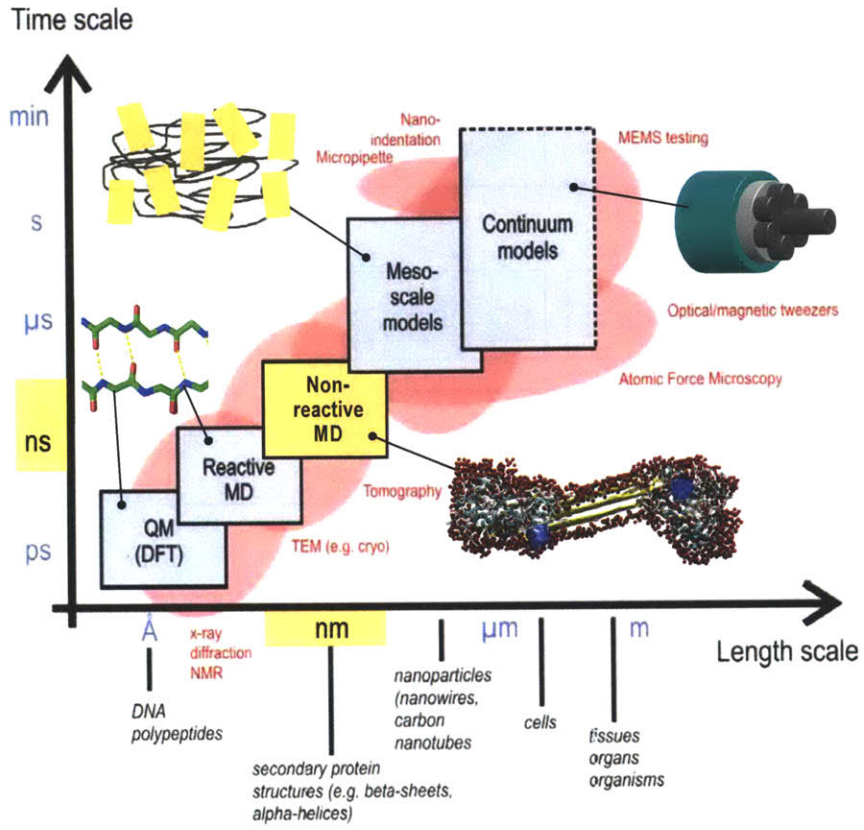

Figure 2-4: While computational modeling can simulate materials at several orders of magnitude in length and time scales, different computational tools are appropriate for a finite scale dictated by the scale and goals of the modeling. Protein simulations, such as those in this study, typically use non-reactive molecular dynamics to cover the nanoscale, *i.e.* simulate proteins of several nanometers in length for several nanoseconds, in order to capture secondary structure changes and solvent-mediated mechanisms. Figure reprinted from [77].

REMD simulations periodically exchange the structures of certain replicas in order to to explore a large conformational space. If the conformational potential energy of a high temperature replica is lower than that of the replica one the temperature step below, the conformational structures of the replicas may exchange. The probability of an exchange is governed by a Metropolis-Hastings algorithm. Therefore, replicas must be close enough in temperature for the conformational potential energy functions to overlap and allow exchange. Here, exchange events are scheduled every 0.5 ps. The

EEF1.1 implicit solvation model becomes necessary for simulating each of 64 replicas with 900 residues for 10 ns within a convenient user timeframe.

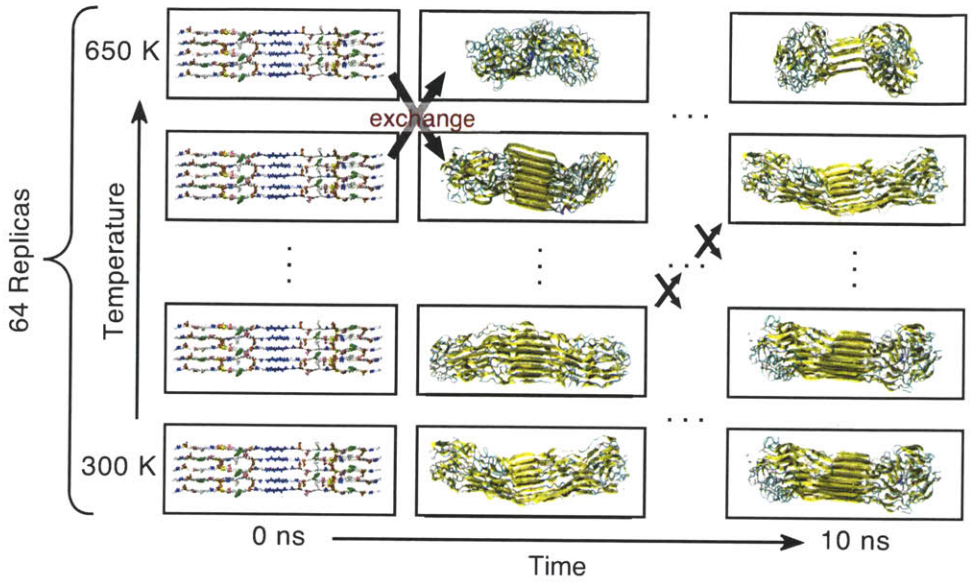

Figure 2-5: Schematic of the REMD protocol. 64 replicas of the initial input structure are simulated at different temperatures from 300 K to 650 K. Each replica is simulated for 10 ns, corresponding to a total effective simulation time of 640 ns for the protein assembly.

The k-means clustering algorithm from the MMTSB Toolset is performed on the last 1 ns of the 300 K replica timeline. In k-means clustering, n observations (here, discrete conformational structures during simulation) are partitioned into k sets or clusters according to the distance from a cluster's mean or center. For proteins, this distance from a cluster's center is represented by the root mean square deviation (RMSD) of each observed structure from the structure of the closest cluster center. RMSD is computed by averaging the distance δ of the alpha-carbon (Cα) of each of N residues from the corresponding Cα of the cluster center structure:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_{i,\mathrm{C}\alpha}^{2}} \tag{2.1}$$

Due to the nature of replica exchanges, structures with lower potential energies are exchanged less often and are predictions of more near-native structures than others. The cluster centers therefore represent model structures to which others are

compared. The centers of the five largest clusters are exported for analysis as the test structures for the wildtype case of *N. clavipes* with 6 alanine residues in the poly-alanine region of the MaSp1 repeat unit. The creation of the initial lattice, REMD, and k-means clustering is performed again for cases of 2, 4, and 12 alanine residues in the poly-alanine region. The twenty resulting structures constitute the test case ensemble (see Figure 3-1). The relative cluster sizes (*i.e.* time representation in the 300 K replica timeline) is used to weigh property averages.

2.3 Explicit solvation equilibration

To obtain more realistic molecular conformation and tertiary protein structure, the "principal" structure, *i.e.* that from the largest cluster, for each test case is equilibrated for 20 ns in a wrapping periodic waterbox of TIP3P explicit solvent using NAMD with the CHARMM-22 force field. Refer to Appendix A.2 for an example of the input file. Studies of structural change during this equilibration show that 20 ns provides sufficient convergence of both the root mean square deviation (RMSD) and the solvent-accessible surface area. To prevent image interactions when using periodic boundary conditions, the waterbox pads the protein by at least 10 A. Equilibration is performed with Langevin dynamics at 300 K and with Particle Mesh Ewald (PME) electrostatics to more accurately capture solvent interactions. Table 2.1 illustrates the additional cost of simulations with explicit solvent. In particular, the addition of TIP3P water molecules increases the total atom count N ten-fold. Since simulation time increases exponentially at least as N^3, only the principal structures are simulated with explicit solvent in order to maintain a convenient time-frame for the user.

Table 2.1: Atom count N for each studied poly-Ala length and solvation condition.

Poly-Ala	Imp.	Exp.	Ratio
2-Ala	5 955	52 149	8.76
4-Ala	6 135	56 295	9.18
6-Ala	6 315	63 510	10.06
12-Ala	6 855	64 185	9.36

2.4 Analysis methods of structure predictions

The secondary structures of each structure is determined using the STRIDE algorithm built into the VMD Molecular Graphics Viewer [78] and customized .tcl scripts. The STRIDE algorithm holds an advantage over DSSP and other secondary structural algorithms by employing pattern recognition of statistically-derived backbone dihedral angle information [79]. Using the secondary structure results, Ramachandran density plots are created for the principal (that is, the most represented) cluster center using a 10 bin size. Propensity for certain secondary structures along each strand are predicted by analyzing the peptide sequence for each test case in the Protein Plot window in MATLAB for local Hydrophobicity and Total β-strand properties. The definition of hydrophobicity as defined by Kyte & Doolittle [80] is based on an index of relative hydropathicity ranging from -4.5 to +4.5. The value for an individual amino acid is a weighted average of the normalized transfer free energy from water to vapor, the fraction of side-chains found 100% buried in a sample of nearly 1300 experimentally studied proteins, and the fraction of side-chains found 95% buried in the same sample. Amino acids with a negative hydropathicity are considered hydrophilic, those with a positive hydropathicity are hydrophobic, and those near zero are ambivalent. For an entered peptide sequence, the hydropathicity of an individual residue is a weighted average of the hydropathicities of adjacent residues within a sliding window. Thus, a hydrophobic segment reinforces its own hydrophobicity. The Total β-strand preference [81] is a relative index ranging between 0 and 2 based on strand-strand interactions instead of the dihedral angles of a single residue. Index values for each natural amino acid were determined for antiparallel and parallel β-sheets in an experimental sampling of 30 proteins. In order to emphasize the predictions of both the Hydropathicity index and the TBP index on β-sheet distribution within the MaSp1 repeat sequence, we multiply the values of both indices at each residue along the sequence.

2.5 Nanomechanical testing

Nanomechanical testing was performed on MaSp1 test cases with poly-Ala lengths of 2-, 6-, and 12-Ala with two distinct solvent conditions, *i.e.* EEF1.1 implicit solvent and TIP3P explicit solvent; and two force-control loading conditions, *i.e.* "Stretch" and "Pull-out" tests. Due to the computational cost of simulations with explicit solvent, testing was only performed on equilibrated principal structures and not on the entire test ensemble. The solvation and loading conditions are described in more detail below.

2.5.1 Stretch Test

The Stretch test loads the β-sheet crystal in multi-lap shear and is based on previous computational studies of the the wildtype 6-Ala case [35]. Figure 2-6 shows a schematic of the Stretch test loading conditions: 14 strands are loaded in opposite directions for multi-lap shear. To maintain a null net force, an even number of strands are loaded, and the remaining 15^{th} strand is left free.

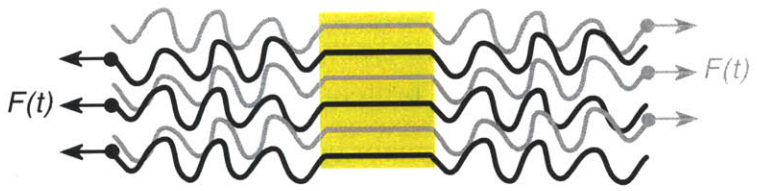

Figure 2-6: Stretch test loading conditions. 14 strands are loaded in opposite directions for multi-lap shear. To balance forces, the 15^{th} strand is free.

Implicit Solvent

The Stretch test is performed on the five largest cluster center structures of the REMD simulation run (Section 2.2) using the CHARMM molecular modeling program with the CHARMM-19 and EEF1.1 implicit solvent force fields. Each simulation is performed at 300 K with a Nosé-Hoover themostat. The force *F(t)* on each strand terminus is increased stepwise by 2 pN every 20 ps of equilibration. Refer to Appendix

A.3 for an example of the input file. The simulation is complete when the protein has failed in shear, *i.e.* when one or more strands have separated from the main body.

Explicit Solvent

The Stretch test is also performed on the principal structure, *i.e.* that of the largest REMD k-means cluster center, after equilibration in explicit solvent for 20 ns (Section 2.3) using the NAMD molecular modeling program using the CHARMM-22 force field. Each simulation is performed at 300 K. To prevent image interaction after large deformation, the simulation box, with fully wrapping periodic boundary conditions, is kept larger than the expected deformed size and pressure control is turned off. While some water molecules become gaseous during the simulation, most maintain a liquid state surrounding the protein. Particle Mesh Ewald (PME) electrostatics are employed to better capture solvent effects and changes in solvent-mediated changes in secondary structure. The force *F(t)* on each strand termini is increased by a step of 20 pN every 20 ps of equilibration. The loading rate must be ten times faster than that of the implicit solvent simulations in order to maintain a convenient user timeframe. Refer to Appendix A.4 for an example of the input file and all arguments. The simulation is complete when the protein has failed in shear, *i.e.* when one or more strands have separated from the main body.

2.5.2 Pull-out Test

The Pull-out test loads the β-sheet crystal in double shear and is chosen to better represent the probable loading conditions of a fringed micelle poly-crystalline material. While the Pull-out test assumes that alternating strands will be loaded equally, the Pull-out test treats the strands as if they divert and connect to separate crystals further in the fibrillar network. Figure 2-7 shows a schematic of the Pull-out test loading conditions: The termini of 13 strands are fixed, while the temrini of the remaining 2 adjacent strands are loaded.

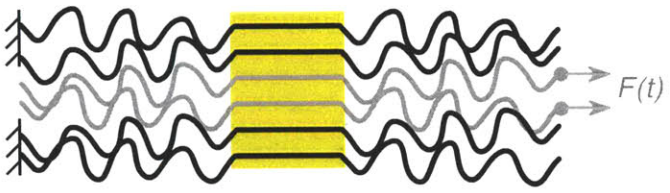

Figure 2-7: Pull-out test loading conditions. The termini of 13 strands are fixed, while the remaining 2 adjacent strands are loaded.

Implicit Solvent

The Pull-out test is performed on the five largest cluster center structures of the REMD simulation run (Section 2.2) using the CHARMM molecular modeling program with the CHARMM-19 and EEF1.1 implicit solvent force fields. Each simulation is performed at 300 K with a Nosé-Hoover themostat. The force *F(t)* on each strand terminus is increased stepwise by 2 pN every 20 ps of equilibration. The simulation is complete when the protein has failed in shear, *i.e.* when one or more strands have separated from the main body.

Explicit Solvent

The Pull-out test is also performed on the principal structure, *i.e.* that of the largest REMD k-means cluster center, after equilibration in explicit solvent for 20 ns (Section 2.3) using the NAMD molecular modeling program using the CHARMM-22 force field. Each simulation is performed at 300 K. To prevent image interaction after large deformation, the simulation box, with fully wrapping periodic boundary conditions, is kept larger than the expected deformed size and pressure control is turned off. While some water molecules become gaseous during the simulation, most maintain a liquid state surrounding the protein. Particle Mesh Ewald (PME) electrostatics are employed to better capture solvent effects and changes in solvent-mediated changes in secondary structure. The force *F(t)* on each strand termini is increased by a step of 20 pN every 20 ps of equilibration. The loading rate must be ten times faster than that of the implicit solvent simulations in order to maintain a convenient user timeframe.The simulation is complete when the protein has failed in shear, *i.e.* when

one or more strands have separated from the main body.

Chapter 3

Sequence-Structure Correlations in MaSp1 Protein

This chapter presents structural predictions as a function of poly-Ala repeat length variation as well as implicit and explicit water solvation during equilibration for a direct comparison between these two approaches. We focus primarily on dihedral angles and secondary structure of the predicted structures shown in Figure 3-1. Since the majority of MaSp1 (and silk in general) comprises glycine and alanine amino acids, analysis directly compares the dihedral ϕ-ψ angles of the glycine and alanine groups separately for the test cases with experimental data on spider silk proteins.

3.1 Secondary Structure

The secondary structure distributions are presented visually in Figure 3-1. Glycine residues (Figure 3-2a) in implicit solvent show symmetry about the origin and a wide distribution about a peak at (-75°, +75°) for all cases of poly-Ala length. Glycine in explicit solvent shows very different peaks at (±90, 0) and (±90°, 180°). The peaks with explicit solvent are in better agreement with experimental findings [17] as well as allowed Ramachandran regions. Alanine residues (Figure 3-2b) in implicit solvent show a single peak at (-90°, +75°) for the 2-Ala principal structure. As the poly-Ala length is increased, the peak shifts to around (-140°, +140°). Alanine in explicit

solvent show a similar progression towards (-140°,+150°), which corresponds to a predominantly anti-parallel β-sheet conformation for the 6-Ala and 12-Ala principal structures and is in excellent agreement with experimental findings with peaks at (-135°, +150°) [17]. Note that as the number of simulated amino acids increases, the peaks soften, seen as a greater number of contour lines on the density plots. While the implicit and explicit water solvation models make simplifications for side-chain and hydrogen bond interactions in favor of computational efficiency, our secondary structure and dihedral angle analysis agree well with experimental findings.

