UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO STATE UNIVERSITY

# Understanding the Native Spider Silk Spinning Process Using Recombinant Silk Proteins

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in

Chemistry

by

Dillan Stengel

Committee in charge:

University of California San Diego:
- Judy Kim
- Robert Pomeroy

San Diego State University
- Gregory Holland, Chair
- Erica Forsberg
- Marina Kalyuzhnaya
- John Love

2021

This dissertation of Dillan Stengel is approved, and it is acceptable in the quality and form for publications on microfilm and electronically.

_____

_____

_____

_____

_____

_____

Chair

University of California San Diego

San Diego State University

2021

Dedication

To my mother and father, Mehrnaz and Jess Stengel,

who have always provided me with what I needed to succeed.

Epigraph

“The best way to have a good idea is to have a lot of ideas.”

– Dr. Linus Pauling

"What's miraculous about a spider's web?" said Mrs. Arable.

"I don't see why you say a web is a miracle-it's just a web.”

"Ever try to spin one?" asked Dr. Dorian.

— E.B. White, Charlotte’s Web

Table of Contents

List of Figures

List of Tables

Acknowledgements

Chapter 2 is a full reprint of “Investigating the Atomic and Mesoscale Interactions That Facilitate Spider Silk Protein Pre-Assembly” published in Biomacromolecules 2021. Onofrei, D.; Stengel, D.; Jia, D.; Johnson, H. R.; Trescott, S.; Soni, A.; Addison, B.; Muthukumar, M.; Holland, G. P. David Onofrei was the primary author of this paper.

Chapter 6 is a full reprint of “Spider Prey-Wrapping Silk is an α-Helical Coiled-Coil/β-Sheet Hybrid Nanofiber” Published in Chem. Commun. 2018. Addison, B.; Onofrei, D.; Stengel, D.; Blass, B.; Brenneman, B.; Ayon, J.; Holland, G. P. Bennett Addison was the primary author of this paper.

Chapter 7 is a full reprint of “Hydration-Induced β-Sheet Crosslinking of α-Helical-Rich Spider Prey-Wrapping Silk” published in Advanced Function Materials in 2021. Dillan Stengel, J. Bennett Addison, David Onofrei, Nha Uyen Huynh, George Youssef, and Gregory P. Holland. Dillan Stengel was the primary author of this paper.

# Vita

2012 BS in Biochemistry and Cell Biology, UCSD

2021 PhD in Chemistry, UCSD/SDSU

Abstract of the Dissertation

Understanding the Native Spider Silk Spinning Process Using Recombinant Silk Proteins

by

Dillan Stengel

Doctor of Philosophy in Chemistry

University of California San Diego, 2021

San Diego State University, 2021

Professor Gregory Holland, Chair

Spider silks are a biological protein polymer with an incredibly diverse range of mechanical properties and functions. Spider silk has been recognized as having mechanical properties that

rival man-made materials, fueling attempts to make synthetic versions. However, there have been only a few examples of synthetic silks meeting or surpassing natively-spin materials to date - none of which have been successfully commercialized. This is likely due to the numerous factors working in concert at length scales between the atomic and macroscale to create a continuous fiber with native-like properties. The goal of the research presented in this dissertation is to improve our understanding of the fiber spinning process at these length scales and to understand the material properties of the resulting fibers. These processes were studied using advanced liquid and solid-state nuclear magnetic resonance spectroscopy techniques, solid-phase peptide synthesis, molecular dynamics simulations, dynamic light scattering, and electron microscopy on native and recombinant silk proteins in solution and as fibers. In Chapter 2, we investigate the atomic and mesoscale interactions that facilitate spider silk pre-assemblies and find that dragline silk proteins in 4 M urea show the greatest deviation from glandular conditions at the peripheral ends of the poly(Ala) motifs. These proteins also display a critical aggregation concentration near 4 wt% in 4 M urea. In Chapter 3, we utilize molecular dynamics (MD) simulations to assist in developing a more accurate structural depiction of a native-sized spider silk proteins in solution, and find consistent residues located on the surface of the simulated structure which may help to drive pre-assembly formation. In Chapter 4, we synthesize peptide mimics to study the hierarchical pre-assembly of native silk proteins and show that sequences taken from the amorphous region of dragline silks are capable of hierarchical structure formation and liquid-liquid phase separation. In Chapter 5, we discuss our results from a collaboration with Prof. Thomas Scheibel's Lab at the University of Bayreuth in Germany where we investigate the secondary structure of isotopically-labeled recombinant silk as concentrated dopes and fibers. In Chapters 6 and 7, we discuss a novel silk used during prey wrapping that is capable of crosslinking when wetted with water, where the as-spun mainly $\alpha$-helical, coiled-coil silk transitions to having increased $\beta$-sheet character with matted morphology following water treatment.

**Chapter 1:** Introduction to Spider Silk

## Spider Silks

Silk is one of many biological polymers that fulfill structural roles in the natural world. To date, humans have successfully utilized this protein-based biopolymer for textiles and niche uses in the medical industry.[1, 2] Despite the diversity of silks in nature, only silkworm silk has historically seen large-scale industrial and commercial scale utilization. Unfortunately, silkworms have a single silk gland which produces only one silk type during their lifecycle. In contrast, an individual spider can produce up to seven types of silks which are used on a daily basis for a range of applications that suits the spider's needs.[3, 4] These protein-based fibers have a diverse range of material properties and are highly specialized to be used in a number of places in web construction, prey wrapping, egg cases, and during a fight or flight response (**Figure 1**).[5-7] Depending on application, they range from highly extensible or elastic, to tough and brittle.[8, 9] Because these fibers offer a wide range of mechanical properties and are biodegradable, they are favored in the pursuit of replacing petroleum-based polymers as we look forward towards a greener future.

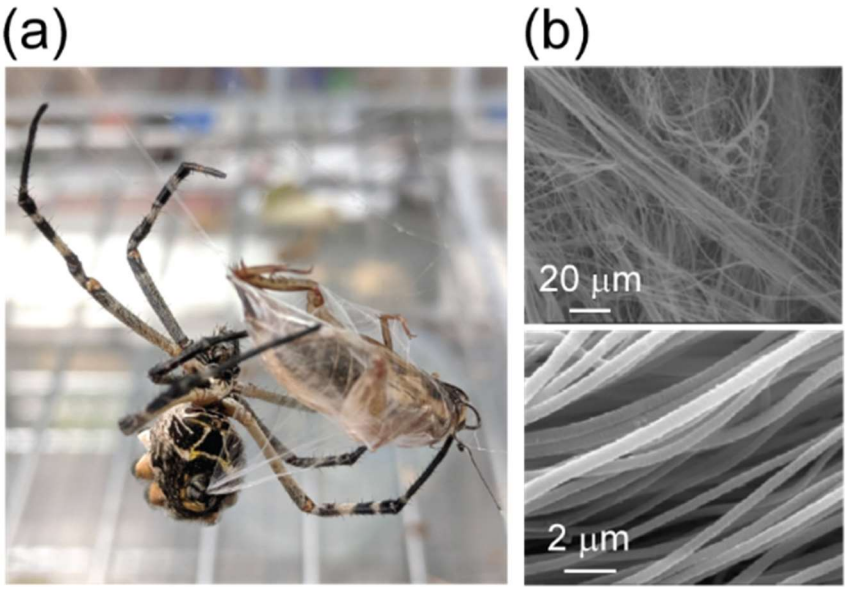



**Figure 1.** (a) *A. argentata* spinning a cricket with aciniform (Ac, prey wrapping) silk. (b) SEM images of aciniform fibers.

Before looking at each silk in detail, it would be helpful to discuss the broader characteristics of these proteins. The most-widely studied spider silk is dragline silk, getting this name from the spider "dragging" the fiber behind it as it repels during a fight or flight response. This silk is produced in the major ampullate (Ma) gland and also goes by the name major ampullate silk. Ma silk is primarily composed of two spidroin (Sp) proteins – MaSp1 and MaSp2, although recently it has been shown that other MaSp proteins may be present and are currently being discovered.[10-12] The dragline silk is not entirely comprised of Ma silk however, and a small amount of minor ampullate (Mi) silk is also incorporated in small amounts. Prey wrapping silks are synthesized in the aciniform (Ac) gland. The spidroin proteins in these glands are extruded by the spider to form the final fiber, where the final name of the protein is a joining of both the original gland and protein name (e.g. MaSp, AcSp, MiSp, etc).

To date, all spider silk proteins seem to share three common features:

1. They are high (250-550 kDa) molecular weight.[13-15]
2. The majority of the protein sequence is made from iterated repeats with flanking N- and C-termini which form dimers at different stages of the fiber forming process.[16-19]
3. The proteins have varying degrees of intrinsic disorder stemming from their low sequence complexity.

The two primary MaSp proteins (MaSp1, MaSp2) that comprise the Ma silk of black widow spiders *(L. hesperus*) are 250 and 310 kDa, with alternative sequences of GGX, GPGGX (MaSp2), or poly(Ala), where X can be Ala, Gln, Tyr, or Pro in the case of MaSp2 (**Figure 2**). Glycine and alanine make up approximately 70% of MaSp1 and 65% in MaSp2.[13] Apart from the well-folded termini in both MaSp1 and MaSp2, the protein is generally disordered and displays essentially no secondary structure in the gland before extrusion akin to an Intrinsically Disordered Protein (IDP's).[20] After fiber extrusion however, the majority of Ala located in the poly(Ala) repeats

as well as Gly and Ser residues flanking these motifs are found within nanocrystalline β-sheet domains that are dispersed throughout a disordered GGX matrix.[21-24] These poly(Ala) β-sheet domains impart strength to the fiber, where the GGX regions give the fiber extensibility and

…YGRGGYGQGGAGQGGAGAAAAATAAGAGQGGQGGYG
QGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQ
GGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAGGA
GQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQ
GGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGY
GRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGA…

…GGQSGYG**P**GGAGAAAAAAAGGAG**P**GRQQEYG**P**GGSG
AAAAAAAAAGSG**P**SGYG**P**GAAG**P**IG**P**GGAGAAAAGGSG**P**
VGYGQG**P**SGYGASGTGGEQDYG**P**GGSGAAAAAAAAAGG
AG**P**GRQQGYG**P**GGSGAAAAAAAGG**P**GYGGQQGYG**P**GG
AGAAAAAAAGGAG**P**GRQQ**P**YG**P**GGAGAAAAAAGGSG**P**G
GYGQG**P**SGYGASG**P**GGQQGFG**P**GGSGAAAAAAAGGAG**P**
G…

**Figure 2.** Partial MaSp1 (top) and MaSp2 (bottom) sequences taken from *L. hesperus* (black widow). Ala is found in long runs in both sequences. Pro occurs frequently in MaSp2, shown in bold.

toughness.

As another example, AcSp proteins used for prey-wrapping, are believed to exhibit partial secondary structure as they are stored in the gland.[25] These proteins have iterated repeats of approximately 200 amino acids that make up the majority of the AcSp protein (**Figure 3**).[15] Each of these iterated repeats have been shown to display loose α-helical character at the core of the sequence, with disorder mostly occurring in the Gly/Pro-rich tail.[25-27] Together, the entire structure would look something like “beads-on-a-string” where the folded region is the bead, and the disordered linker region is the string. After extrusion from the spider, the final Ac fiber exhibits

…LISRVANALSNTSTLRTVLRRGVSQQIASSILRRA
AQTLASTLGVDGNNLSRVALQAISQVPTGSDTSAY
AQAFSSALFNAGVLNASNIDTLGSRVLSAVLNGVS
SAAQGLGINVDTGSVQSDISSSSSFLSTSSSASSFS
SQASASSTSGAGYTGPSGYTGPVGGGAQFGSAS
GQSSFGQTSGLTASSGGQAAFGGTSGASAG…

**Figure 3.** Schematic of *Argiope* AcSp1 prey-wrapping silk. In the *A. argentata* species, there are 20 iterated repeats of the wrapping (“W”) subunit.

high α-helical character where there is additional evidence that it is partially coiled-coiled.[28] It should also be noted that, due to the folded nature of the "bead" region of this protein, it has greater extensibility and in fact surpasses the toughness of other silks, including dragline silk.[8]

**Molecular Structure of Silk Proteins and Fibers**

The structure of silk proteins in fibers continues to be studied by a variety of techniques including solid-state NMR, X-ray diffraction, and Raman spectroscopy.[29-32] A simplistic model for the structure of dragline fibers has been proposed and is widely accepted in the silk community.[33] This model shows that dragline silk is a semi-crystalline biopolymer containing nanocrystalline regions dispersed throughout an amorphous or glassy matrix (**Figure 4, C**). The fiber core is surrounded by two layers: an outer lipid layer and an inner layer which maintains the pressure exerted by the amorphous region.[34, 35] The inner core is made of smaller fibrils oriented along the main fiber axis. In Ma fibers, ~82% of the Ala residues in the sequence are in β-sheet structures

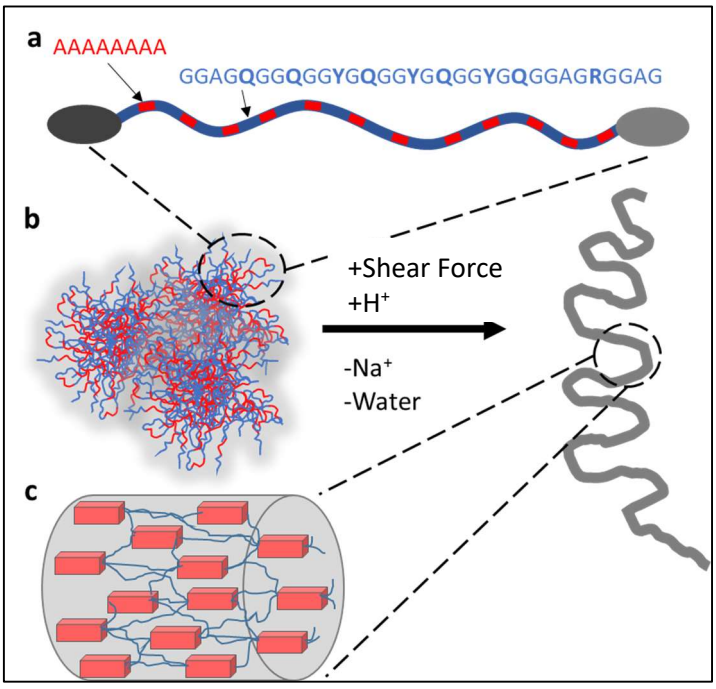



**Figure 4.** Summary of dragline silk formation and fiber models based on previous studies. a), model of MaSp1 silk primary sequence. b), these Ma proteins exist as intrinsically disordered proteins in the gland and through physiological changes are converted to fibers. c), the poly(Ala) regions in the primary sequence come together to form nanocrystalline β-sheet domains (red) where the GGX regions (blue) are generally reported to be disordered.