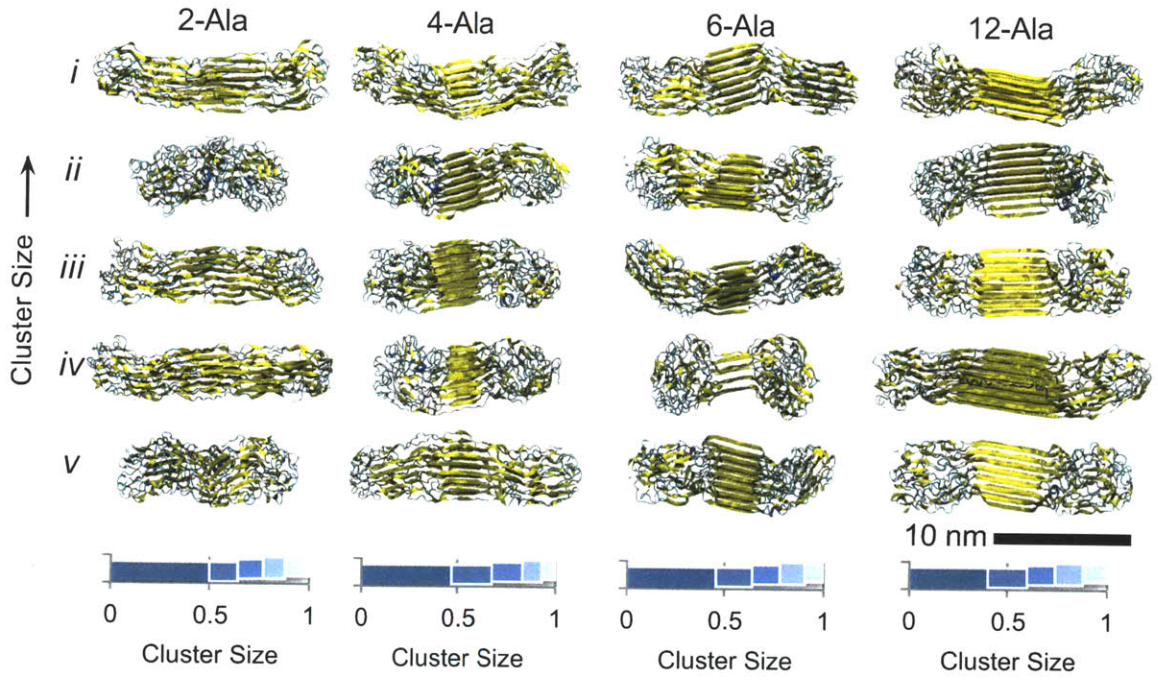

Figure 3-1: The test case ensemble after REMD with implicit solvation. The test models, in a cartoon representation colored by secondary structure, shows the five largest k-means cluster centers after REMD. It can already be seen visually that the 4-Ala case represents a critical length for poly-Ala crystal formation. Relative k-means cluster sizes are shown underneath for each test case. The largest clusters, *i.e.* the principal*i*structures, are most likely to represent native structures for each test case of poly-Ala length.

3.1.1 Distribution of β-sheets

We now focus on relative secondary structure ratios and the spatial distribution of β-sheet conformations for each test case considered here, visualized with cartoon

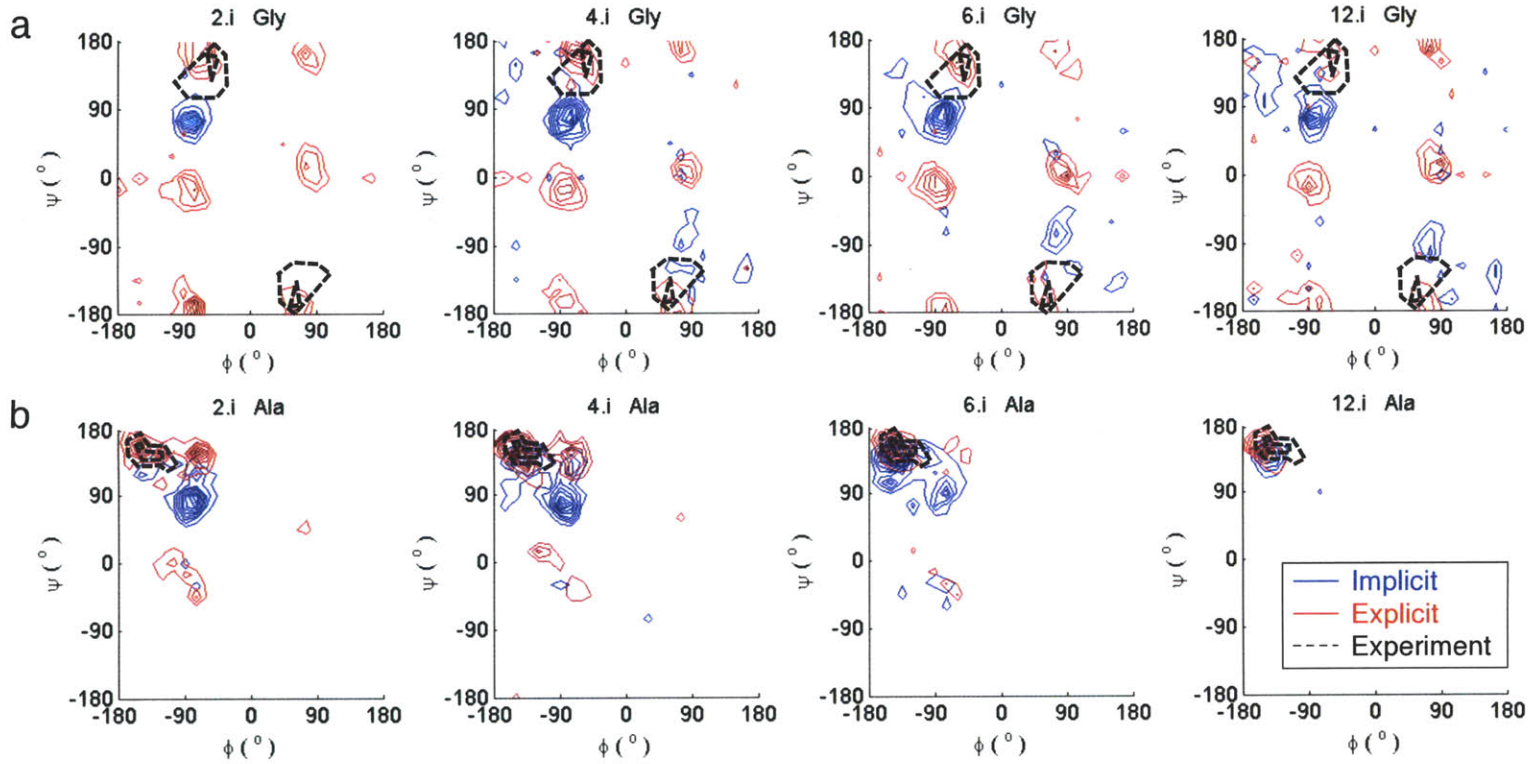

Figure 3-2: Experimental validation of the *in silico* silk structure predictions based on implicit and explicit water models. Ramachandran density plots of the glycine and alanine in the principal *i*structures. Blue shows conformation with implicit solvent, red with explicit solvent, and dotted black shows experimental peaks [17]. (a) For glycine, all test cases have similar peak locations and distribution for each solvation. As the system gets larger, *i.e.* longer poly-Ala repeat, the glycine peaks are less defined. (b) As poly-Ala length increases, the alanine peak shifts toward (-140°, +140°), very close to what is observed experimentally. For both glycine and alanine, explicit solvent results in peaks closer to both allowed Ramachandran regions and to experimental peaks. This result implies that the wildtype 6-Ala case is the minimum poly-Ala length that forms the anti-parallel β-sheet nanocrystals observed in in vivo dragline silk. It also demonstrates the viability of REMD in determining native protein structures, but only when coupled with explicit solvent equilibration.

representation in Figure 3-3a. With implicit solvent, we find an average, weighted according to relative cluster size, of approximately 50% β-sheet, 30% turn, and 20% random coil conformation for each test case of poly-Ala length. No helix conformation is reported by the STRIDE algorithm definition. Although the weighted average ratios do not vary more than 5% among the test cases, the distribution of discrete ratios decreases as the poly-Ala repeat length increases. The most consistent (*i.e.* tightest distribution) of β-sheet and turn conformations are observed for the wildtype 6-Ala test case. Consistency among conformational structures for each test case signals an implicit stability of the respective poly-Ala length.

After equilibration of the principal structures in explicit solvent, the semi-amorphous region shows an average of only 10% β-sheet for all test cases. The spatial distribution of residues with β-sheet conformation differs according to the poly-Ala repeat length. By plotting the occurrence of β-sheet for each residue ID among the fifteen chains of the test lattice (Figure 3-3b), we directly observe β-sheet grouping for each test case. The 2-Ala test case in explicit solvent shows a soft peak in β-sheet occurrence centered at the poly-Ala region. The 46% peak occurrence implies inconsistent and weak crystal formation. For the 4-, 6-, and 12-Ala test cases, 90–100% of chains have β-sheet conformation in the poly-Ala region, in sharp contrast to an average 10% occurrence outside this region. This shows very strong crystal definition for all cases with poly-Ala repeat length of 4 alanines or longer. Changes in total β-sheet contents after equilibration are shown in Figure 3-4.

The spatial distribution of residues with β-sheet conformation in explicit solvent (Figure 3-5a) agrees well with the Hydropathicity and Total β-sheet Preference (TBP) predictions in Figure 3-5b. Crystal definition and consistency is made more readily apparent through direct observation of β-sheet occurrence among only the poly-Ala residues (Figure 3-5c), again with averages weighted by relative cluster size for implicit solvent test cases. Explicit solvent cases refer only to principal structures. For both solvation models, the 2-Ala test group averages 45-65% β-sheet conformation in the poly-Ala region, with a very wide distribution of discrete points. On the other hand, the 4-, 6-, and 12-Ala test groups average more than 90% β-sheet conformation for

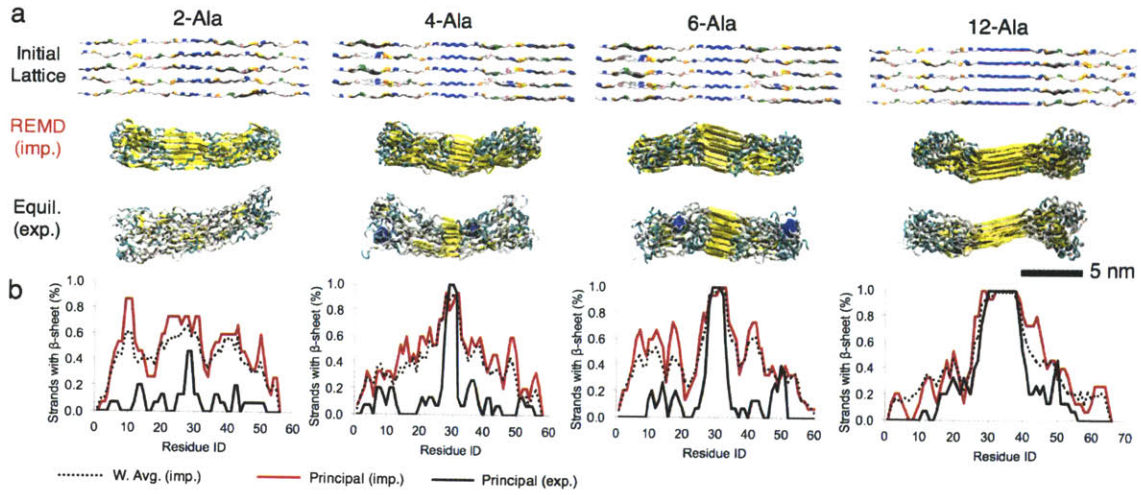

Figure 3-3: Analysis of the β-sheet content, which depends heavily on solvent choice, for implicit and explicit water models. (a) The initial lattice structure before REMD shows the location of the poly-Ala in blue. This also illustrates the hidden length afforded by the semi-amorphous regions after REMD and equilibration. (b) Crystal definition is inferred by the percent of strands with β-strand conformation at each residue for the principal structure for each test case. These results confirm that a critical minimum of 4 alanines is needed for defined β-sheet nanocrystal formation (*i.e.* a full-height peak). Solvent effects are also clearly observable; after diffusion of explicit solvent, most β-sheets in the semi-amorphous regions are disrupted, while the hydrophobic poly-Ala crystal remains defined.

both solvation models. The wildtype 6-Ala test case shows the highest average β-sheet content and smallest distribution. The 12-Ala test case is almost as defined, but shows that larger crystals may be degraded by detrimental effects at this length scale, such as hydrogen bond saturation.

3.2 Discussion and Conclusion

We presented results from atomistic REMD simulations on MaSp1 protein segments of the dragline spider silk of *N. clavipes*. We have illustrated that critical conditions for particular secondary structure formation can be found through systematic variation of the peptide sequence alone, and that the length of the poly-Ala repeat unit is critical in defining identifiable stable β-sheet nanocrystals. Specifically, there exists a strong scaling effect where a minimum length of poly-Ala repeats is found. Averaging

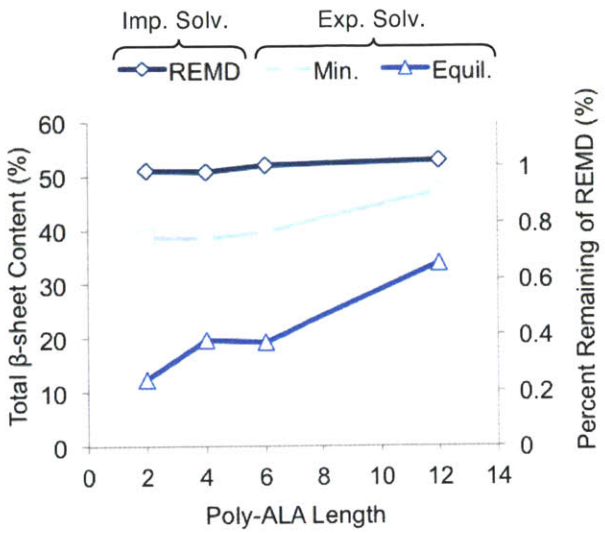

Figure 3-4: Total β-sheet content of the principal structures after REMD with implicit solvent and minimization and equilibration with explicit solvent. REMD predicts a β-sheet content between 51–53% regardless of poly-Ala length. Equilibration with explicit solvent and PME electrostatics predicts a much lower β-sheet content, between 12–34%, that depends heavily on the poly-Ala length due to the refined clarity of the central crystal.

over all probable structures of each test case shows a minimum poly-Ala length of at least 4 alanine residues for consistent crystal formation. However, this assumes perfect alignment of alanine side-chains before agglomeration. During natural spinning, a minimum of 6 alanines is more realistic for the formation of robust beta-sheet nanocrystals, and this is indeed the wildtype poly-Ala length for *N. clavipes*. Other species that produce MaSp1 and MaSp2 proteins feature 8-Ala, seen in Figure 3-6. While a longer poly-Ala region also results in consistently defined crystals, the generation of poly-Ala regions longer than 8 alanines may be too energetically expensive and thus prohibitive during the evolution of this species and without any further mechanical payoff as suggested in earlier work [35].

3.2.1 Geometric Interference of Side-chains

The observation of a minimum poly-Ala length is explained by a simple biophysical model. Alanine side-chains (*i.e.* nonpolar methyl groups) alternate sides of the

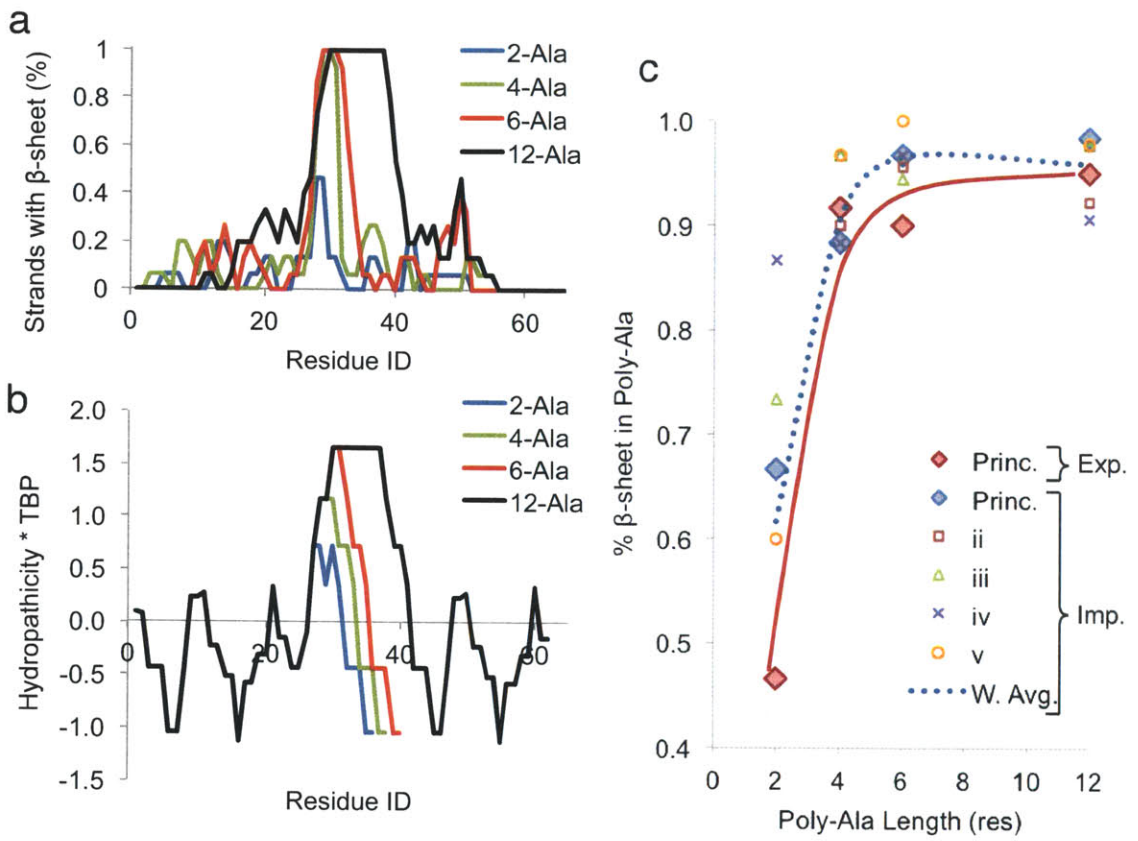

Figure 3-5: The β-sheet content within only the poly-Ala region for each test case, for implicit and explicit water models. (a) When the Hydropathicity and Total β-sheet Preference indices are multiplied, it is observed that poly-Ala repeats of 6-Ala and longer reinforce their own hydrophobicity and β-sheet preference. Only the values for the 12-Ala case are shown past the poly-Ala region for clarity; the values of the other cases are identical beyond their cutoffs. (b) β-sheet distribution and crystal definition after equilibration with explicit solvent shows very similar peaks along the peptide sequence, particularly in the poly-Ala repeat and at (GGL) segments around residues 20 and 50. (c) The 2-Ala cases show a wide distribution, indicating poor crystal clarity. 4-Ala cases begin to show defined crystals, while wildtype 6-Ala cases show the most defined crystals. 12-Ala cases also show well-defined crystals, but longer alanine repeats in the MaSp1 sequence may be more biologically expensive and thus prohibitive. This result implies that the evolutionary process may have optimized the MaSp1 protein per the environmental conditions of *N. clavipes* for the most consistent crystal formation while consuming the least energy and food resources.

backbone along a β-strand. This alignment and the small size of the alanine side-chain allow the side-chains of adjacent β-sheets to zip together, in turn allowing aligned poly-Ala β-sheets to closely stack out-of-plane. Geometric interference of

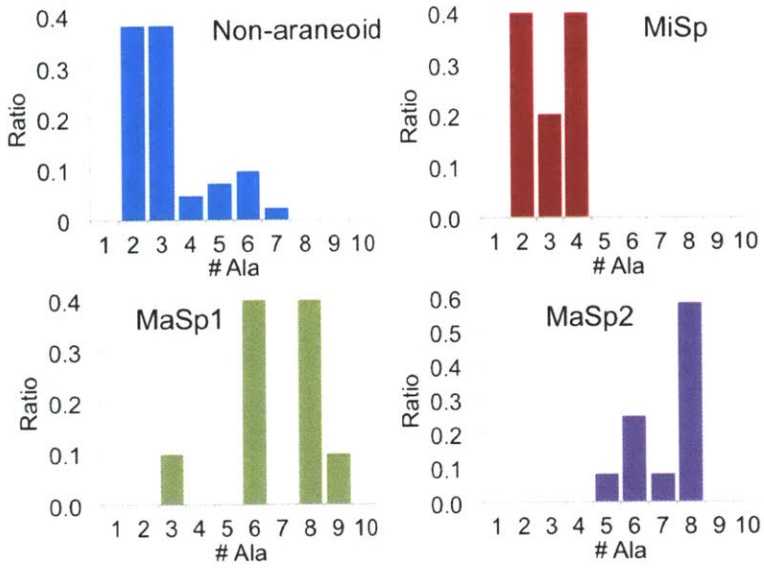

Figure 3-6: Poly-Ala length distribution among spider silk proteins across several species. Non-araneoid and MiSp (*e.g.* capture) silk feature short repeats 2-4 Ala in length. Major ampuallate (*e.g.* dragline) silk, the toughest and strongest silk, features longer 6-Ala or 8-Ala repeats, allowing larger cross-linking β-sheet crystals. Ratios are normalized to unity for each category. Values taken from the consensus sequences listed in [8].

the side-chains after stacking provides additional bending and torsional rigidity to the multi-layer crystal. In addition, the hydrophobic and nonpolar nature of alanine reinforces crystal stability by preventing water diffusion between the stacked β-sheets. Indeed, after equilibration of the principal structures in explicit solvent, water was found to have diffused within the semi-amorphous region, but no water was observed within the β-sheet crystal, illustrated in Figure 3-7.

Side-chain zipping also offers resistance to peeling, as deformed β-strand backbones force side-chains to mechanically clamp onto other adjacent side-chains, as shown in Figure 3-8. The 2-Ala β-strands are not long enough to zip together and are easily cleaved by water or peeled by boundary conditions of the less-dense semi-amorphous region. This explains the low β-sheet content of the poly-Ala region of the 2-Ala test case. Also, 3-Ala or 4-Ala β-strands must be highly aligned to allow side-chain zipping, but can resist cleavage in this case. The 6-Ala (wildtype) case may be misaligned during amyloidization and still result in some side-chains zipping. This margin of error may be structurally worth the energetic cost of additional poly-Ala

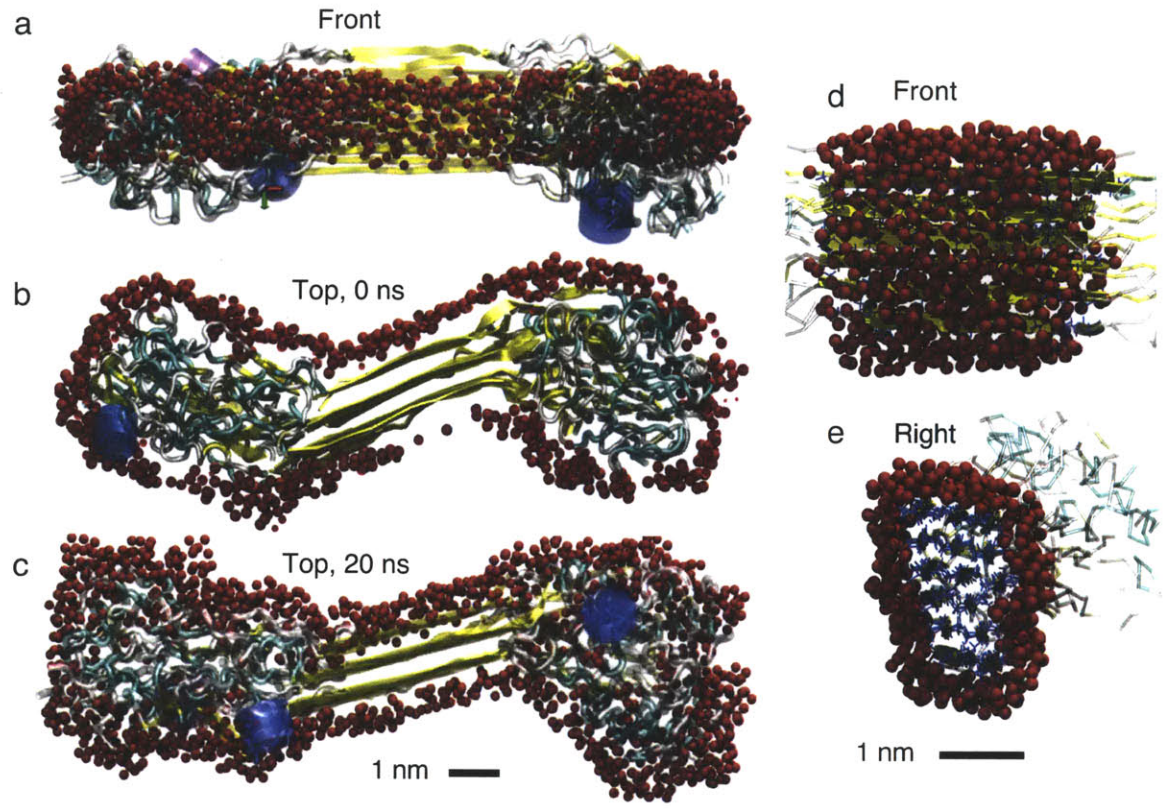

Figure 3-7: Diffusion of the explicit water molecules (red spheres) after equilibration of the 12-Ala case. Only water within 5 Å of the protein is shown for clarity. (a) By hiding the water on each side of the protein, the diffusion of water within the protein can be better visualized (b) before and (c) after 20 ns equilibration, after which water molecules are found within the semi-amorphous regions but not within the central crystal. (d) By focusing only on water near the crystal, (d) it is seen that the efficient packing of the small alanine side-chains (blue sticks) prevents diffusion of solvent into the crystal.

synthesis.

The simulation results of poly-Ala length test cases in explicit solvent illustrate these trends in the β-sheet nanocrystal stability in Figure 3-9. For the 2-Ala test case, a single β-sheet of only three strands is found in the interior of the protein. Another β-sheet is not present for stacking in the side-chain direction. In contrast, the 4-Ala test case shows a very aligned two-layer crystal that is open to exterior water. The 6-Ala case shows a similar crystal and also illustrates the tolerance of misalignment. In the center of the crystal, several β-strands are misaligned by two residues. However, the core of the crystal remains 4-Ala in width and thus remains

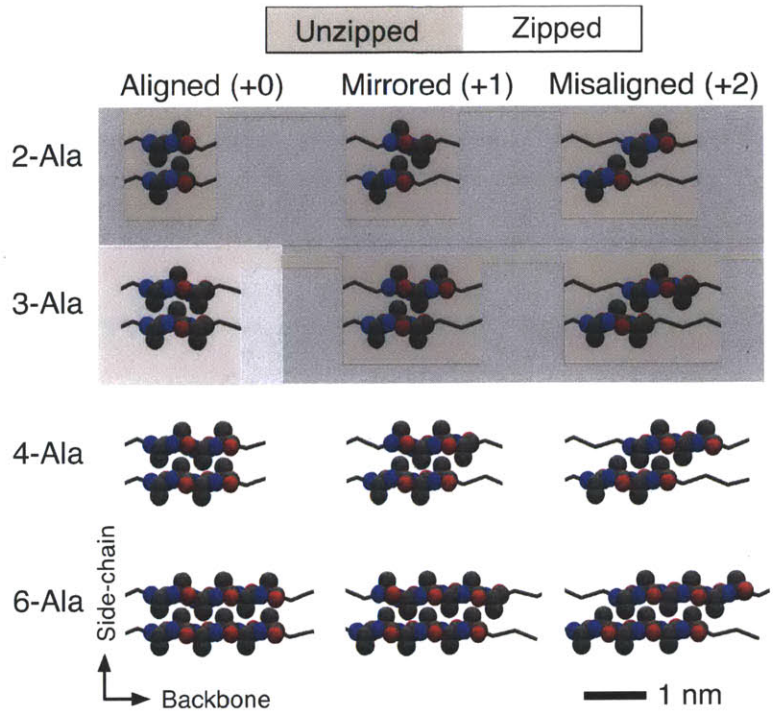

Figure 3-8: Efficient stacking of alanine side-chains determines a critical poly-Ala length for crystal stability. Ala side-chains alternate direction along a β-strand, allowing poly-Ala β-sheets to stack out-of-plane to reinforce hydrophobicity and provide rigidity through geometric interference. Color key: C (black), O (red), N (blue). H is hidden for clarity.

stable in the presence of nearby water molecules. The 12-Ala test case, omitted from Figure 3-9, shows highly ordered stacking and is in some places three layers thick. Therefore, MD simulation of MaSp1 poly-Ala amyloidization demonstrates that 4-Ala repeats are sufficient for the formation of β-sheet nanocrystals in the final silk. However, repeats of 6-Ala (wildtype) or longer allow misalignment of the poly-Ala during amyloidization.

The most important findings of this study is that the critical conditions for particular secondary, tertiary, and quartenary structure formation, in particular of stable β-sheet nanocrystals, can be found through systematic variation of the peptide sequence alone, and that the length of the poly-Ala repeat unit is critical in defining identifiable β-sheet nanocrystals. Specifically, there exists a strong scaling effect where a minimum length of poly-Ala repeats is required. Our results also confirm that the glycine-rich regions form semi-extended 3_{10}-helix type structures and not

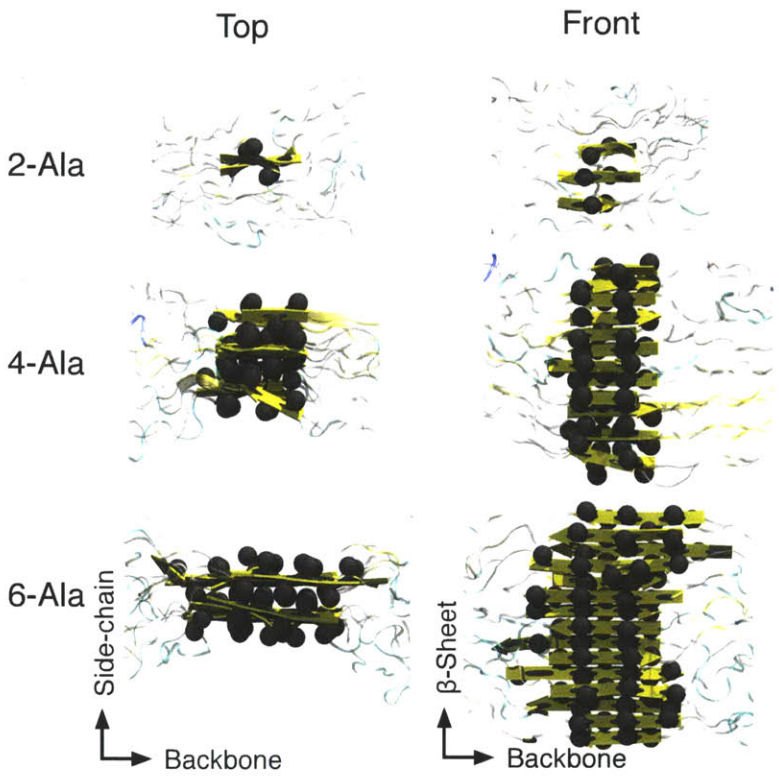

Figure 3-9: Efficient stacking of alanine side-chains determines a critical poly-Ala length for crystal stability. Molecular simulation results directly illustrate trends in crystal stability and β-sheet stacking. Ala side-chain carbon atoms are shown in grey to illustrate alignment in the β-sheet direction and stacking in the side-chain direction.

alpha-helix or beta-helix structures, in agreement with experimental NMR studies [36]. We showed that the poly-Ala region agglomerates under elongation during the spinning process into dominantly anti-parallel β-sheet nanocrystals.

The efficacy of the REMD simulation method can most readily be seen in the final shapes of the β-sheet nanocrystals in our ensemble of cases considered. While the initial lattice structure arranges three layers of strands in the anti-parallel direction, the majority of the final nanocrystals are only two layers in depth. This illustrates the ability of the higher temperature replicas in disrupting initial hydrogen bond networks in order to find structures with lower potential energy. These lower-energy structures are favored in replica exchanges and in turn form the basis of the final lower-temperature structures. While our choice of a three-by-five initial lattice structure is based on intuition of physical conditions, it is limited by current

computational constraints. However, we assume that the high-temperature replicas are able to explore conformations beyond those in the neighborhood of a user-defined, unnatural initial structure.

3.3 Conclusion

The test model ensemble (Figure 3-1) shows that the clearly defined crystals are 2–4 nm in length, depending on poly-Ala length, and consistently 3.1–3.4 nm in width (*i.e.* in the side-chain direction), no matter the poly-Ala length. With identical simulation conditions, each test case produces a nanocrystal that self-assembles into a critical width at which hydrogen bonds within the β-sheet gain a strong character through cooperativity [20]. In addition to being more energetically expensive to create, longer nanocrystals may also prohibit certain mechanisms at a higher hierarchical level. To test the macroscale effects of crystal size and connectivity, the atomistic structure predictions of REMD simulations may be used to train a coarse-grain model of the protein network within the core of a spider silk strand. Such a network would be too large to simulate with atomistic resolution, but the deformation and failure of the network would depend heavily on the shear behavior of the relatively small nanocrystals and the extensible hidden length of the amorphous regions.

Chapter 4

Nanomechanical Testing of MaSp1 Test Cases

The following sections discuss results from the Stretch and Pull-out tests for case of implicit and explicit solvent. Effects of the loading conditions and solvent conditions are compared. Since explicit solvent simulations yield more accurate secondary structure predictions, especially β-sheet content and strain-induced changes in that content, test results with explicit solvent will be discussed in more detail. Test results with implicit solvent are presented for comparison and are discussed briefly.

4.1 Implicit Solvent

The following sections discuss results from the Stretch and Pull-out tests with implicit solvent, then compares the effects of the loading conditions on deformation and failure.

4.1.1 Stretch Test

The failure forces of the Stretch test with implicit solvent show a wide but similar range for each case of poly-Ala length. Failure forces for the 2-Ala case range between 1.30–2.33 nN; for the 6-Ala case, between 1.03–2.04 nN; and for the 12-Ala case, between 1.50–2.12 nN. By dividing the failure force by 15 strands, each strand shows

a failure force near the range of 100–300 pN as predicted for β-sheet-rich proteins in Section 1.2.2. Although the failure forces show a wide range within each test case, the deformation profiles are almost identical after a yielding point, shown in Figure 4-1. This yield point occurs at approximately 250 pN for all structures in all test cases. However, the total length of the protein at this yield point varies with the poly-Ala length. While the structures begin at different total lengths with a range of 5 nm, the yield point occurs at the same total length for each poly-Ala case. For 2-Ala and 6-Ala cases, the yield point occurs when the total length of the protein is nearly 18 nm. However, for 12-Ala cases, it occurs at nearly 23 nm due to the much longer poly-Ala repeat sequence. Beyond this yield point, the total β-sheet content increases rapidly as the strands are straightened and aligned in parallel. After this alignment, a greater number of H-bond donors and acceptors are close enough to form new H-bonds between the newly formed β-strands that have transitioned from more randomly coiled conformations.