as reported by NMR.[21] Additionally, Gly and Ser residues flanking the poly(Ala) motifs are also found in β-sheets structures. The Gly-Gly-X regions have been reported to adopt a number of structures, from random coils, to β-turns, to some β-sheets but, is generally believed to be disordered or amorphous.[21, 22, 36] In MaSp2, the proline-rich regions partially exist in elastin-like type II β-turn structures that are hypothesized to be responsible for some of the extensibility of dragline silks.[37]

The structure and dynamics of spider silk proteins prior to spinning has been one of the most challenging aspects of spider silk investigations. Prior to spinning, the proteins are housed in a storage gland (**Figure 5, B**) where they remain essentially disordered (random coils) except for the helical termini which are well -defined helical bundles in a number of protein types.[38, 39] Because the silk is converted rapidly (less than a second) from soluble dope to fiber, this suggests that pre-assembly may be present even in the liquid state. Due to the lack of secondary structure reported from our group as well as others on the stored proteins, we suggest this order is probably supramolecular – the result of several protein monomers coalescing into disordered assemblies.

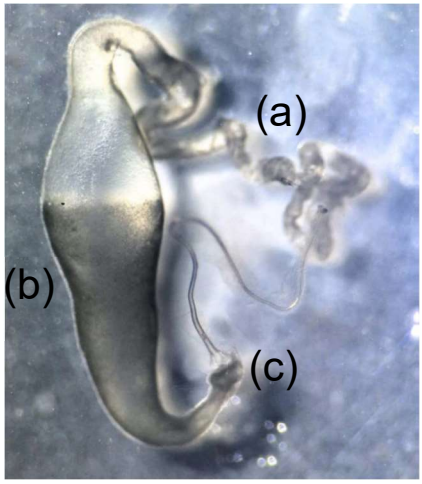



**Figure 5.** Black widow Ma gland. (a), this is where the fiber spinning dope is created. (b) is the storage sac where the spidroins are stored at very high concentration (~25-50 wt.%). The sac opens to the spinning ducts (c) which leads to the spinnerets. The transition from soluble protein to insoluble fiber happens very quickly and occurs through the removal of water, exchange of $Na^+$ ions for $K^+$ ions, and continuously decreasing pH gradient as the dope moves towards the spinnerets where the fiber is extruded.

Since the concentration of spidroin proteins in the gland are extremely high (25-50 wt%),[40] this is key to understanding how silk is stored and spun on-demand into a fiber.

The current understanding of the spinning process is based on several principles which originated after studying silkworm silk. In silkworms, the large repetitive core domain of the silk protein is dominated by hydrophobic residues while the termini contain hydrophilic ones.[41] It was hypothesized that the silk proteins form roughly spherical micelle-like assemblies as stored in the gland solution.[41] This highly concentrated spinning dope has also been described as a type of liquid crystal where dope flow through the duct induces an elongated character.[42] Interactions with solvent are expected to drive the formation of these structures by burying the core domains and leaving exposed the termini as an outer shell. The hydropathy plot of *L. Hesperus* MaSp proteins are similar, and show the more hydrophilic residues in the termini and the more hydrophobic ones in the core, but the core repeat has a greater distribution of hydrophilic residues than silkworms.[13] A proposed mechanism for the assembly of silkworm silk, and further extended to spider silk, relies heavily on this similarity. As the dope is forced through the narrowing duct leading to the spinneret, removal of water, lowered pH, and ion exchange would drive these micelle like structure into oblong fiber-like structures (**Figure 4**).[43] Shear force applied during silk spinning could cause these larger globules to lose their symmetrical spherical character in favor of prolate spheroids which then pack together along a common axis on their way to forming a final solid silk fiber.[42] The stacking of these structures then would give rise to the final fiber where an acidic pH gradient in the duct is thought to induce fibrillization.[42-46]

**Mechanical Properties of Silk Fibers**

Understanding the mechanical properties of spider silks and their correlation to the underlying atomic, molecular and nanoscale structures of the silk proteins is a critical step to generating synthetic spider silk fibers with a range of desired physical and mechanical properties. Much of the interest in the material properties of spider silk is founded on testing Ma silk - one of the toughest materials known having an incredible strength for a biopolymer fiber compared to

**Table 1.** Mechanical properties of spider dragline silk and other common materials.[ref 47, 53-55]

| Material | Strength (MPa) | Stiffness (GPa) | Toughness (MJ $m^{-3}$) |
|---|---|---|---|
| **Dragline Silk** | 140-1,600 | 3-10 | 16-350 |
| **Bombyx mori Silkworm Silk** | 500-600 | 7 | 70 |
| **Elastin** | 2 | 0.00011 | 1.6 |
| **Tendon Collagen** | 120 | 1.2 | 6 |
| **Wool** | 200 | 2-3 | 60 |
| **Kevlar** | 3600 | 130 | 60 |
| **Nylon 06** | 750-950 | 2-3.6 | 80 |
| **High tensile steel** | 1650 | 190-210 | 6 |
| **Carbon fiber** | 4000 | 300 | 25 |
| **E-glass fiber** | 2200-3580 | 63-83 | - |

man-made materials.[47, 48] However, AC silks have both high strength and extensibility making them significantly tougher.[8, 49] The wide range of spider silk's mechanical properties makes it a highly versatile and tunable biopolymer.[43, 50-53] This broad range is compared with other common fibers and materials (**Table 1**).[48, 54-56] It may be useful therefore to discuss some of the terms – strength, extensibility, Young's modulus (stiffness), and toughness – commonly used to classify these fibers.

To test the mechanical properties of a single fiber, an increasing tensile force is applied to the fiber while the elongation of the fiber is measured. This is plotted as force per unit area (stress) as a function of percent elongation (strain). An example of this stress/strain curve is shown in **Figure 6**. The strength of a material is defined as the stress at fracture. The extensibility of a material is defined as the strain at fracture. The stiffness of a material, also called the Young's modulus, is the slope of the initial portion of the curve (**Figure 6,** arrows). The toughness of the

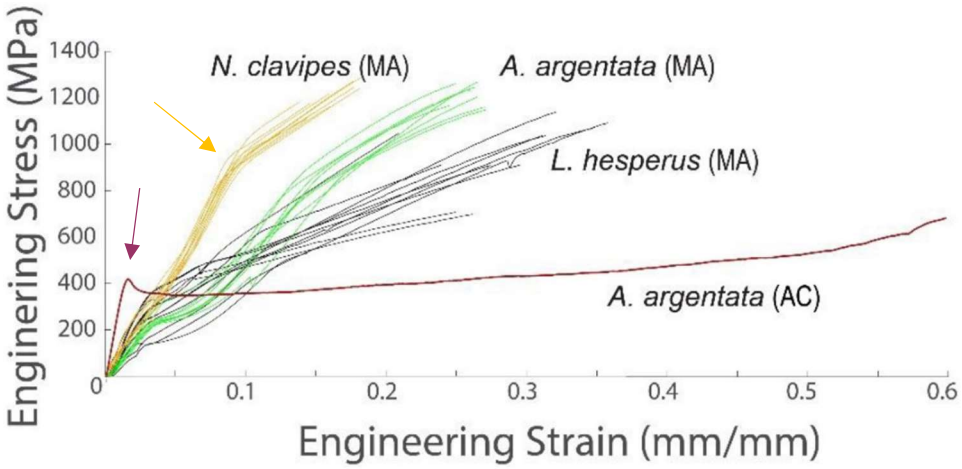



**Figure 6.** Mechanical stress-strain curves for dragline (major ampullate, MA) silk and prey wrap (aciniform, AC) silk collected from different species of spider. The strands were tested multiple times to account for variations across different individuals. The yellow and brown arrows indicate two examples of the yield point, or end of the linear region which defines the Young's modulus (stiffness).

material is the area under the curve. These material properties for dragline silk change substantially if the fibers are wet or in humid conditions.[57] When the fibers become wet they increase in diameter and shrink in length, a phenomenon referred to as supercontraction where the fiber displays enhanced extensibility.[58-60]

**Recombinant Silk Fibers**

For decades, the goal of the silk community has been to commercialize processes of spinning spider silk-like materials that are similar in mechanical properties to their native counterparts. Synthetic production of recombinant spider silk that effectively mimics the mechanical properties of its natural counterpart suffers from several key knowledge gaps. These gaps lie between the molecular/biochemical level understanding of the protein structure and how these translate to the nano- and macroscale material properties. The challenge here lies in the fact that we do not fully understand the molecular and physical state of the Sp in the dope and the processes involved in the formation of silk as it transitions from a soluble protein solution to a nanocrystalline fiber. A fiber that contains a broad range of secondary structures and a combination of highly ordered, partially ordered, and disordered domains between the nano- to

macro-scale lengths. Here we discuss the utility of recombinant silks and notable attempts to generate materials that are similar in their mechanical properties to native fibers.

Due to the large molecular weight of native silk proteins and their propensity for aggregation, several aspects of recombinantly expressing the proteins are met with difficulty. First, the DNA plasmid needed for recombinant expression is not available commercially due to the repetitive nature of the nucleic acids. These plasmids are generally synthesized in-house, using a stepwise approach where the repetitive regions are copied and cloned back into the plasmid again and again until reaching the desired number of repeats.[61-63] This is a laborious process. Several silk types have been recombinantly expressed using this method, including cylindrical (Cy), tubuliform (Tu), minor ampullate (Mi), and major ampullate (MaSp1 and MaSp2) silks which are low complexity and repeated several times in native sequences.[61-64] Other silk protein types with a greater complexity have also been expressed using commercially available plasmids, including prey-wrapping (aciniform, Ac) silk.[39, 65] To convert the silk dope into fibers, a general process includes concentrating the recombinant proteins in a solvent or buffer to create a dope, and extruding the dope through a small orifice into a coagulation bath. Generally, there are two methods to produce these materials. One is to use an organic solvent to dissolve the proteins, the other is to use native-like solutions.[66] Both have their merits and below we discuss some of the benefits and groups utilizing them.

Several groups are working on the spider silk problem and have made substantial contributions to the understanding of the spinning process itself. The Kaplan group at Tufts University, Boston has a number of projects related to silk materials from 3D printed fibers, silk-derived tools, and silk-based therapeutics.[61, 67-72] Although the Kaplan Lab focuses on generating silk-based materials, they do not utilize a native-like spinning process. To generate fibers with properties that are similar to native fibers, recombinant silk proteins are dissolved into hexafluoroisopropanol (HFIP), a strong denaturant, which is then extruded into an aqueous coagulation bath containing alcohol.[61, 73-75] The benefit of this method is that HFIP can easily

**ADF-3**

| Module A | Module Q |
| --- | --- |
| GPYGPGASAAAAAAGGYGPGSGQ | GPGQQGPGQQGPGQQGPGQQ |

**Figure 7.** Amino acids sequences of designed silk modules from the Scheibel group.

dissolve large, aggregate-prone proteins and generate materials that are similar in mechanical properties to native silks.[61, 71] Of course, HFIP is not part of the mechanism of native silk formation and also suffers from being rather expensive, corrosive and toxic. This prevents this process from being scalable or green. Together, the Kaplan Lab has explored new frontiers for silk and has developed a strong understanding of working with these large proteins, generating materials that are silk-based for materials applications and therapeutics.

Using an alternative approach, several groups utilize native-like conditions in processing silk materials into fibers and other materials.[65, 76-80] Because of this, the resulting fibers are able to be directly compared to the native fibers. This method precludes the use of expensive, toxic solvents making it scalable and green for commercialization. One example is the Scheibel group from Bayreuth University in Germany who utilizes a MaSp2-like protein that originates from the *Araneus diadematus* species, dubbed ADF-3, to explore the mechanism of silk fiber formation.[63] The protein itself is made from two modules, A and Q (**Figure 7**), repeated several times.[63] This material has been used as the basis for several studies, again using native-like spinning conditions, to explore the effects of buffers, extrusion, post-stretching, and repeat length.[77, 78, 81, 82] In one paper, the group simultaneously examines the influence of spinning conditions and protein length on mechanical properties where greater toughness is achieved for fibers spun from phosphate-containing buffers over Tris/HCl buffers.[78] This result is in line with other current work examining the effect of phosphate-containing buffers to encourage liquid-liquid phase separation which drives protein-protein interactions forming large, hierarchical protein pre-assemblies.[42, 82-84] The Scheibel group has also been successful in commercializing silk fibers through their

company. AMSilk (www.AMSilk.com). Together, the Holland and Scheibel Labs are both interested in exploring pre-assembly as a key mechanism in spider silk formation, as well as the molecular structure of the resulting fibers. In Chapter 5 of this dissertation, we explore the secondary structure of the recombinant $(AQ)_{12}NR3$ protein solubilized under native-like conditions, as well as the subsequent fibers provided during our collaboration.

## Biophysical Tools for Characterizing Silk Proteins and Fibers

### Introduction to Nuclear Magnetic Resonance Techniques for Proteins

Nuclear magnetic resonance (NMR) spectroscopy is a technique that is ubiquitously applied throughout all fields of chemistry, structural biology, and material science. It is the only technique that provides atomic-level detail of biomolecular systems in solution. Certain atoms on the periodic table have a property called “spin”, arising from the quantized nuclear angular momentum which gives rise to a nuclear magnetic moment. The proportionality between the nuclear angular momentum and the nuclear magnetic moment is called the gyromagnetic ratio. When a strong magnetic field is applied to a nuclear spin, it begins to precess at a frequency

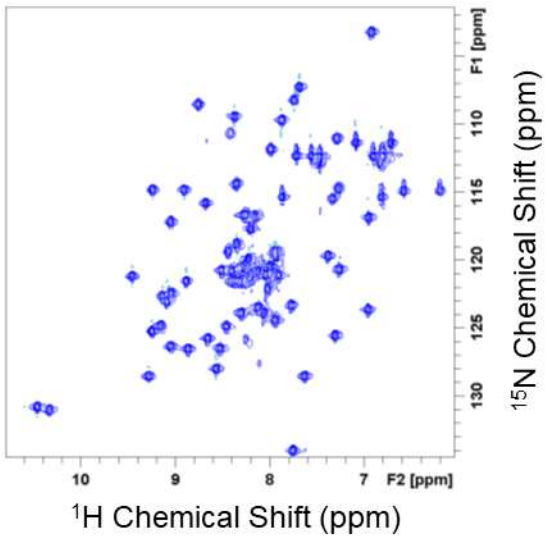



**Figure 8.** $^{1}H$-$^{15}N$ HSQC of uniformly-labeled $^{13}C/^{15}N$ GB1. GB1 is well-folded 56 residue protein that serves as a setup sample for solution NMR. With the exception of a few amide side-chains, each peak corresponds to an H-N bond for each individual amino acid residue of the protein.