4.1.2 Pull-out Test

The failure forces of the Pull-out tests with implicit solvent show a wide range for each case of poly-Ala length and a dependance on crystal size: while the failure force range is similar for the 2-Ala and 6-Ala cases, the 12-Ala cases show failure forces nearly twice as high. In particular, failure forces for the 2-Ala case range between 1.26–2.14 nN; for the 6-Ala case, between 1.25–2.18 nN; and for the 12-Ala case, between 2.67–3.44 nN. The deformation profiles are similar to those of the Stretch tests and also display a similar yield point. However, the location of the yield point depends on the poly-Ala length. While the yield point of the 2-Ala cases is difficult to define, the yield point of the 6-Ala occurs around 0.50 nN, and for 12-Ala cases, around 1.0 nN.

The delayed yielding and superior strength of the 12-Ala cases is attributed to the fundamental differences in the Pull-out test loading conditions compared to those of the Stretch test. Strands in the Stretch test are uniformly stretched and aligned, and β-sheet transitions are therefore similarly uniform. On the other hand, only two

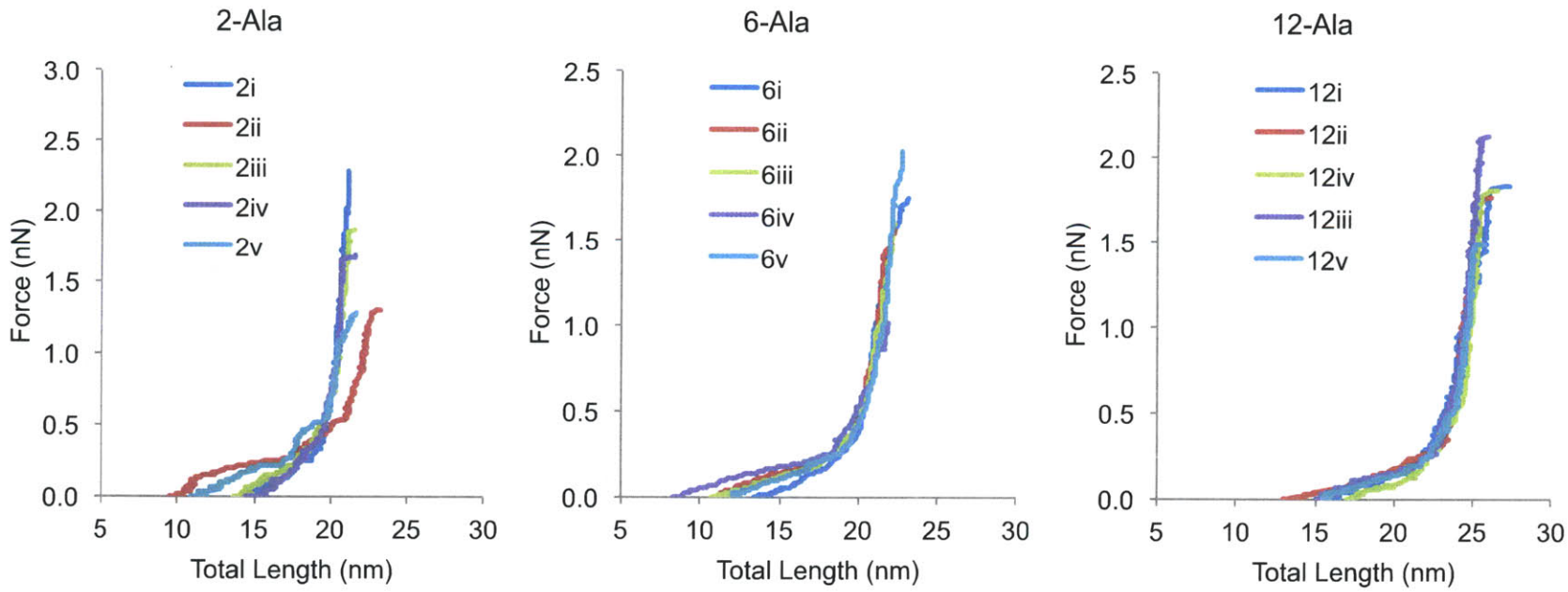

Figure 4-1: Stretch test results for the five largest REMD cluster center structures with implicit solvent.

central strands are loaded in the Pull-out tests, and deformation is confined within the central region *via* β-sheet transitions as a molecular analog of the plasticity normally observed in bulk, continuum materials. For the 12-Ala case, the loaded central strands are confined a larger "plastic" zone of the larger crystal, and more work-to-failure is required to transfer the load through the crystal to other, non-loaded strands. The connections between β-sheet transitions and pseudo-plasticity are discussed in more detail in the following sections.

4.1.3 Distribution of β-sheet Content During Deformation with Implicit Solvent

The Stretch and Pull-out test results reveal that ultimate strength of the MaSp1 unit cell depends on more than just the initial crystals size. Because of the strain-hardening mechanism, *i.e.* the strain-induced transition from random coil conformation to β-sheet, a higher β-sheet content is present closer to failure than in the initial structure. To visualize not only how much the β-sheet content changes with deformation, but especially where and when these transitions occur, Figure 4-3 shows a colormap of β-strand conformation for each residue, averaged across the 15 strands of the unit cell, during deformation. Since the strands alternate direction in the initial lattice used to create the test structure ensemble, the colormap presents only a pseudo-spatial distribution and not a true distribution in Cartesian coordinates. However, displaying β-strand content by residue reveals the specific amino acids and sequences the are more prone to β-sheet transition. This information can then be used to tailor customized MaSp sequences *via* genetic modification to produce superior β-sheet crystals and crystal distribution.

Crystal distribution, and not crystal size alone, may be more important than was previously thought. For the principal structures, the colormap indicates that the failure forces of Pull-out tests depend primarily on total crystal size, while the failure force of Stretch tests depend more on global β-sheet distribution. While the 6-Ala and 12-Ala Stretch tests show nearly identical failure forces and very similar β-sheet

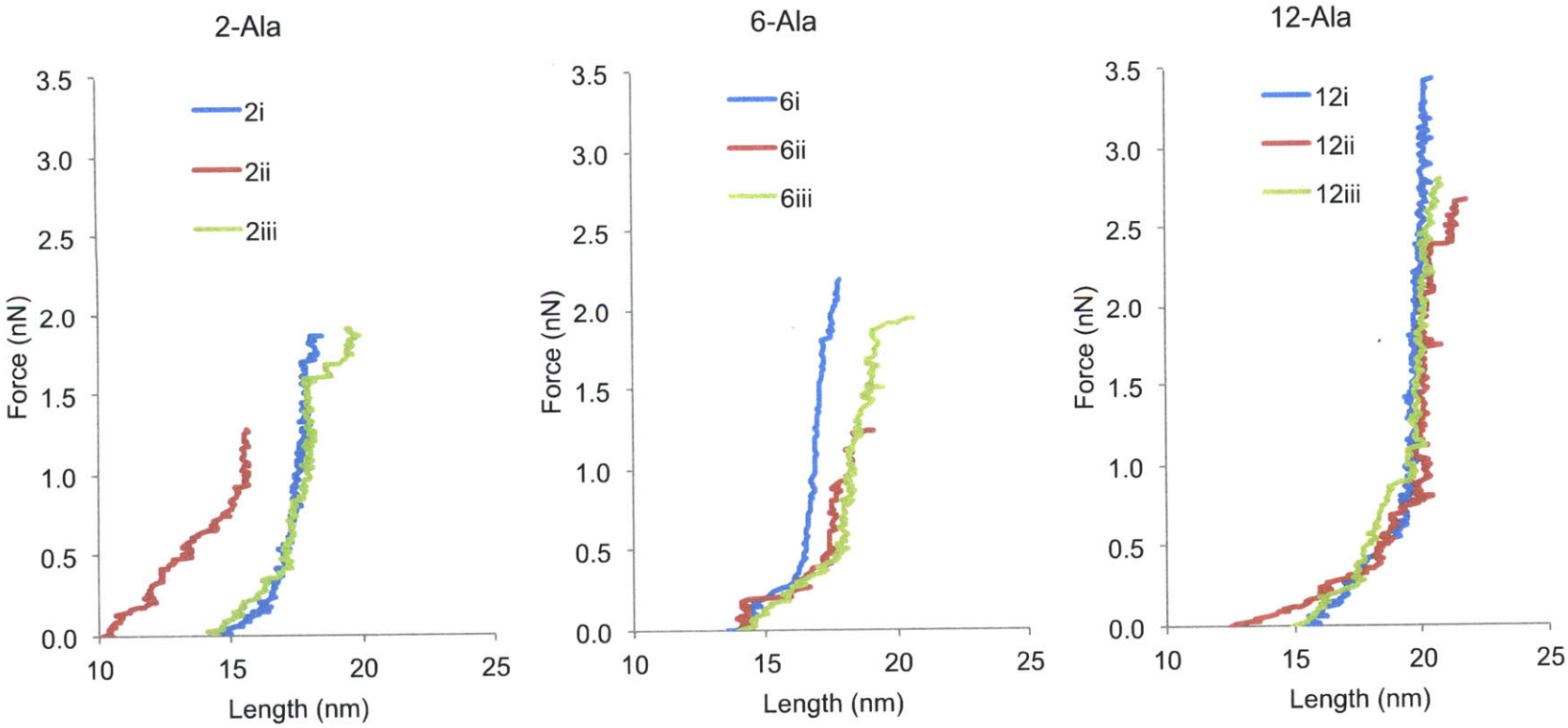

Figure 4-2: Pull-out test results for the five largest REMD cluster center structures with implicit solvent.

distribution at failure, the 2-Ala Stretch test shows a 30% increase in failure force with nearly the same overall β-sheet content at failure.

For all cases, the highest concentration of β-sheet is within the poly-Ala segment, shown as a grey box in 4-3. For the stretch tests, the β-sheet crystal distribution preceding failure of the 2-Ala case features two small, adjacent crystals in contrast to the single, larger crystal of the 6-Ala and 12-Ala cases. This finding of small adjacent crystals having a higher effective shear strength that a single large crystal is in agreement with prior findings of H-bond cooperativity and saturation discussed in Section 1.2.2 and shown again in Figure 4-8. The 6-Ala and 12-Ala cases also show the formation of secondary crystals. For 2-Ala and 6-Ala cases, this is almost immediate, but for the 12-Ala Stretch case, it is delayed until F=0.5 nN. The secondary crystal formation is most evident in the 6-Ala case near residues 11, 21, 44, and 54. These residue IDs correspond to Leucine in (GGL) groups. Similar to Alanine, the side-chain of Leucine is also a hydrocarbon, and thus Leucine is also considered hydrophobic. Although the amyloidization of the poly-Ala repeats into a cohesive crystal is considered the main cross-linking agent, Leucine may also be investigated further as an agent of β-sheet formation. The main secondary crystal of the 6-Ala case coalesces with the main poly-Ala crystal preceding failure, while the main secondary crystal of the 2-Ala case remains separate to a much higher force. This suggests that failure in the wildtype 6-Ala case may be delayed by interrupting the poly-Ala segment, thereby producing two smaller, adjacent crystals that resist coalesces to a much higher failure force than is observed here.

The Pull-out tests suggest similar implications. While the 2-Ala case shows much larger secondary crystals than the 6-Ala case with Pull-out loading conditions, these crystals collapse preceding failure, resulting in a β-sheet distribution and failure force nearly identical to the 6-Ala case. On the other hand, the 12-Ala case maintains a large central crystal to a failure force almost twice that of the other cases. With the Pull-out loading conditions, only two central strands are loaded, and deformation is more localized in the center of the crystal and of the unit cell as a whole than in the Stretch test. This allows a molecular analog of ductility to prevent catastrophic

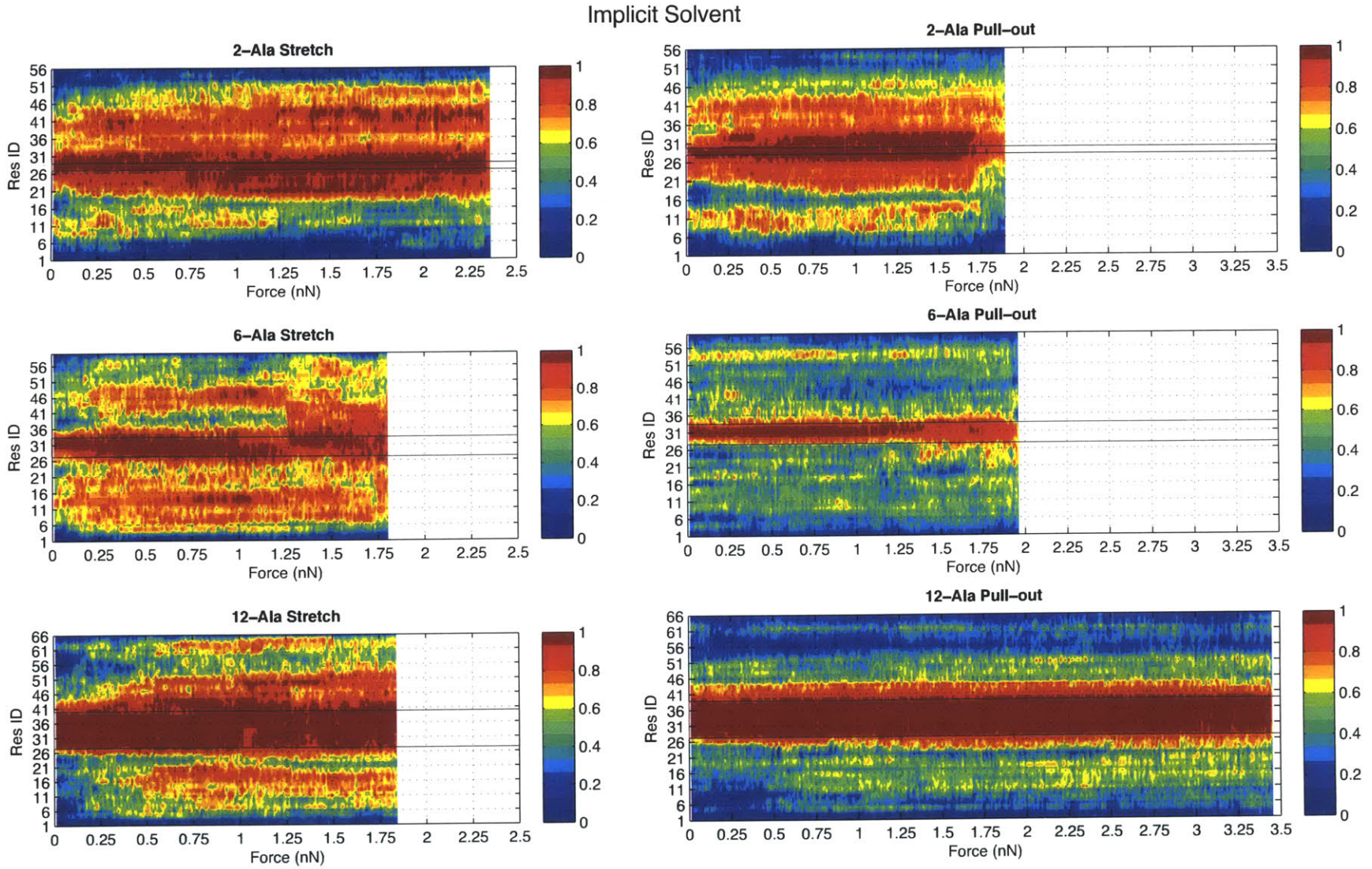

Figure 4-3: Colormap showing averaged β-sheet distribution as a percentage of the 15 strands during Stretch and Pull-out tests for principal structures with implicit solvent. Grey lines indicate the Poly-Ala residues.

failure in favor of deformation. Because only the middle strands of the central crystal are loaded, fewer secondary crystals form, and the failure force is thus more dependent on the size of the central crystal. Therefore, interrupting the poly-Ala segment and forming two smaller, adjacent crystals may still take advantage of the lessons of H-bond cooperativity shown in Figure 4-8.

4.2 Explicit Solvent

The following sections discuss results from the Stretch and Pull-out tests with explicit solvent, then compares the effects of the loading conditions on deformation and failure.

4.2.1 Stretch Test

The failure forces of the Stretch test with explicit solvent fall at the low end of the range seen in the implicit solvent test results. In particular, failure forces are found to be 1.35 nN for the 2-Ala case; 1.23 nN for the 6-Ala case; and 1.45 nN for the 12-Ala case (Figure 4-4. By dividing the failure force by 15 strands, each strand again shows a failure force near the range of 100–300 pN as predicted for β-sheet-rich proteins in Section **??**. The deformation profiles are similar to those with implicit solvent and show a similar strain-hardening around 0.5 nN. However, the yield point is now less defined, and a third, softened regime is observed between the strain-hardened regime and failure. This is especially evident in the 6-Ala case and is attributed to a solvent-mediated molecular analog of ductility that prevents catastrophic failure in favor of deformation. This also greatly increases the work-to-failure, *i.e.* the area under the force-deformation curve. Although the 6-Ala case fails at a lower force than the 2-Ala case, it extends approximately 5 nm longer before failure and thus requires a much larger work-to-failure.

Interestingly, the Stretch tests with explicit solvent fail at nearly the same force as with implicit solvent despite featuring a much lower range in total β-sheet content: 10–35% compared to 50–70% with implicit solvent. This confirms that the hierarchical arrangement of β-sheets is far more important than merely β-sheet content alone.