proportional to the strength of the magnetic field based on their gyromagnetic ratio. This frequency from the spin is known as the Larmor frequency. The frequency is proportional to the energy associated with the interaction between the nuclear magnetic moments and the magnetic field. However, the effective magnetic field felt by each individual nuclei is attenuated by the influence of other nearby electronic environments. Because of this, the frequency of these spins vary depending on the atomic bonding environment. For example, the connectivity of carbon atoms in the molecule ethanol ($CH_3CH_2OH$) produces two different resonances; one corresponding to the $CH_2$ bound to the –OH group, and the one from the $CH_3$. To record this difference in frequency, a second oscillating magnetic field is applied as a "pulse" to the sample which excites the spins and promotes them from low to high energy, allowing one to record the frequency of energy given off as the spin relaxes from high to low energy. This pulse sequence is repeated several times, where the signals are averaged to improve signal-to-noise (S/N). The data are recorded as frequency, but with one additional step to account for the strength of the magnetic field. Since the precession frequency of nuclei change with magnetic field strength ($^1H$ at 600 MHz precesses at 600 million times/sec vs. $^1H$ at 400 MHz precesses at 400 million times/sec), it is useful to make the frequency values obtained relative to the magnetic field so it becomes easier to compare data across different field strengths. This is done with the implementation of the "chemical shift", which is in parts per million (ppm) and is calculated by dividing the obtained frequency by the magnetic field strength, multiplied by one million. For example, if a $^1H$ frequency is recorded as 1800 Hz on a 600 MHz spectrometer, the chemical shift would be 3 ppm (0.001800 MHz * 1e6/600e6 MHz = 3 ppm). On a 400 MHz instrument, 3 ppm = 1200 Hz (3 ppm * 400e6 MHz/1e6 = 1200 Hz). In the case of amino acids, these chemical shifts are helpful in solving protein structure questions. For example, the chemical shifts of the atoms ($^1H/^{13}C/^{15}N$) within an amino acid are influenced by its phi/psi angles. After obtaining the chemical shifts for each amino acid in a protein, one can compare them to a set of published chemical shifts to determine if the amino acid is in a helical, random coil, or β-sheet structure.[85]

## 2D NMR and Backbone Assignment Strategies

The previous example is of a 1-dimensional experiment where only a single frequency is recorded. However, there are modified versions of these experiments allowing us to record two, three, or even more nuclei simultaneously in a single experiment. For example, the most common two-dimensional (2D) NMR experiment utilized in a number of fields is the heteronuclear single quantum coherence (HSQC) experiment (**Figure 8**).[86] In this experiment, $^{1}H$ nuclei are excited and through spin-spin coupling J-transfer their magnetization to $^{13}C$ nuclei (or any other active nuclei) one bond away are also observed. A $^{1}H$-$^{13}C$ bond pair is displayed as a single peak, where the chemical shifts of the $^{1}H$ and $^{13}C$ are on different axis. This experiment is often faster at generating data than a direct experiment because of the utilization of $^{1}H$ magnetization transfer making it possible to collect data on natural abundance samples. However, this mapping technique is rarely sufficient to confidently assign residues in proteins. Furthermore, sequential

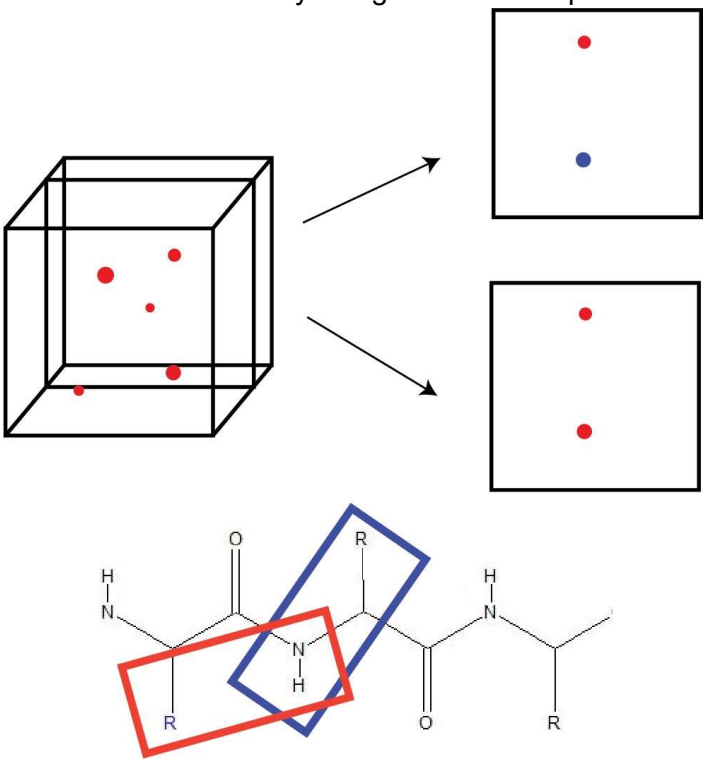



**Figure 9.** Graphical representation of the 3D backbone walk strategy. The cube contains data collected from an HNCACB experiment. The extracted slice from the cube is at a $^{15}N$ intersection, and correlates the HNCACB experiment (blue) with data from the CBCACONH experiment (red).

assignments are not possible with only two dimensions. To solve this, there are a variety of triple resonance ($^{1}H/^{13}C/^{15}N$) 3D experiments that pass magnetization along the backbone thus, providing chemical shift information from neighboring amino acids.

First, a $^{1}H/^{15}N$ HSQC is generally recorded to map the dispersity of backbone $^{15}N$ amides. This is then followed up with a number of 3D experiments that pass magnetization along the backbone to sidechains and provide greater specificity when doing assignments (**Figure 10**).[87] One example of this is the HNCACB experiment, where magnetization is passed as follows: $^{1}H$ > $^{15}N$ > $^{13}C\alpha$ + $^{13}C\beta$ > $^{15}N$ > $^{1}H$ (**Figure 9**).[88, 89] The resulting resonances correlate $^{1}H$ and $^{15}N$ with $^{13}C\alpha$ and $^{13}C\beta$ peaks, including a weak set of resonances that bleed through from the preceding (i-1) residue. A second experiment, called the CBCAcoNH, passes magnetization from the amide

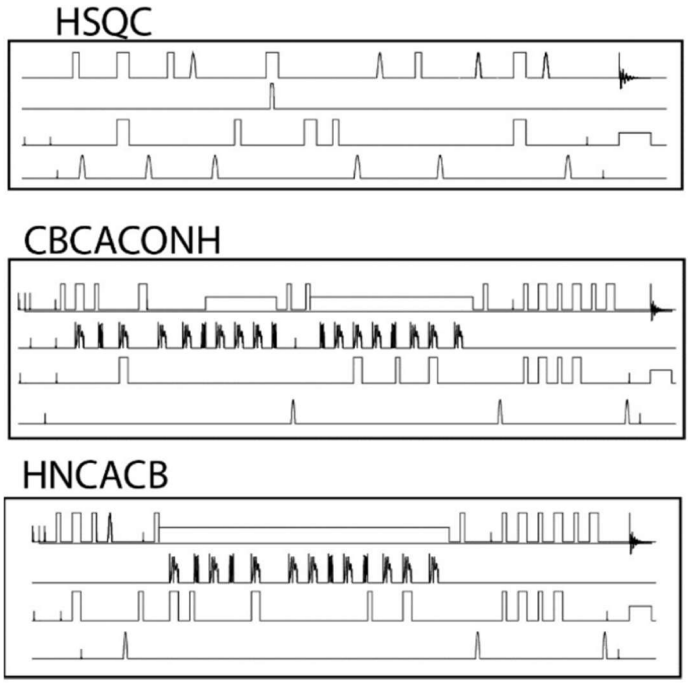



**Figure 10.** Pulse sequences used for the “backbone walk” strategy to correlate sequential peaks for protein sequencing.

$^{1}H$ through the adjacent CO group to correlate the i-1 $C\alpha$ and $C\beta$ resonances. These $C\alpha$ and $C\beta$ peaks are then matched with the i residue from the HNCACB experiment and used to identify the

$^{15}$N chemical shift of the i+1 amino acid (**Figure 10**).[90, 91] This "backbone walk" is continued until the entire protein can be assigned.

**Protein Dynamics**

Segmental processes occur on a wide timescale for proteins, from aromatic ring flips (ms-μs), to full-body tumbling (μs-ns), to the wiggling and jiggling of local bond vectors (ns-ps). These motions cover several orders of magnitude and their perturbations can be a diagnostic of protein folding, protein-protein and/or protein-ligand interactions. Measuring these in both the liquid and solid-state therefore, helps to further reveal atomic information regarding the protein through measured local molecular dynamics. Technically in these cases, what is measured is the spins returning to equilibrium. Nuclear spin relaxation has two components – spin-lattice relaxation ($T_1$), and spin-spin relaxation ($T_2$) – and is perturbed by the aforementioned processes. Measuring the rate at which this signal intensity returns to equilibrium, or decays, provides insight into these motional processes. Other experiments can exchange nuclear magnetization and are distance dependent, like the Nuclear Overhauser Effect (NOE), and can therefore be used to determine structural restraints for solving protein structures. Finally, we can measure bulk (Brownian) dynamics of particles in solution using pulsed-field gradient diffusion oriented spectroscopy (PFG-DOSY), giving us clues to protein-protein interactions that occur at the nanoscale.

**Relaxation**

Typically when spectroscopists refer to NMR relaxation they are referring to $T_1$ and $T_2$ relaxation (**Figure 11**). Broadly speaking, the motional processes that occur with $T_1$ report on the overall tumbling, or reorientation of the molecule in ps-ns regime. $T_2$ reports on slower, local motions such as chemical exchange, side chain rotations or flexible loop motions that occur on the μs-ms timescale. $T_1$ relaxation is usually measured using the "inversion recovery" method, where a 180° pulse is applied to the sample, followed by an incremented delay, then a 90° pulse.[92] The result is that net magnetization grows in the z-direction during the incremented delay that

becomes tipped into the plane of detection when the 90° pulse is applied. The signal builds over the course of the experiment, reaching a plateau. The data is fit to the equation:

$$I = I_0(1 - 2e^{-\frac{t}{T_1}})$$

where *t* is the incremented time and $I_0$ is the signal intensity at the given time (**Figure 11, A**). $T_2$ relaxation is measured using the Carr-Purcell-Meiboom-Gill (CPMG) experiment, where a 90° pulse is applied to the spin of interest followed by an array of (delay-180°)$_n$ pulses.[93, 94] The signal decays as a function of (delay-180°)$_n$ pulses, and is fit to an exponential decay:

$$I = I_0(e^{-\frac{t}{T_2}})$$

where *t* is the delay time (**Figure 11, B**). $T_1$ relaxation is most effective (lowest) when molecular tumbling is at the Larmor frequency and increases as the sample either becomes more liquid, or solid. $T_2$ relaxation is governed by these same motional processes, as well as other more local processes and decreases as sample tumbling slows. Together, both of these processes can be helpful in determining dynamic changes within proteins and provide information of protein folding, protein-protein interactions and ligand-protien interactions.

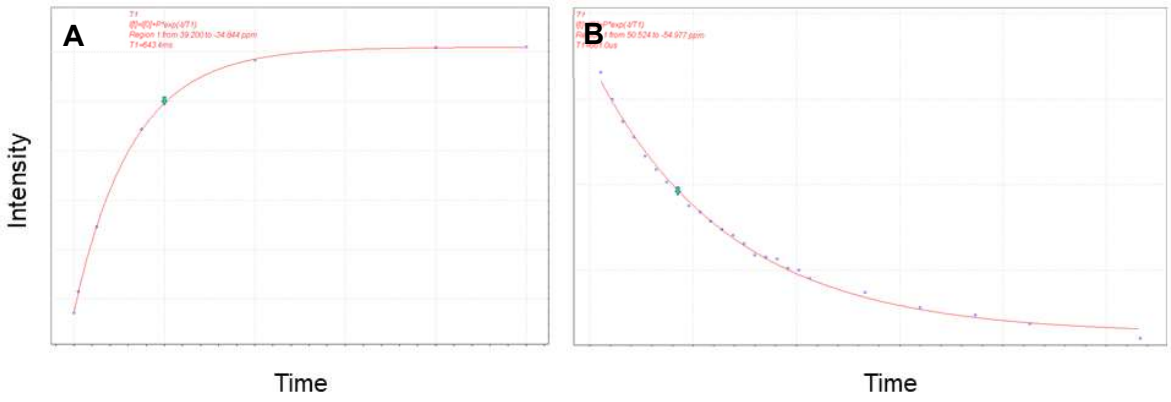



**Figure 11.** Example $T_1$ (A) and $T_2$ (B) curves. $T_1$ is determined using the "inversion-recovery" method where signal builds over time. $T_2$ is measured using the CPMG sequence where signal decreases over time. Both processes follow first-order kinetics and can be fit by an exponential.

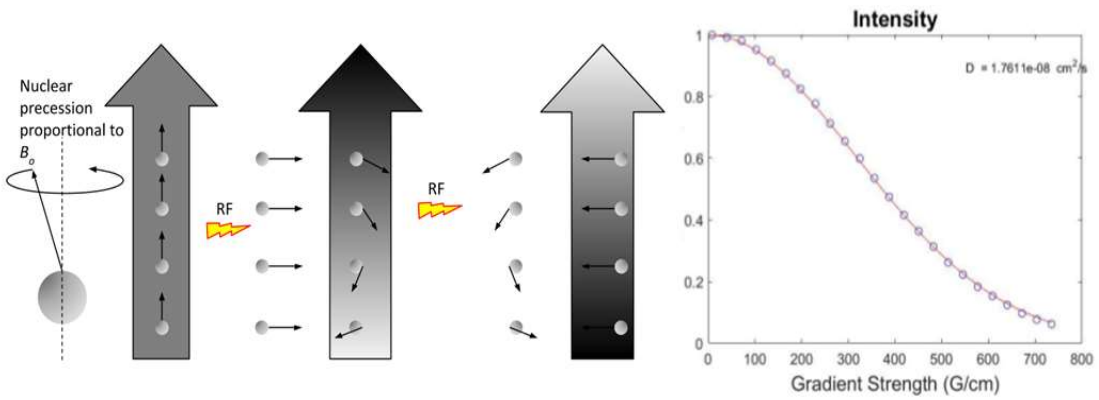



**Figure 12.** PFG-DOSY experiment scheme is displayed on the left. Data taken measuring the diffusion of glycerol is displayed on the right.