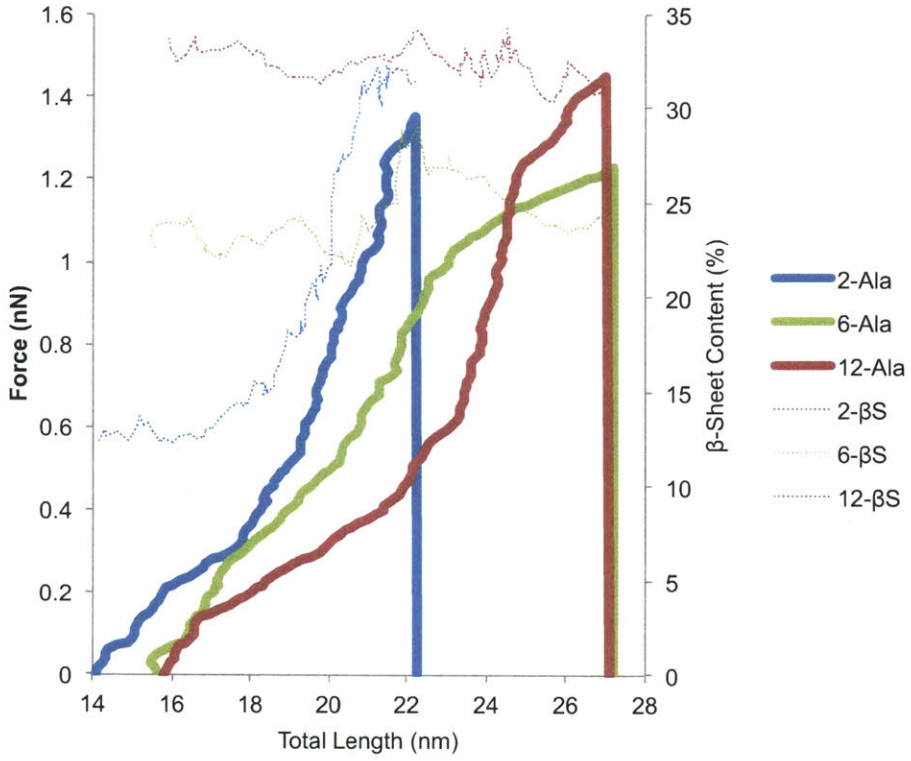

Figure 4-4: Stretch test results for principal structures with explicit solvent.

Despite beginning between 12–33% β-sheet, the β-sheet content of each test case shows a peak near 30% preceding failure due to the loading conditions of the Stretch test and the strain-induced transition to β-sheet. Further analysis of the correlation between β-sheet content and stiffness is discussed in later sections.

4.2.2 Pull-out Test

The failure forces of the Pull-out tests with explicit solvent show a similar trend as those with implicit solvent and fall at the high end of the range seen in the implicit solvent test results. In particular, failure forces are found to be 2.62 nN for the 2-Ala case; 2.60 nN for the 6-Ala case; and 4.12 nN for the 12-Ala case, a 60% increase over the other poly-Ala cases (Figure 4-5. The deformation profiles are similar to those with implicit solvent and show a similar strain-hardening, though later here at 1.50 nN. The yield point is less defined and is attributed to the ability of explicit TIP3P solvent and PME electrostatics to capture secondary structure changes better than

implicit EEF1.1 solvent. A third, softened regime is observed between the strain-hardened regime and failure, similar to — but shorter than — the third regime seen in Stretch tests with explicit solvent. The delayed yielding and superior strength of the 12-Ala cases is again attributed to the fundamental differences in the Pull-out test loading conditions compared to those of the Stretch test. Only two central strands are loaded in the Pull-out tests, and deformation is confined within the central region *via* β-sheet transitions as a molecular analog of the plasticity normally observed in bulk, continuum materials. For the 12-Ala case, the loaded central strands are confined a larger "plastic" zone of the larger crystal, and more work-to-failure is required to transfer the load through the crystal to other, non-loaded strands. Seen in Figure 4-5, changes in the total β-sheet content during deformation are less severe than for the Stretch tests, changing by only 5–7% over the entire simulation. This is a direct result of only the crystal and pseudo-plastic regions being sufficiently deformed to trigger β-sheet transitions.

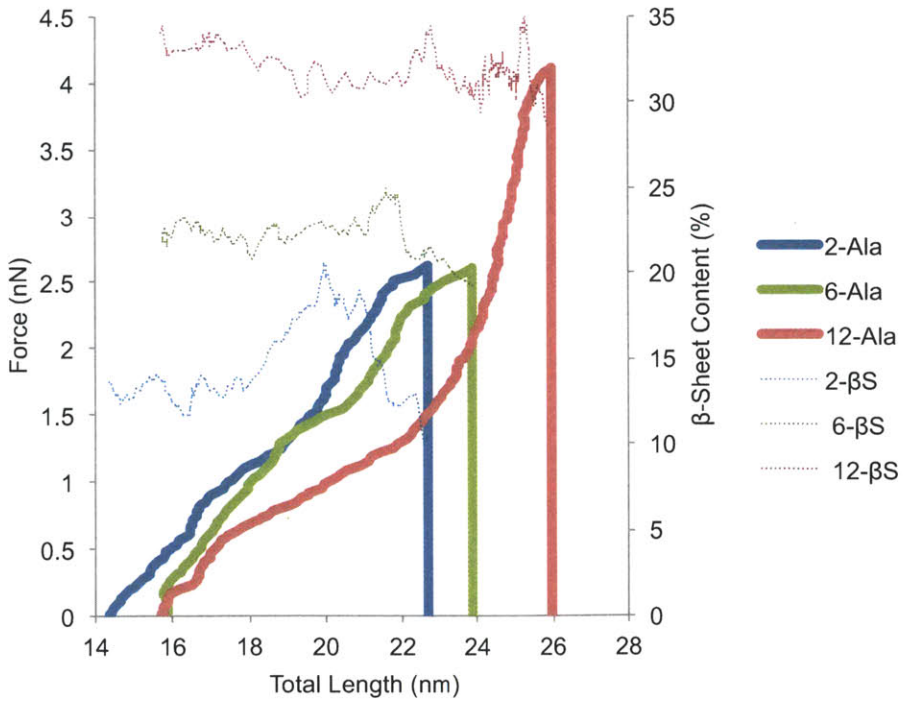

Figure 4-5: Pull-out test results for principal structures with explicit solvent

4.2.3 Distribution of β-sheet Content During Deformation with Explicit Solvent

To reveal at what residues and forces these β-sheet transitions occur, *i.e.* the formation and growth of pseudo-plastic zones preceding failure, a colormap of local β-strand conformation is presented in Figure 4-6. The Stretch and Pull-out test results with explicit solvent reveal that ultimate strength of the MaSp1 unit cell depends on more than just the initial crystals size. Because of the strain-hardening mechanism, *i.e.* the strain-induced transition from random coil conformation to β-sheet, a higher β-sheet content is present closer to failure than in the initial structure.

For all cases, the highest concentration of β-sheet is within the poly-Ala segment, shown as a grey box in 4-6. The largest difference from the tests with implicit solvent is the nearly total absence of β-strand in the semi-amorphous Gly-rich regions. Although some secondary crystals formation is observed, they are much less pronounced than in the implicit solvent tests. With explicit solvent, the failure force appears to hold a direct correlation to the size of the central crystal for both Stretch and Pull-out loading conditions. However, the Pull-out tests show failure forces 2.0–2.7 times higher than for the Stretch tests with the same initial structures and poly-Ala lengths but different loading conditions. For both tests, the failure force depends on the health of the central crystal. For Stretch tests, the central crystal experiences direct loading at a lower force, while for Pull-out tests, a pseudo-plastic zone delays direct loading of the central crystal *via* β-sheet transitions in the semi-amorphous regions.

A visual representation of the health of the central crystal can be inferred from the amount of β-strand conformation within and surrounding the central crystal region in Figure 4-6. For the 2-Ala Stretch case, the Poly-Ala region transitions to a much higher concentration of β-sheet between 0.75–1.0 nN as formerly coiled strands are straightened and aligned, and a large pseudo-plastic zone develops to surround the crystal. After 1.25 nN, the small crystal begins to shear directly, and the crystal deteriorates just before failure. The 6-Ala and 12-Ala cases show similar secondary crystal formation around 0.75–1.0 nN, but less of a pseudo-plastic zone develops. The

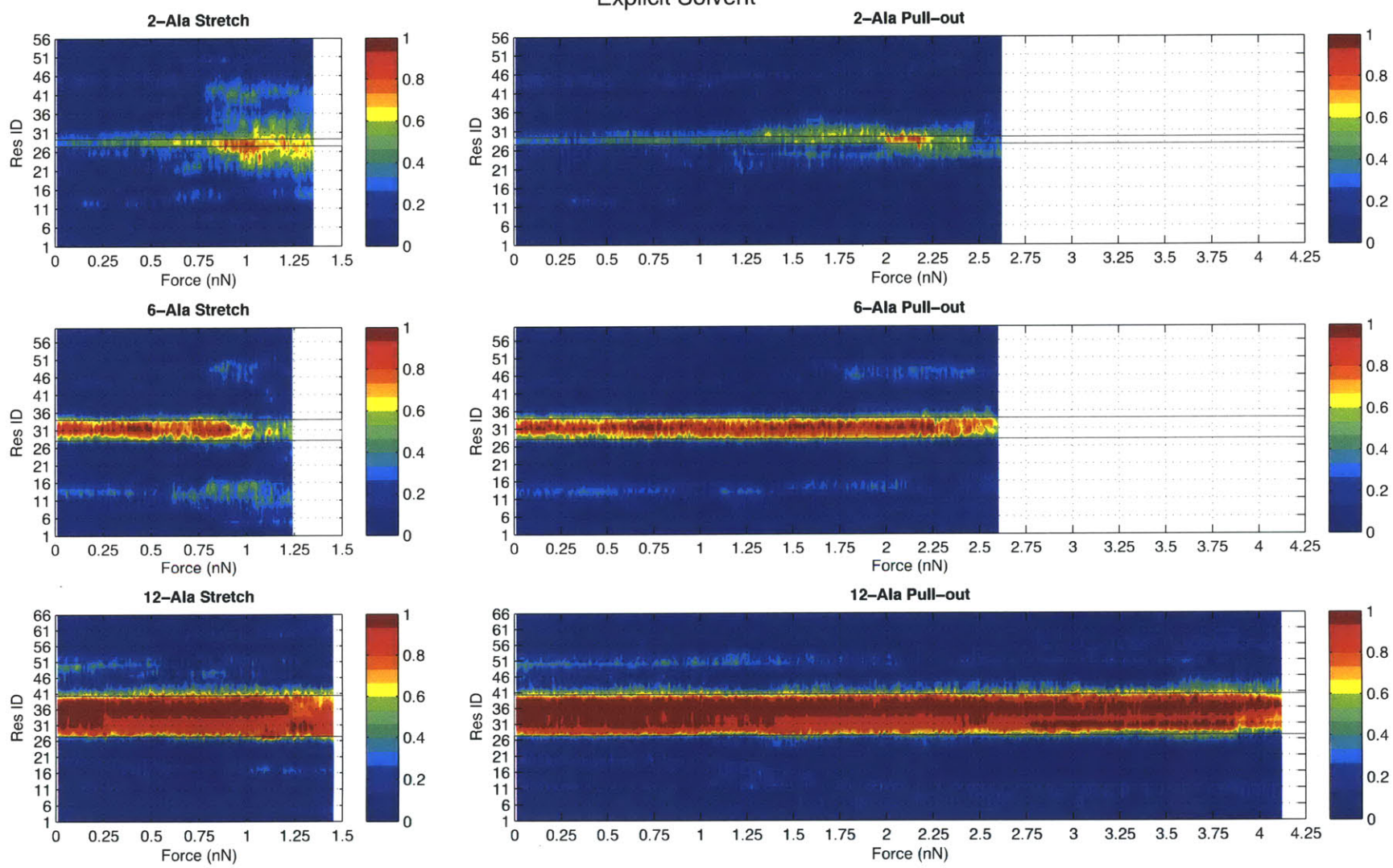

Figure 4-6: Colormap showing averaged β-sheet distribution as a percentage of the 15 strands during Stretch and Pull-out tests with explicit solvent.

β-sheet content and distribution of the 6-Ala case is similar to the 2-Ala case preceding failure, and shows a similar failure force. Because of a negligible development of a pseudo-plastic zone, the central crystal of the 12-Ala case is similarly directly loaded and fails in shear at nearly the same force. Snapshots of the Stretch tests of each test case, shown in Figure 4-7, illustrate the formation and growth of secondary crystals as strands are extended and aligned in parallel.

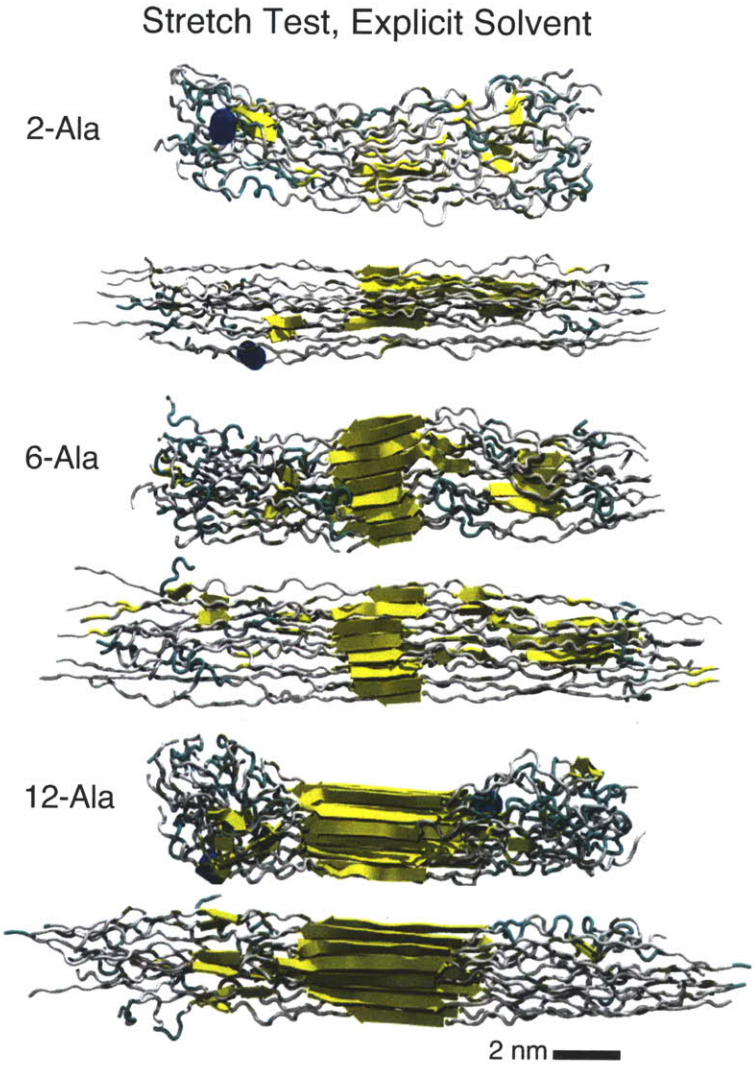

Figure 4-7: Snapshots of the Stretch test with explicit solvent shows each test case before loading and preceding failure. The slight growth of a secondary crystal in the 2-Ala and 6-Ala cases occurs after the strands are extended and aligned, allowing the formation of additional β-sheets. In contrast, the 12-Ala case maintains a defined central crystal until failure. Water is hidden for clarity.

The Pull-out tests results show similar trends, but at forces 2.0–2.7 times higher. For the 2-Ala Stretch case, the Poly-Ala region transitions to a much higher concentration of β-sheet now between 2.0–2.25 nN as formerly coiled strands are straightened and aligned, and a large pseudo-plastic zone develops to surround the crystal. After 2.25 nN, the small crystal begins to shear directly, but the pseudo-plastic zone carries the load to a much higher failure force than in the Stretch test. This loading of the pseudo-plastic zone is the softened third regime preceding failure that was observed in the force/length plots in Figures 4-4 and 4-5. The 6-Ala Pull-out case shows a similar development of a pseudo-plastic zone around 2.0 nN, and deterioration of the central crystal is again more gradual with these loading conditions. The 12-Ala case begins with a very defined central crystal, but slowly loses crystal definition at the edges, seen in Figure 4-6 as a sharp edge from red to blue becoming a wider slope through orange and yellow into light blue. Beyond 2.75 nN, the central crystal appears split into two close, smaller crystals with peaks at residues 31 and 36. This is in agreement with the predictions of H-bond cooperativity, shown visually in Figure 4-7 and schematically in Figure 4-8, and may explain the 60% increase in failure force over the 2-Ala and 6-Ala Pull-out tests.

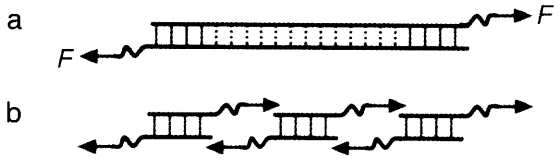

Figure 4-8: H-bond cooperativity decreases with crystal size. (a) Longer crystals reach a maximum shear strength at 3–4 H-bonds, and don't become stronger with additional length. (b) On the other hand, multiple adjacent crystals with 3–4 H-bonds each can take advantage of the increased cooperativity for a much higher effective shear strength. Figure reprinted from [74].