**Diffusion**

One class of experiments that can be used to extract bulk dynamics are pulsed field gradient (PFG) diffusion NMR measurements. These experiments inform on the Brownian diffusion of molecules or proteins in solution which through the Stokes-Einstein relationship can be directly correlated to the size of these molecules or assembly of molecules. In its simplest form, the experiment starts with a standard 90° pulse and then a non-uniform gradient field is applied across the sample which spatially encodes the particle's location within the NMR tube (**Figure 12**). After a delay allowing the particles to move, their location is then decoded and compared to their original signal intensity. This experiment is repeated several times with increasing gradient strengths where the signal attenuation of each experiment is fit to an exponential decay known as the Stejskal-Tanner equation, allowing us to solve for the diffusion coefficient (*D*):

$$S = S_0 \exp\left[-(\gamma\delta g)^2 D\left(\Delta - \frac{\delta}{3}\right)\right]$$

where $\gamma$ is the gyromagnetic ratio of the nucleus, $\delta$ is the length of time the gradient pulse is applied, *g* is the strength of the gradient pulse, $\Delta$ is the time in between the two pulses, and *D* is the diffusion coefficient.[95] The final diffusion coefficient is used in the Stokes-Einstein equation to solve for particle size (**Figure 17**).[96]

**Solid-state NMR**

While solution NMR is implemented to study silk dope under various native-like conditions, we are unable to use this technique to structurally characterize solid fiber samples. This is due to the rigid nature of the material causing strong dipolar-coupling between adjacent spins leading to rapid signal decay and broad resonances. In order to remove this dipolar-coupling, we use a solid-state NMR probe equipped to spin the sample at high speeds (5-110 kHz) at an angle of 54.74° relative to the $B_0$ field, dubbed magic angle spinning (MAS). In doing so, the dipolar network is

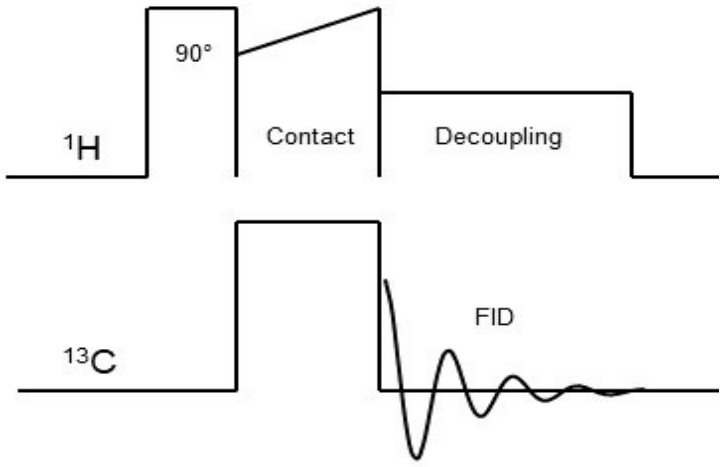



**Figure 13.** Typical cross-polarization (CP) pulse sequence scheme. First, a 90° pulse is applied to the abundant nuclei to tip the spins into the plane of detection. Then, power is applied to both channels to achieve a Hartman-Hahn match, where the spin-lock fields ($B_1^H$ and $B_1^C$) are adjusted so the splitting for $^1$H and $^{13}$C become equal ($\omega_H = \omega_C$). This helps increase signal from the less abundant nuclei. During signal acquisition, the abundant nuclei is decoupled to prevent its interaction from the collected FID signal.

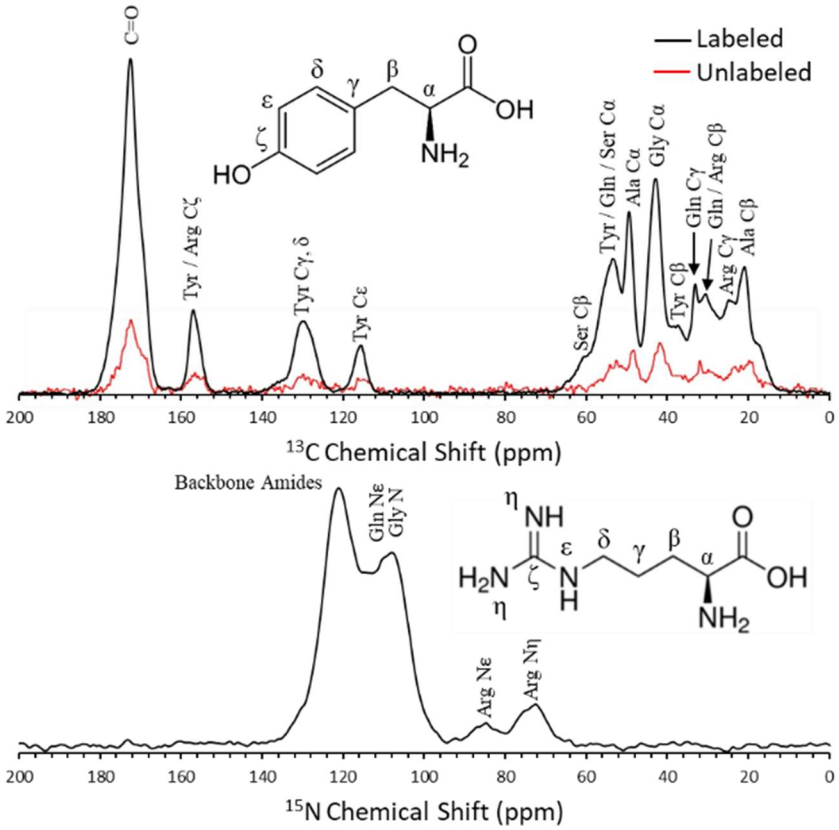



**Figure 14.** Cross-polarization magic-angle spinning (CP-MAS) results from *N. clavipes* (Golden Orb Weaver) dragline silk enriched with uniformly-labeled $^{13}C/^{15}N$-Tyr/Arg. With site-specific labeling, we can enrich particular nuclei to boost their signal. After measuring the chemical shifts of each major peak, we can compare these with published values to determine secondary structure.

greatly reduced and allows us to collect chemical shift information on the structure of solid materials. Some NMR nuclei are 100% abundance and have a high gyromagnetic ratio allowing us to collect data with a single direct pulse. However, several nuclei of interest involved with proteins ($^{13}C$, $^{15}N$) have not only a low abundance, but also have gyromagnetic ratios that are low necessitating more scans resulting in very long experiment times. To increase S/N for low-gyromagnetic ratio nuclei, a cross-polarization (CP) step is added where magnetization is transferred from an abundant nuclei ($^{1}H$) to the nuclei of interest ($^{13}C/^{15}N$) over a contact period.[97, 98] This is done by matching the frequency of the two spins by modifying their spin-lock fields with an applied magnetic field ($B_1$). The pulse sequence first begins with a 90° pulse on the abundant nuclei ($^{1}H$), followed by the application of frequencies on both channels such that they are spin-