β-sheet Content Correlates to Stiffness

The instantaneous β-sheet content correlates closely with stiffness for tests with explicit solvent, as shown in Figure 4-9. As cross-sectional area and volume of the unit cell is difficult to define at the molecular scale in solvated conditions, stiffness is

presented as a more convenient metric than engineering stress. Stiffness is here defined as the instantaneous slope of the force/displacement curves found by a central difference:

$$k(t) = \frac{\delta F(t)}{\delta l(t)} = \frac{F(t + \Delta t) - F(t - \Delta t)}{l(t + \Delta t) - l(t - \Delta t)}, \tag{4.1}$$

where $k(t)$ is the instantaneous stiffness, $F(t)$ is the force within the stepwise loading profile, $l(t)$ is the total length of the protein, and Δt is the sampling time interval. For all cases, peaks in β-sheet content math peaks in stiffness. This is less evident in 12-Ala cases, however, because total β-sheet content ranges within 3–4% of the initial value, as opposed to 2-Ala cases ranging 10–20%. This marks the softened third regime discussed in Sections 4.2.1 and 4.2.2. Differences in stiffness profiles brought by the differences in the loading conditions of the Stretch and Pull-out also become evident. A characteristic valley in stiffness marks the beginning of the strain-hardening mechanism: a sudden increase in β-sheet content as strands are straightened and aligned. All cases also show a coinciding sudden decrease in β-sheet content and stiffness immediately preceding failure as H-bond clusters are ruptured in shear.

There exists a transition in the ways crystal size and loading conditions affect the stiffness profiles. Small crystals such as those of the 2-Ala case are similar in their range in stiffness but not in the shapes of the stiffness profiles. In particular, the 2-Ala cases both range in stiffness from 0.1–0.7 N/m, but the Pull-out case shows two stiff peaks and maintains a higher average stiffness throughout, even though it reaches only 0.6 times the maximum β-sheet content. In contrast, the 12-Ala cases show nearly identical stiffness profiles and ranges in β-sheet content. However, the ranges in stiffness are very different. The Stretch case reaches only 0.6–0.7 N/m, close to the maximums of the 2-Ala and 6-Ala cases in stretch, while the 12-Ala case in Pull-out reaches a stiffness near 2.0 N/m. This superior stiffness regime results from the large crystal and central loading conditions of the Pull-out test, and it results in the superior mechanical properties observed in previous sections.

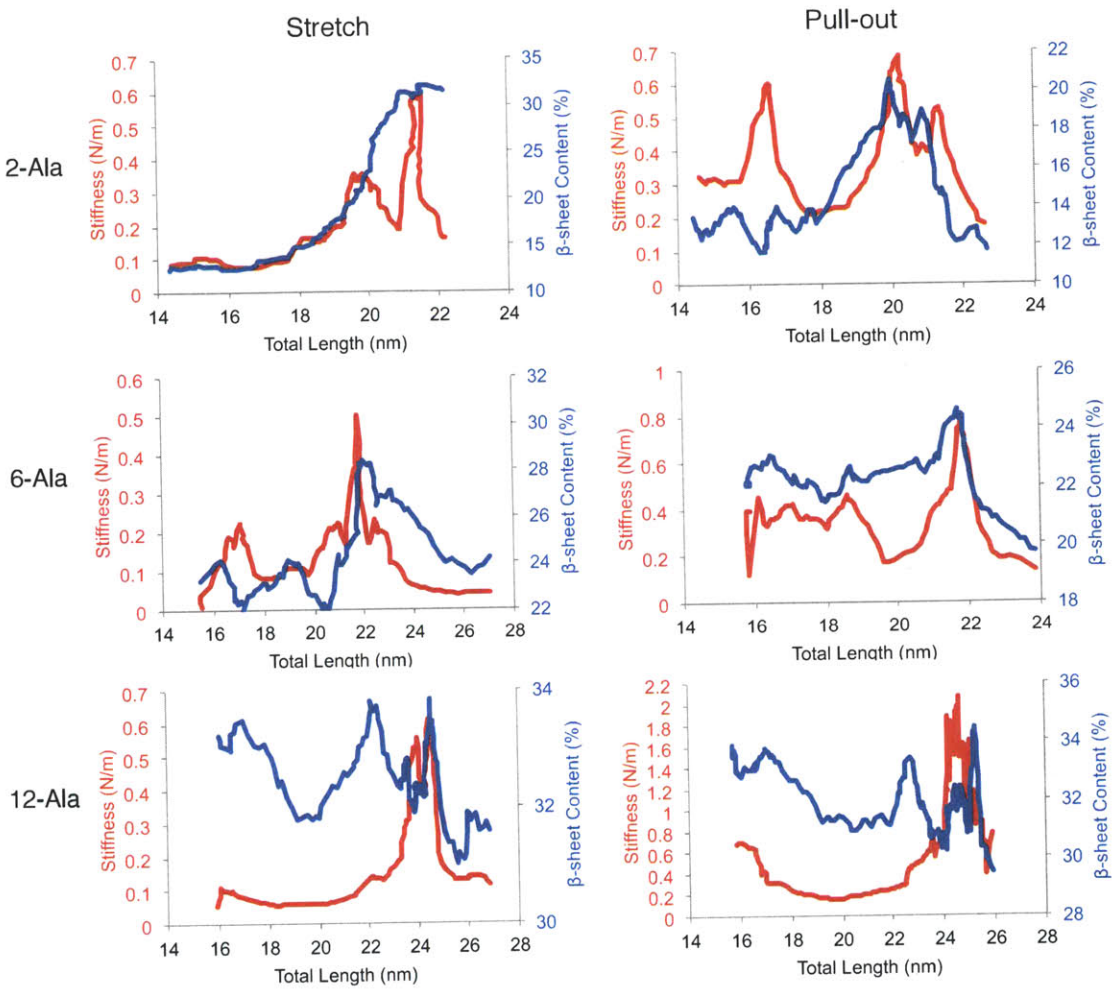

Figure 4-9: β-sheet content (blue) correlates closely to stiffness (red) for testing with explicit solvent. A characteristic drop in stiffness marks the beginning of a strain-hardening mechanism: a sudden increase in β-sheet content as strands are straightened and aligned. Preceding failure, β-sheet content decreases rapidly as H-bond clusters are ruptured in shear.

4.3 3D-Printed Visualization of Stretch Test

Modern molecular visualization programs, such as the Visual Molecular Dynamics (VMD) software package used in this study [78], allow a very user-friendly method of displaying and analyzing the structural data of both biomolecular and synthetic materials at the atomistic scale. However, translating the 3D motion within the VMD display to static media — *e.g.* posters, articles, and slideshow presentations — remains a challenge for the molecular modeling community. While commercial ball-

and-stick molecular model kits may easily be assembled to represent individual amino acids and monomers, no product exists to represent structures of 1,000 amino acids, such as those described in this thesis, or of the cartoon representations of secondary and tertiary structures common to proteins. However, a customized physical model may be constructed using rapid prototyping and additive manufacturing techniques, or "3D printing". The following sections describe the creation of a shadow box used to present 3D printed physical models of snapshots during the Stretch test of the 6-Ala case with implicit solvent in Section 4.1.1.

4.3.1 Rendering of the Surface Geometry

First, the simulation trajectory file (*.dcd) is loaded into VMD with the associated Protein Structure File (*.psf). The trajectory file stores the positions of the atoms at user-set intervals during equilibration or deformation. The PSF file stores bond information among the atoms to properly define amino acids and allows the secondary structure to be calculated. The protein shown in the chosen frame of the trajectory is represented as a cartoon as in Figure 1-5. This represents the backbones as tubes and the β-sheets as arrows, shown in Figure 4-10. The arrow width and arrowpoint size are adjusted to ensure that all of the 15 strands in the protein intersect to create a single solid body. The outer surface geometry of this solid body is rendered by VMD as a stereolithography (*.stl) file. The STL file stores the surface geometry as a list of the vertices and normal unit vector of a triangulated approximation of the surface with a user-defined resolution, shown in Figure 4-11. The STL file is loaded into the CatalystEX computer-aided design (CAD) software used by Dimension 3D printers and is converted into a proprietary CMB file that lists the toolpath information for the printing head. Rendering of the CMB file allows the user to modify the size, resolution, and alignment of the surface geometry before printing.

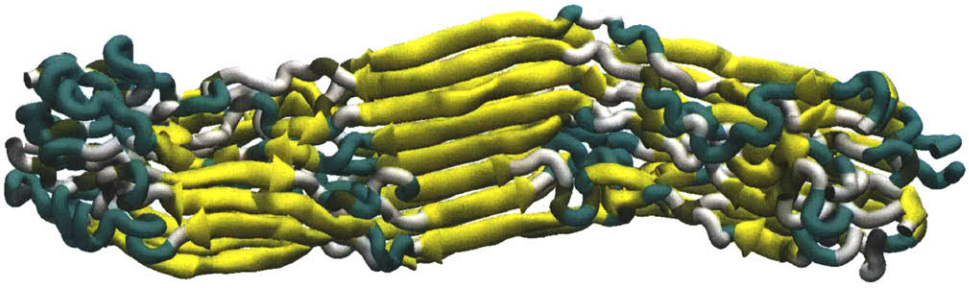

Figure 4-10: A frame of the Stretch test trajectory. The cartoon representation displays the backbones as tubes and the β-sheets as arrows. The arrow width and arrowpoint size are adjusted to ensure that all of the 15 strands in the protein intersect to create a single solid body. The outer surface geometry, but not the color, is exported into an STL file.

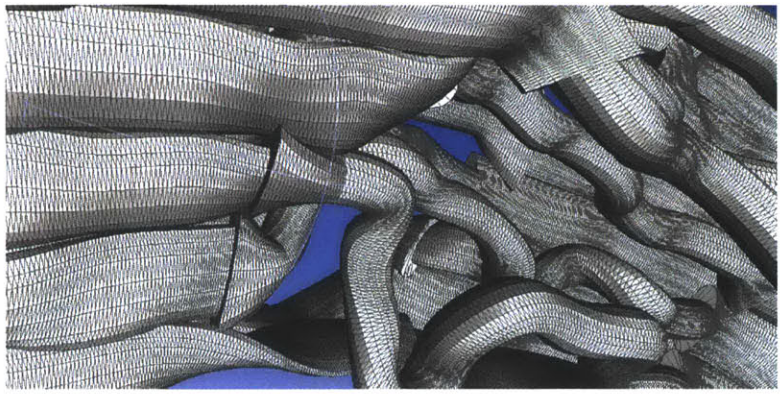

Figure 4-11: A zoomed view of the STL triangle mesh. The surface must represent a single solid body without defects in order to be converted to a toolpath file.

4.3.2 Printing and Assembly

The physical models for the shadow box were printed using the Dimension Elite 3D printer at the Edgerton Center Student Shop, Room 44-023 of the MIT campus. The Dimension Elite series uses Fused Deposition Modeling (FDM), wherein heated ABS thermoplastic is extruded and fused into layers. A separate polymer, the"support material," is also extruded to support ABS layers at an overhang or slope. Once the model is printed, the support material is dissolved in a heated solution of sodium hydroxide and water. After the model is dried, the β-sheet arrows are painted yellow to match the VMD coloring scheme as in Figure 4-10. Black steel wire holds the models in the wood shadow box but allows the models to be easily removed and

handled. The final shadow box and labels are seen in Figure 4-12 and hangs outside Room 1-235A&B at the time of writing.

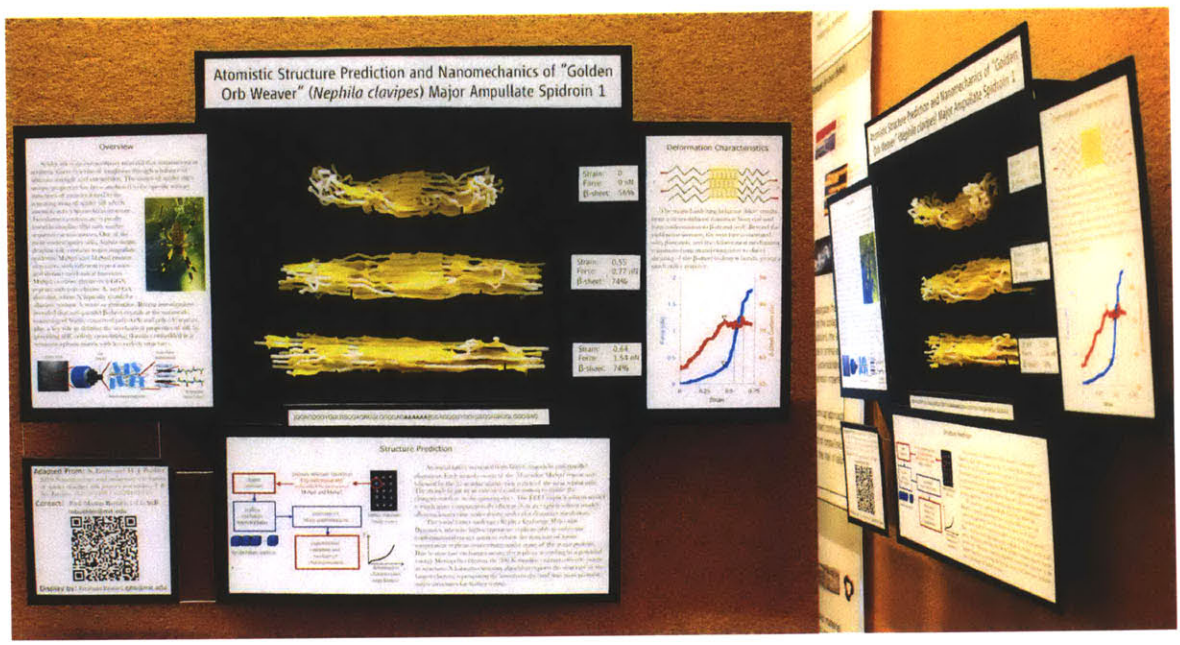

Figure 4-12: β-sheet content (blue) correlates closely to stiffness (red) for testing with explicit solvent. A characteristic drop in stiffness marks the beginning of a strain-hardening mechanism: a sudden increase in β-sheet content as strands are straightened and aligned. Preceding failure, β-sheet content decreases rapidly as H-bond clusters are ruptured in shear.

4.4 Conclusion

The failure forces of the nanomechanical test results for both implicit and explicit solvent are summarized in Figure 4-13. While the initial total β-sheet contents and crystal size and definition are very different for test cases with poly-Ala lengths of 2-, 6-, and 12-Ala, the unit cells' mechanical behaviors and forces at failure are very similar for the Stretch test loading conditions, no matter the solvent condition. Failure forces of the 2-Ala and 6-Ala cases in Pull-out with implicit solvent fall within the same range, but the 12-Ala case shows a much higher failure force. The Pull-out tests results with explicit solvent show a similar trend with poly-Ala length but with failure forces consistently 30% higher. The combination of the 12-Ala case crystal and

Pull-out loading conditions results in a clearly superior unit cell by using a hierarchy of strong β-sheets and soft, extensible semi-amorpous regions to overcome a predicted H-bond saturation. Therefore, future parametric studies in peptide sequence to optimize bulk fiber properties must involve changes in simulated nanomechanical loading conditions to properly assess the effects of the changes in peptide sequence.

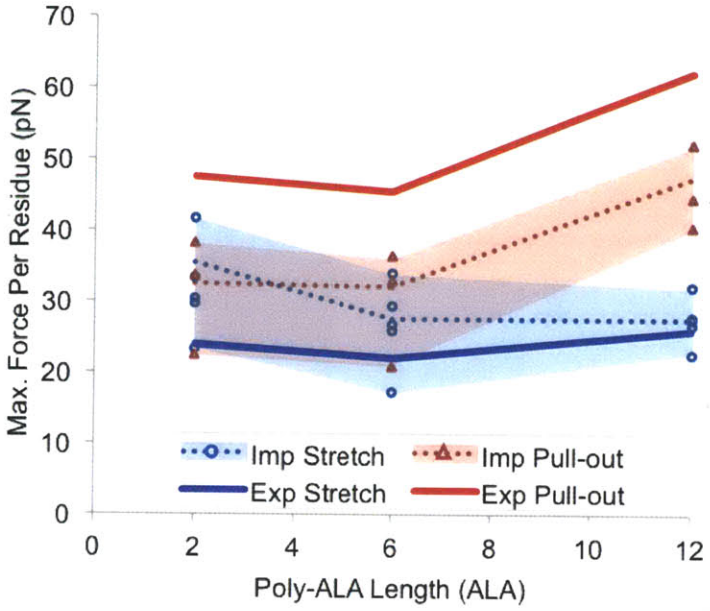

Figure 4-13: Combined summary of failure forces for Stretch and Pull-out test results. The force at failure is normalized by the number of residues per polypeptide strand for each case of poly-Ala length. Averaged implicit solvent results, as dotted lines, are weighted by relative cluster size.

Chapter 5

Conclusions and Outlook on Future Research

5.1 Impact and Contributions

The atomistic simulations discussed in this parametric study have explored the effects of the poly-Ala length of the *N. clavipes* MaSp1 peptide sequence, solvent conditions, and nanomechanical loading conditions on secondary and tertiary structure predictions as well as the nanomechanical behavior of a unit cell of 15 strands with 900–1000 total residues used to represent a cross-linking β-sheet crystal node in the network within a fibril of the dragline silk thread. Understanding the behavior of this node at the molecular scale is critical for potentially bypassing strength limits at this length scale and vastly improving silk for medical and textile purposes as well as synthetic elastomers and polymer or aramid fiber composites with a similar molecular structure and noncovalent bonding for aerospace, armor, and medical applications.

This work constitutes the most comprehensive study to-date of the molecular structure prediction and nanomechanical behavior of dragline silk. While other computational studies have used similar methods for structure prediction and mechanical analysis, *e.g.* REMD and force-control loading or more course methods [35, 20, 37], this work presents:

- the first results of the near-native structures determined by REMD after equilibration in TIP3P explicit solvent,

- the first parametric study of the effects of modifying the wildtype poly-Ala segment length to values outside the range naturally observed for MaSp on structure prediction and nanomechanical behavior,

- the first comparison between previously published loading conditions, *i.e.* the Stretch test, and the novel Pull-out loading conditions that are hypothesized to be more appropriate for modeling of the *in situ* loading of the cross-linking β-sheet crystal, and

- the first 3D-printed models of silk that allow more direct visualization of the molecular mechanics of the nanoscale fibrillar network during deformation.

The test model ensemble (Figure 3-1), determined by REMD, shows that the clearly defined crystals are 2–4 nm in length, depending on poly-Ala length, and consistently 3.1–3.4 nm in width (*i.e.* in the side-chain direction), no matter the poly-Ala length. With identical simulation conditions, each test case produces a nanocrystal that self-assembles into a critical width at which hydrogen bonds within the β-sheet gain a strong character through H-bond cooperativity. The simulation results of poly-Ala length test cases in explicit solvent illustrate these trends in the β-sheet nanocrystal stability in Figure 3-9. For the 2-Ala test case, a single β-sheet of only three strands is found in the interior of the protein. Another β-sheet is not present for stacking in the side-chain direction. In contrast, the 4-Ala test case shows a very aligned two-layer crystal that is open to exterior water. The 6-Ala case shows a similar crystal and also illustrates the tolerance of misalignment. In the center of the crystal, several β-strands are misaligned by two residues. However, the core of the crystal remains 4-Ala in width and thus remains stable in the presence of nearby water molecules. The 12-Ala test case shows highly ordered stacking and is in some places three layers in thickness. Therefore, MD simulation of MaSp1 poly-Ala amyloidization demonstrates that 4-Ala repeats are sufficient for the formation of

β-sheet nanocrystals in the final silk, and that repeats of 6-Ala (wildtype) or longer allow misalignment of the poly-Ala during amyloidization.

This study also demonstrates that the critical conditions for particular secondary, tertiary, and quartenary structure formation, in particular of stable β-sheet nanocrystals, can be found through systematic variation of the peptide sequence alone, and that the length of the poly-Ala repeat unit is critical in defining identifiable β-sheet nanocrystals. Specifically, there exists a strong scaling effect where a minimum length of poly-Ala repeats is required. The results also confirm that the glycine-rich regions form semi-extended 3_{10}-helix type structures and not alpha-helix or beta-helix structures, in agreement with experimental NMR studies [36].

The failure forces of the nanomechanical test results for both implicit and explicit solvent are summarized in Figure 4-13. While the initial total β-sheet contents and crystal size and definition are very different for test cases with poly-Ala lengths of 2-, 6-, and 12-Ala, the unit cells' mechanical behaviors and forces at failure are very similar for the Stretch test loading conditions, no matter the solvent condition. Failure forces of the 2-Ala and 6-Ala cases in Pull-out with implicit solvent fall within the same range, but the 12-Ala case shows a much higher failure force. The Pull-out tests results with explicit solvent show a similar trend with poly-Ala length but with failure forces consistently 30% higher. The combination of the 12-Ala case crystal and Pull-out loading conditions results in a clearly superior unit cell by using a hierarchy of strong β-sheets and soft, extensible semi-amorpous regions to overcome a predicted H-bond saturation. Therefore, future parametric studies in peptide sequence to optimize bulk fiber properties must involve changes in simulated nanomechanical loading conditions to properly asses the effects of the changes in peptide sequence.

5.2 Remaining Challenges

In addition to being more energetically expensive to synthesize, longer poly-Ala segments (and thus larger nanocrystals) may also prohibit certain deformation mechanisms at a higher hierarchical level of the fibril protein network. To test the macroscale

effects of crystal size and connectivity, the atomistic structure predictions of REMD simulations may be used to train a coarse-grain bead-spring model of the protein network within the core of a spider silk strand. Such a network would be too large to simulate with atomistic resolution, but the deformation and failure of the network would depend heavily on the shear behavior of the relatively small nanocrystals and the extensible hidden length of the amorphous regions.

Using the design strategy of REMD structure prediction, structure refinement *via* explicit solvent equilibration, and nanomechanical testing presented in Figure 2-1, parametric studies can be performed on other published consensus peptide sequences of other silks, such as minor ampullate spidroin or the cocoon silk of the silkworm *Bombyx mori*. Experimental synthesis of the modified sequences through genetic modification or microfluidic spinning or of synthetic polymer fibers is necessary for validation of the predicted improvements, but molecular modeling is an significant and novel tool for the design of new materials at the nano- and microscales.

Bibliography

[1] Y. Termonia, "Molecular modeling of spider silk elasticity," *Macromolecules*, vol. 27, no. 25, pp. 7378–7381, 1994.

[2] A. H. Simmons, C. A. Michal, and L. W. Jelinski, "Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk," *Science*, vol. 271, no. 5245, pp. 84–87, 1996.

[3] F. Vollrath and D. P. Knight, "Liquid crystalline spinning of spider silk," *Nature*, vol. 410, no. 6828, pp. 541–548, 2001.

[4] Z. Z. Shao and F. Vollrath, "Materials: Surprising strength of silkworm silk," *Nature*, vol. 418, no. 6899, pp. 741–741, 2002.

[5] N. Becker, E. Oroudjev, S. Mutz, J. P. Cleveland, P. K. Hansma, C. Y. Hayashi, D. E. Makarov, and H. G. Hansma, "Molecular nanosprings in spider capture-silk threads," *Nat. Mater.*, vol. 2, no. 4, pp. 278–283, 2003.

[6] C. Y. Hayashi, N. H. Shipley, and R. V. Lewis, "Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins," *Int. J. Biol. Macromol.*, vol. 24, no. 2-3, pp. 271–275, 1999.

[7] J. Gosline, P. Guerette, C. Ortlepp, and K. Savage, "The mechanical design of spider silks: from fibroin sequence to mechanical function," *J Exp Biol*, vol. 202, no. 23, pp. 3295–3303, 1999.

[8] J. Gatesy, C. Hayashi, D. Motriuk, J. Woods, and R. Lewis, "Extreme diversity,

conservation, and convergence of spider silk fibroin sequences," *Science*, vol. 291, no. 5513, pp. 2603–2605, 2001.

[9] C. Y. Hayashi and R. V. Lewis, "Evidence from flagelliform silk cdna for the structural basis of elasticity and modular nature of spider silks," *Journal of Molecular Biology*, vol. 275, no. 5, pp. 773–784, 1998.

[10] A. E. Brooks, H. B. Steinkraus, S. R. Nelson, and R. V. Lewis, "An investigation of the divergence of major ampullate silk fibers from nephila clavipes and argiope aurantia," *Biomacromolecules*, vol. 6, no. 6, pp. 3095–3099, 2005.

[11] G. P. Holland, M. S. Creager, J. E. Jenkins, R. V. Lewis, and J. L. Yarger, "Determining secondary structure in spider dragline silk by carbon-carbon correlation solid-state nmr spectroscopy," *J. Am. Chem. Soc.*, vol. 130, no. 30, pp. 9871–9877, 2008.

[12] M. B. Hinman and R. V. Lewis, "Isolation of a clone encoding a 2nd dragline silk fibroin - nephila-clavipes dragline silk is a 2-protein fiber," *J. Biol. Chem.*, vol. 267, no. 27, pp. 19320–19324, 1992.

[13] P. A. Guerette, D. G. Ginzinger, B. H. F. Weber, and J. M. Gosline, "Silk properties determined by gland-specific expression of a spider fibroin gene family," *Science*, vol. 272, no. 5258, pp. 112–115, 1996.

[14] A. Sponner, B. Schlott, F. Vollrath, E. Unger, F. Grosse, and K. Weisshart, "Characterization of the protein components of nephila clavipes dragline silk," *Biochemistry*, vol. 44, no. 12, pp. 4727–4736, 2005.

[15] F. Vollrath, "Strength and structure of spiders' silks," *Reviews in Molecular Biotechnology*, vol. 74, no. 2, pp. 67 – 83, 2000.

[16] B. L. Thiel, K. B. Guess, and C. Viney, "Non-periodic lattice crystals in the hierarchical microstructure of spider (major ampullate) silk," *Biopolymers*, vol. 41, no. 7, pp. 703–719, 1997.

[17] J. D. van Beek, S. Hess, F. Vollrath, and B. H. Meier, "The molecular structure of spider dragline silk: Folding and orientation of the protein backbone," *P. Natl. Acad. Sci. USA*, vol. 99, no. 16, pp. 10266–10271, 2002.

[18] T. Lefevre, M. E. Rousseau, and M. Pezolet, "Protein secondary structure and orientation in silk as revealed by raman spectromicroscopy," *Biophys. J.*, vol. 92, no. 8, pp. 2885–2895, 2007.

[19] S. Keten and M. Buehler, "Atomistic model of the spider silk nanostructure," *Appl. Phy. Lett.*, 2010.

[20] S. Keten, Z. Xu, B. Ihle, and M. J. Buehler, "Nanoconfinement controls stiffness, strength and mechanical toughness of beta-sheet crystals in silk," *Nat. Mater.*, vol. 9, p. 359367, 2010.

[21] C. Dicko, F. Vollrath, and J. M. Kenney, "Spider silk protein refolding is controlled by changing ph," *Biomacromolecules*, vol. 5, no. 3, pp. 704–710, 2004.

[22] S. Rammensee, U. Slotta, T. Scheibel, and A. R. Bausch, "Assembly mechanism of recombinant spider silk proteins," *P. Natl. Acad. Sci. USA*, vol. 105, no. 18, pp. 6590–6595, 2008.

[23] M. C. Philip, A. F. Stephen, A. A. Margaret, W. S. John, L. K. David, W. W. Adams, K. E. Ronald, M. David, and L. V. Deborah, "Mechanical and thermal properties of dragline silk from the spider ¡i¿nephila clavipes¡/i¿," *Polym. Advan. Technol.*, vol. 5, no. 8, pp. 401–410, 1994.

[24] D. Porter, F. Vollrath, and Z. Shao, "Predicting the mechanical properties of spider silk as a model nanostructured polymer," *Eur. Phys. J. E*, vol. 16, no. 2, pp. 199–206, 2005.

[25] C. Riekel and F. Vollrath, "Spider silk fibre extrusion: combined wide- and small-angle x-ray microdiffraction experiments," *Int. J. Biol. Macromol.*, vol. 29, no. 3, pp. 203–210, 2001.

[26] J. E. Trancik, J. T. Czernuszka, F. I. Bell, and C. Viney, "Nanostructural features of a spider dragline silk as revealed by electron and x-ray diffraction studies," *Polymer*, vol. 47, no. 15, pp. 5633–5642, 2006.

[27] J. D. van Beek, J. Kummerlen, F. Vollrath, and B. H. Meier, "Supercontracted spider dragline silk: a solid-state nmr study of the local structure," *Int. J. Biol. Macromol.*, vol. 24, no. 2-3, pp. 173–178, 1999.

[28] J. E. Jenkins, M. S. Creager, R. V. Lewis, G. P. Holland, and J. L. Yarger, "Quantitative correlation between the protein primary sequences and secondary structures in spider dragline silks," *Biomacromolecules*, vol. 11, no. 1, pp. 192–200, 2010.

[29] M. E. Rousseau, T. Lefevre, L. Beaulieu, T. Asakura, and M. Pezolet, "Study of protein conformation and orientation in silkworm and spider silk fibers using raman microspectroscopy," *Biomacromolecules*, vol. 5, no. 6, pp. 2247–2257, 2004.

[30] M. E. Rousseau, T. Lefevre, and M. Pezolet, "Conformation and orientation of proteins in various types of silk fibers produced by nephila clavipes spiders," *Biomacromolecules*, vol. 10, no. 10, pp. 2945–2953, 2009.

[31] D. Porter and F. Vollrath, "The role of kinetics of water and amide bonding in protein stability," *Soft Matter*, vol. 4, no. 2, pp. 328–336, 2008.

[32] C. L. Brooks, "Methodological advances in molecular-dynamics simulations of biological-systems," *Curr. Opin. Struc. Biol.*, vol. 5, no. 2, pp. 211–215, 1995.

[33] B. Y. Ma and R. Nussinov, "Simulations as analytical tools to understand protein aggregation and predict amyloid conformation," *Curr. Opin. Chem. Biol.*, vol. 10, no. 5, pp. 445–452, 2006.

[34] M. J. Buehler, S. Keten, and T. Ackbarow, "Theoretical and computational hierarchical nanomechanics of protein materials: Deformation and fracture," *Prog. Mater. Sci.*, vol. 53, no. 8, pp. 1101–1241, 2008.

[35] S. Keten and M. J. Buehler, "Nanostructure and molecular mechanics of spider dragline silk protein assemblies," *Journal of The Royal Society Interface*, 2010.

[36] J. Kummerlen, J. D. van Beek, F. Vollrath, and B. H. Meier, "Local structure in spider dragline silk investigated by two-dimensional spin-diffusion nuclear magnetic resonance," *Macromolecules*, vol. 29, no. 8, pp. 2920–2928, 1996.

[37] M. Cetinkaya, S. Xiao, B. Markert, W. Stacklies, and F. Grter, "Silk fiber mechanics from multiscale force distribution analysis," *Biophysical journal*, vol. 100, no. 5, pp. 1298–1305, 2011.

[38] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chem. Phys. Lett.*, vol. 314, no. 1-2, pp. 141–151, 1999.

[39] P. Bradley, K. M. S. Misura, and D. Baker, "Toward high-resolution de novo structure prediction for small proteins," *Science*, vol. 309, no. 5742, pp. 1868–1871, 2005.

[40] Y. Zhang, "Progress and challenges in protein structure prediction," *Curr. Opin. Struc. Biol.*, vol. 18, no. 3, pp. 342–348, 2008.

[41] K. Y. Sanbonmatsu and A. E. Garcia, "Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics," *Proteins-Structure Function and Genetics*, vol. 46, no. 2, pp. 225–234, 2002.

[42] M. Feig, A. D. MacKerell, and C. L. Brooks, "Force field influence on the observation of pi-helical protein structures in molecular dynamics simulations," *J. Phys. Chem. B*, vol. 107, no. 12, pp. 2831–2836, 2003.

[43] F. Rao and A. Caflisch, "Replica exchange molecular dynamics simulations of reversible folding," *J. Chem. Phys.*, vol. 119, no. 7, pp. 4035–4042, 2003.

[44] Y. M. Rhee and V. S. Pande, "Multiplexed-replica exchange molecular dynamics method for protein folding simulation," *Biophys. J.*, vol. 84, no. 2, pp. 775–786, 2003.

[45] N. Miyashita, J. E. Straub, D. Thirumalai, and Y. Sugita, "Transmembrane structures of amyloid precursor protein dimer predicted by replica-exchange molecular dynamics simulations," *J. Am. Chem. Soc.*, vol. 131, no. 10, pp. 3438–3239, 2009.

[46] M. A. Meyers, P.-Y. Chen, A. Y.-M. Lin, and Y. Seki, "Biological materials: Structure and mechanical properties," *Progress in Materials Science*, vol. 53, no. 1, pp. 1–206, 2008.

[47] J. Black and G. Hastings, *Handbook of Biomaterial Properties.* Springer - Verlag, 2006.

[48] U. G. K. Wegst and M. F. Ashby, "The mechanical efficiency of natural materials," *Philosophical Magazine*, vol. 84, no. 21, pp. 2167 – 2186, 2004.

[49] T. Ackbarow, D. Sen, C. Thaulow, and M. J. Buehler, "Alpha-helical protein networks are self-protective and flaw-tolerant," *PLoS ONE*, vol. 4, no. 6, p. e6015, 2009.

[50] A. Kai-Nan, S. Yu-Long, and L. Zong-Ping, "Flexibility of type i collagen and mechanical property of connective tissue," *Biorheology*, vol. 41, no. 3, pp. 239–246, 2004.

[51] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell.* Garland Science, 2002.

[52] M. Gao, M. Sotomayor, E. Villa, E. H. Lee, and K. Schulten, "Molecular mechanisms of cellular mechanics," *Physical Chemistry Chemical Physics*, vol. 8, no. 32, pp. 3692–3706, 2006.

[53] A. Gautieri, S. Uzel, S. Vesentini, A. Redaelli, and M. J. Buehler, "Molecular and mesoscale mechanisms of osteogenesis imperfecta disease in collagen fibrils," *Biophysical journal*, vol. 97, no. 3, pp. 857–865, 2009.

[54] M. F. Ashby, L. J. Gibson, U. Wegst, and R. Olive, "The mechanical properties of natural materials. i. material property charts," *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, vol. 450, no. 1938, pp. 123–140, 1995.

[55] L. J. Gibson and M. F. Ashby, "The mechanics of three-dimensional cellular materials," *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, vol. 382, no. 1782, pp. 43–59, 1982.

[56] M. J. Buehler, "Tu(r)ning weakness to strength," *Nano Today*, vol. 5, no. 5, pp. 379–383, 2010.

[57] P. Fratzl and R. Weinkamer, "Nature's hierarchical materials," *Progress in Materials Science*, vol. 52, no. 8, pp. 1263–1334, 2007.

[58] T. Ackbarow and M. J. Buehler, "Hierarchical coexistence of universality and diversity controls robustness and multi-functionality in protein materials," *Journal of Computational and Theoretical Nanoscience*, vol. 5, no. 7, pp. 1193–1204, 2008.

[59] A. Gautieri, S. Vesentini, A. Redaelli, and M. J. Buehler, "Hierarchical structure and nanomechanics of collagen microfibrils from the atomistic scale up," *Nano Letters*, vol. 11, no. 2, pp. 757–766, 2011.

[60] X. Hu, K. Vasanthavada, K. Kohler, S. McNary, A. Moore, and C. Vierra, "Molecular mechanisms of spider silk," *Cellular and Molecular Life Sciences*, vol. 63, no. 17, pp. 1986–1999, 2006.

[61] M. Denny, "The physical properties of spider's silk and their role in the design of orb-webs," *J Exp Biol*, vol. 65, no. 2, pp. 483–506, 1976.

[62] D. Huemmerich, T. Scheibel, F. Vollrath, S. Cohen, U. Gat, and S. Ittah, "Novel assembly properties of recombinant spider dragline silk proteins," *Current biology : CB*, vol. 14, no. 22, pp. 2070–2074, 2004.

[63] C. Foo, W. Po, E. Bini, J. Huang, S. Lee, and D. Kaplan, "Solution behavior of synthetic silk peptides and modified recombinant silk proteins," *Applied Physics A*, vol. 82, pp. 193–203, 2006.

[64] N. Du, Z. Yang, X. Y. Liu, Y. Li, and H. Y. Xu, "Structural origin of the strain-hardening of spider silk," *Advanced Functional Materials*, vol. 21, no. 4, pp. 772–778, 2010.

[65] A. Nova, S. Keten, N. M. Pugno, A. Redaelli, and M. J. Buehler, "Molecular and nanostructural mechanisms of deformation, strength and toughness of spider silk fibrils," *Nano Letters*, vol. 10, no. 7, pp. 2626–2634, 2010.

[66] D. T. Grubb and L. W. Jelinski, "Fiber morphology of spider silk: The effects of tensile deformation," *Macromolecules*, vol. 30, no. 10, pp. 2860–2867, 1997.

[67] J. Gosline, M. Lillie, E. Carrington, P. Guerette, C. Ortlepp, and K. Savage, "Elastic proteins: biological roles and mechanical properties," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 357, no. 1418, pp. 121–132, 2002.

[68] B. Isralewitz, M. Gao, and K. Schulten, "Steered molecular dynamics and mechanical functions of proteins," *Current Opinion in Structural Biology*, vol. 11, no. 2, pp. 224–230, 2001.

[69] J. Sulkowska and M. Cieplak, "Mechanical stretching of proteins - a theoretical survey of the protein data bank," *Journal of Physics: Condensed Matter*, vol. 19, no. 28, p. 283201, 2007.

[70] M. Sotomayor and K. Schulten, "Single-molecule experiments in vitro and in silico," *Science*, vol. 316, no. 5828, pp. 1144–1148, 2007.

[71] T. Ackbarow, X. Chen, S. Keten, and M. J. Buehler, "Hierarchies, multiple energy barriers, and robustness govern the fracture mechanics of -helical and -sheet protein domains," *Proceedings of the National Academy of Sciences*, vol. 104, no. 42, pp. 16410–16415, 2007.

[72] S. Keten and M. J. Buehler, "Asymptotic strength limit of hydrogen-bond assemblies in proteins at vanishing pulling rates," *Physical Review Letters*, vol. 100, no. 19, p. 198301, 2008.

[73] T. E. Fisher, A. F. Oberhauser, M. Carrion-Vazquez, P. E. Marszalek, and J. M. Fernandez, "The study of protein mechanics with the atomic force microscope," *Trends in Biochemical Sciences*, vol. 24, no. 10, pp. 379–384, 1999.

[74] S. Keten and M. J. Buehler, "Geometric confinement governs the rupture strength of h-bond assemblies at a critical length scale," *Nano Letters*, vol. 8, no. 2, pp. 743–748, 2008.

[75] S. Penel, R. Morrison, P. D. Dobson, R. J. Mortishire, Smith, and A. J. Doig, "Length preferences and periodicity in b-strands. antiparallel edge b-sheets are more likely to finish in non-hydrogen bonded rings," *Protein Engineering*, vol. 16, no. 12, pp. 957–961, 2003.

[76] L. Eisoldt, J. G. Hardy, M. Heim, and T. R. Scheibel, "The role of salt and shear on the storage and assembly of spider silk proteins," *Journal of Structural Biology*, vol. 170, no. 2, pp. 413–419, 2010.

[77] M. J. Buehler and Y. C. Yung, "Deformation and failure of protein materials in physiologically extreme conditions and disease," *Nat Mater*, vol. 8, pp. 175–188, Mar. 2009.

[78] W. Humphrey, A. Dalke, and K. Schulten, "Vmd - visual molecular dynamics," *J. Mol. Graphics*, vol. 14, pp. 33–38, 1996.

[79] D. Frishman and P. Argos, "Knowledge-based protein secondary structure assignment," *Proteins: Structure, Function, and Bioinformatics*, vol. 23, no. 4, pp. 566–579, 1995.

[80] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydropathic character of a protein," *J. Mol. Biol.*, vol. 157, no. 1, pp. 105–132, 1982.

[81] S. Lifson and C. Sander, "Antiparallel and parallel beta-strands differ in amino acid residue preferences," *Nature*, vol. 282, no. 5734, pp. 109–111, 1979.

Appendix A

Appendix

A.1 REMD job submission script

```
#!/bin/tcsh
#PBS -l walltime=165:30:00
#PBS -l nodes=16:ppn=4
#PBS -V
#PBS -N REX3

cd /cfs/scratch/users/ghb/silk-REX3

mvapich2-start-mpd

###setenv NP `wc -l ${PBS_NODEFILE} | cut -d'/' -f1`
#cat $PBS_NODEFILE | sort -u | awk ' { print $1, "8 ",dirh } ' dirh=/cfs/scratch/users/ghb/silk-REX3
> hosts.abe

awk '{printf("%s 4 /cfs/scratch/users/ghb/silk-REX3 \n",$1);}' < $PBS_NODEFILE|sort -u > hosts.abe
setenv MV2_SRQ_SIZE 4000

aarex.pl -n 20000 -hosts hosts.abe -charmmlog charmm.log -log server.log \
         -par archive,psf=out.psf,em_out.crd  \
         -mdpar prnlev=3,nogb,shake=1,shakemode='hyd',param=19x,eef1file=/cfs/scratch/users/ghb/
         silk-REX3/solvpar.inp \
         -mdpar lang=1,langfbeta=1.0,xpar=/cfs/scratch/users/ghb/silk-REX3/param19_eef1.1.inp \
         -mdpar echeck=99999999.0,xtop=/cfs/scratch/users/ghb/silk-REX3/toph19_eef1.1.inp \
         -mdpar explicit=0,ewald=0,cuton=7,cutoff=9,cutnb=10,dielec=rdie,trunc=switch,dynupdimg=10,
         dynoutfrq=100,
         dynsteps=250 \
         -temp 64:300:650 em_out.crd

mpdallexit
```

A.2 NAMD script for explicit solvent equilibration

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Minimization and Equilibration of
# masp1.6.1 in a Water Box

#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          masp161_wb10.psf
coordinates        masp161_wb10.pdb

set temperature    300
set outputname     masp161_wb10_eq
firsttimestep      0

#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm     on
parameters          ~/bin/par_all27_prot_lipid.inp
temperature         $temperature

# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12.
switching           on
switchdist          10.
pairlistdist        13.5

# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       10

# Constant Temperature Control
langevin            on    ;# do langevin dynamics
langevinDamping     5     ;# damping coefficient (gamma) of 5/ps
langevinTemp        $temperature
langevinHydrogen    off    ;# don't couple langevin bath to hydrogens

# Periodic Boundary Conditions
cellBasisVector1    64.0850    0.   0.
cellBasisVector2     0.   63.6420   0.
cellBasisVector3     0.    0   166.5040
cellOrigin          1.82950   2.2860  212.6730

wrapAll             on

# PME (for full-system periodic electrostatics)
PME                 yes
PMEGridSizeX        72
PMEGridSizeY        64
PMEGridSizeZ        192

# Constant Pressure Control (variable volume)
useGroupPressure      yes ;# needed for rigidBonds
useFlexibleCell       no
useConstantArea       no

langevinPiston        on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  100.
langevinPistonDecay   50.
langevinPistonTemp    $temperature

# Output
outputName          $outputname

restartfreq         25000     ;# 25000steps = every 0.05 ns
dcdfreq             25000
xstFreq             25000
outputEnergies      25000
```

```
outputPressure      25000

#####################################################################
## EXTRA PARAMETERS                                                ##
#####################################################################

#####################################################################
## EXECUTION SCRIPT                                                ##
#####################################################################

# Minimization
#minimize            25000
reinitvels          $temperature
run 10000000 ;# 20ns
```

A.3 CHARMM script for stretch test

```
* Stretching MaSp1 with EEF1 using temperature jump
! RTF AND PARAM FILES
open unit 2 read card name ~/bin/toph19_eef1.1.inp
read rtf card unit 2
close unit 2
open unit 2 read card name ~/bin/param19_eef1.1.inp
read param card unit 2
close unit 2

open unit 10 read card name a.pdb
read sequ pdb unit 10
gener A setup warn
rewind unit 10

open unit 11 read card name b.pdb
read sequ pdb unit 11
gener B setup warn
rewind unit 11

open unit 12 read card name c.pdb
read sequ pdb unit 12
gener C setup warn
rewind unit 12

open unit 13 read card name d.pdb
read sequ pdb unit 13
gener D setup warn
rewind unit 13

open unit 14 read card name e.pdb
read sequ pdb unit 14
gener E setup warn
rewind unit 14

open unit 15 read card name f.pdb
read sequ pdb unit 15
gener F setup warn
rewind unit 15

open unit 16 read card name g.pdb
read sequ pdb unit 16
gener G setup warn
rewind unit 16

open unit 17 read card name h.pdb
read sequ pdb unit 17
gener H setup warn
rewind unit 17

open unit 18 read card name i.pdb
read sequ pdb unit 18
gener I setup warn
rewind unit 18

open unit 19 read card name j.pdb
read sequ pdb unit 19
gener J setup warn
rewind unit 19

open unit 20 read card name k.pdb
read sequ pdb unit 20
gener K setup warn
rewind unit 20

open unit 21 read card name l.pdb
read sequ pdb unit 21
gener L setup warn
rewind unit 21

open unit 22 read card name m.pdb
read sequ pdb unit 22
gener M setup warn
rewind unit 22

open unit 23 read card name n.pdb
read sequ pdb unit 23
gener N setup warn
rewind unit 23

open unit 24 read card name o.pdb
read sequ pdb unit 24
gener O setup warn
```

```
rewind unit 24

open unit 1 write card name out.psf
write psf card unit 1

read coor pdb unit 10 offset  -33
close unit 10

read coor pdb unit 11 offset 60
close unit 11

read coor pdb unit 12 offset 120
close unit 12

read coor pdb unit 13 offset 147
close unit 13

read coor pdb unit 14 offset 207
close unit 14

read coor pdb unit 15 offset 267
close unit 15

read coor pdb unit 16 offset 360
close unit 16

read coor pdb unit 17 offset 420
close unit 17

read coor pdb unit 18 offset 447
close unit 18

read coor pdb unit 19 offset 507
close unit 19

read coor pdb unit 20 offset 567
close unit 20

read coor pdb unit 21 offset 660
close unit 21

read coor pdb unit 22 offset 720
close unit 22

read coor pdb unit 23 offset 747
close unit 23

read coor pdb unit 24 offset 807
close unit 24

ic purge   ! CLEANUP IC TABLE
ic param   ! GET MISSING BONDS AND ANGLES FROM PARAMETER FILE
ic build   ! PLACE ANY MISSING COORDS, E.G. TERMINAL O ON CO2-
! CHECK FOR MISSING HEAVY ATOM COORDS
define test sele ( .not. type H* ) .and. ( .not. init ) show end
! SAVE THE IC TABLE FILLED WITH XTAL DATA
ic fill
open unit 1 write card name out.ic
write ic card unit 1
! the six subunits of pertussus toxin; initial CHARMM ic table
! USE HBUILD TO REBUILD H ATOMS; SPINS METHYLS, ETC. TO LOCAL MINIMUM
coor init sele type H* end
hbuild sele type H* end
! CHECK FOR ANY MISSING COORDS
define test sele .not. init show end
eef1 setup temp 298.15 unit 93 name ~/bin/solvpar.inp
update ctonnb 7. ctofnb 9. cutnb 10. group rdie

mini sd nstep 1000
mini abnr nstep 2000
!This command prints out solvation free energy for each atom
eef1 print nprint 10

open unit 1 write card name em_out.crd
write coor card unit 1
close unit 1

open unit 100 write card name em_out.pdb
write coor pdb unit 100
close unit 100

SET time 20   !ps
SET ts 0.001    !time step in picoseconds
CALC nstep = @time/@ts !number of steps for equilibration
SET cycles  = 2500
```

```
set cnt = 1
label dopull

!ADD Force Constraints to ends to keep molecule aligned

!LAST RESIDUE OF EACH CHAIN
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR -1 SELE segid A .AND. resid 34 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR 1 SELE segid C .AND. resid 1 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR -1 SELE segid E .AND. resid 34 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR 1 SELE segid G .AND. resid 1 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR -1 SELE segid I .AND. resid 34 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR -1 SELE segid K .AND. resid 34 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR 1 SELE segid M .AND. resid 1 .AND. type CA END

!FIRST RESIDUE OF EACH CHAIN
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR -1 SELE segid B .AND. resid 60 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR 1 SELE segid D .AND. resid 93 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR 1 SELE segid F .AND. resid 93 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR -1 SELE segid H .AND. resid 60 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR 1 SELE segid J .AND. resid 93 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR -1 SELE segid L .AND. resid 60 .AND. type CA END
PULL FORCE 2 XDIR 0 YDIR 0 ZDIR 1 SELE segid N .AND. resid 93 .AND. type CA END

open unit 42 write card name ./@cnt.rst !To restart after crash
open unit 43 write file name ./@cnt.dcd !Trajectory
open unit 44 write card name ./@cnt.tpc !Thermostat data
open unit 45 write card name ./@cnt.ene !Energy data

tpcontrol nther 1 ther 1 tref 300 tau 0.1 select all end !Turn on Nose-Hoover

dyna vv2 start nstep @nstep timestep @ts -
        nprint 50000 iprfrq 1000 nsavc 10000 -
        nsavv 10000 isvfrq 10000 -
        iunrea -1 iunwri 42 iuncrd 43 iuno 44 -
        iunvel -1 kunit 45 -
        firstt 300 finalt 300 nsnos 100 -
        ntrfrq 100

!This command prints out solvation free energy for each atom
eef1 print

close unit 42
close unit 43
close unit 44
close unit 45

incr cnt by 1
if cnt le @cycles goto dopull

open unit 100 write card name MD_out.pdb
write coor pdb unit 100
close unit 100

stop
```

A.4 NAMD script for stretch test

```
#############################################################
## JOB DESCRIPTION                                         ##
#############################################################

# Minimization and Equilibration of
# masp1.6.1 in a Water Box


#############################################################
## ADJUSTABLE PARAMETERS                                   ##
#############################################################

structure          masp161_wb10.psf
coordinates        6_160.pdb

set temperature    300
set outputname     masp161_stretch
firsttimestep      0


#############################################################
## SIMULATION PARAMETERS                                   ##
#############################################################

# Input
paraTypeCharmm     on
parameters          par_all27_prot_lipid.inp
temperature         $temperature


# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12.
switching           on
switchdist          10.
pairlistdist        13.5


# Integrator Parameters
timestep            2.0  ;# 2fs/step
rigidBonds          all  ;# needed for 2fs steps
nonbondedFreq       1
fullElectFrequency  2
stepspercycle       10


# Constant Temperature Control
langevin            off    ;# do langevin dynamics
langevinDamping     5     ;# damping coefficient (gamma) of 5/ps
langevinTemp        $temperature
langevinHydrogen    off    ;# don't couple langevin bath to hydrogens


# Periodic Boundary Conditions
cellBasisVector1    200.    0.   0.
cellBasisVector2     0.   200.   0.
cellBasisVector3     0.    0   600.
cellOrigin          0. 0. 212.

wrapAll             on

# PME (for full-system periodic electrostatics)
PME                 yes
PMEGridSizeX        216
PMEGridSizeY        216
PMEGridSizeZ        648

# Output
outputName          $outputname

restartfreq         50000     ;# 5000steps = every 0.01 ns
dcdfreq             5000
xstFreq             50000
outputEnergies      10000
outputPressure      10000


#############################################################
## EXTRA PARAMETERS                                        ##
#############################################################
```

```
# Put here any custom parameters that are specific to
# this job (e.g., SMD, TclForces, etc...)

constantforce yes
consforcefile 6ref.pdb

#############################################################
## EXECUTION SCRIPT                                        ##
#############################################################

for { set a 0 } { $a < 150 } { incr a 1 } {
set b [expr {double($a)*0.2879}]
consForceScaling $b
run 10000
}
```

A.5 Amino Acid Side-chain Chart