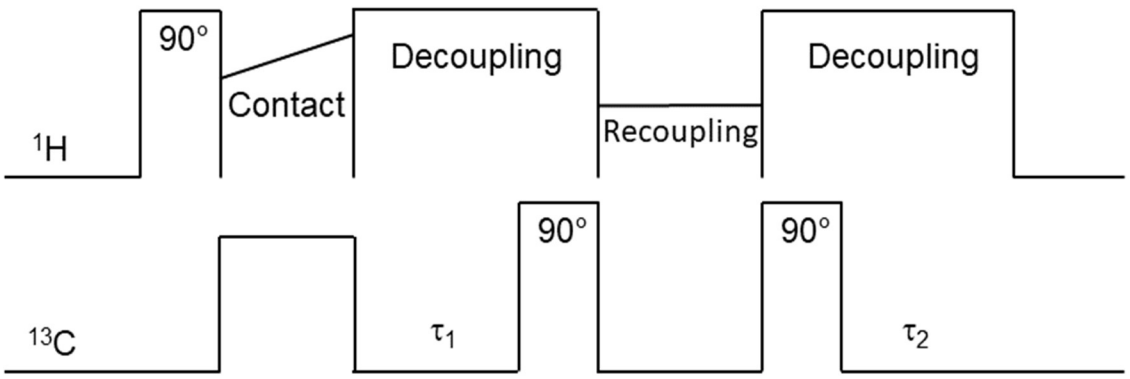



**Figure 15.** Dipolar-assisted rotational resonance (DARR) pulse sequence scheme. After the CP step, signal is collected for the direct dimension and after the spins are inverted where they are recoupled over a mixing period. Then, the signals are reinverted to fall into the plane of detection and detected.

locked. In practice, this is done by fixing the frequency of one channel and "ramping" the frequency of the other in order to excite all nuclei in the frequency range. After this spin exchange, the abundant spin is decoupled from the signal being read (**Figure 13)**. Theoretically the signal enhancement is proportional to gyromagnetic ratios of the two spins. For example, a $^{1}$H-$^{13}$C CP has an enhancement of ~4X since the $^{13}$C is approximately ¼ the gyromagnetic ratio of $^{1}$H (**Figure 14**). In this experiment, we collected $^{1}$H-$^{13}$C and $^{1}$H-$^{15}$N CP MAS data on labelled and unlabeled *N. Clavipes* dragline silk after feeding the spiders with U-$^{13}$C/$^{15}$N-Tyr (**Figure 14**).

Because dipolar interactions are distance-dependent, it is possible to obtain distance information between nuclei by utilizing certain experiments. The dipolar coupling equation describes the relationship between two nuclei:

$$\frac{\mu_0 \gamma_j \gamma_k \hbar}{4\pi \langle r_{jk}^3 \rangle} (1 - 3cos^3\theta)$$

Where $\mu_0$ is the vacuum permeability, $\hbar$ is Plank's constant divided by 2π, $\gamma_x$ are the gyromagnetic ratios of each spin, $r_{jk}^3$ is the internuclear distance between the spins, and $\theta$ is the angle between the two spins with respect to the magnetic field. At the magic angle ($\theta$ = 54.7°) the angular dependence of the dipolar coupling is reduced to zero and signals corresponding to the chemical

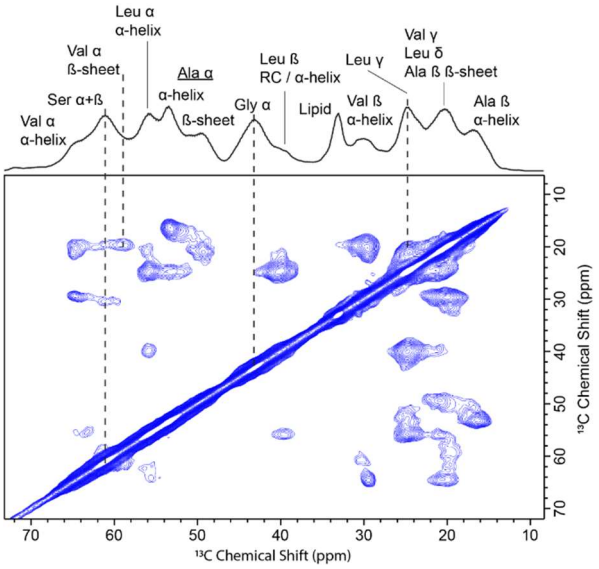



**Figure 16.** 2D dipolar-assisted rotational resonance (DARR) spectra at 14 kHz MAS of native *A. argentata* prey wrap silk labeled with $^{13}$C-Val/Leu/Ala. At a mixing time of 50 ms, short-range contacts can be established to determine amino acid secondary structure, as well as labeling efficiency.

shift dominate. A popular solid-state experiment that exploits this is the dipolar-assisted rotational resonance (DARR) experiment (**Figure 15, 16**).[99, 100] In this experiment, magnetization is transferred from $^{13}$C to $^{13}$C nuclei (or any homonuclear pair), where an adjustable mixing period is added to measure magnetization exchange between spins that can be correlated with distances. This $^{13}$C-$^{13}$C interaction is assisted with the addition of a recoupling scheme, where $^{1}$H-$^{1}$H and $^{1}$H-$^{13}$C dipolar coupling is reintroduced to enhance magnetization exchange. Our group utilizes this experiment regularly to resolve overlapping resonances in CP-MAS NMR spectra of spider silks (**Figure 16**). Together, both the CP and DARR experiments help to answer a number of questions about the overall structure of solid samples and have been utilized throughout several projects in this dissertation.

**Dynamic Light Scattering**

Dynamic light scattering (DLS) is a technique used to measure the size of particles in solution and has been applied to numerous materials including polymers and proteins.[101] When light, usually monochromatic from a laser, is shined on particles in solution they scatter light in proportion to their radius.[101] Because the particles are undergoing Brownian motion, their distance

$$D = \frac{kT}{6\pi\eta R}$$

**Figure 17.** Stokes-Einstein relationship equation for measuring the diffusion of particles in solution, where *k* = Boltzmann constant, T = temperature in Kelvin, η = viscosity of solution, and R = radius of the particle. If the diffusion is known, the equation can be algebraically rearranged to solve for R.

to the detector changes over time and produces fluctuations in this scattered light intensity which causes constructive or destructive interference producing light or dark spots, respectively, on a photodetector.[102] These spots are then correlated to the movement of particles in solution through an auto-correlation function.[103] The data is then deconvoluted to measure the size of particles using a cumulant fit analysis, where the changes in intensity over time are used to calculate diffusion.[104] Finally, using the Stokes-Einstein relationship it is possible to calculate the size of particles in solution (**Figure 17**).[96]

Due to the sensitivity of the detectors implemented in most benchtop DLS systems, it is possible to measure the size of particles present at part-per-million levels (Malvern Zetasizer, www.malvern-talks.com). However, spider silk proteins are stored at native conditions between 25-50 wt%.[105-107] This is far outside the appropriate range of obtaining quality data, where most manufacturers suggest an upper-limit of approximately 1 wt%. If the concentration used is too high, the scattered light and particles interact faster than the detector can record which leads to erroneous data. It is therefore necessary to first measure the size of these particles at a number of concentrations to determine the upper concentration limit.

DLS is an easy and convenient method of measuring particle sizes in solution. We have successfully utilized DLS to measure the size of spider silk proteins in solutions of 4 M urea, observing the pre-assemblies present under native conditions which were also detected by diffusion NMR and cryoEM.[84, 108] DLS has also been utilized by our group as a screening tool

before cryogenic - transmission electron microscopy (cryo-TEM), where we are interested in obtaining images of these pre-assemblies.

**Molecular Dynamics**

Molecular dynamics (MD) is a computational tool used to observe and analyze the movement of atoms and molecules. Velocities are first assigned to the atoms according to the Boltzmann distribution at an initial temperature.[109, 110] Once the simulation begins, restrains are removed and the atoms are allowed to freely move. Their movement is controlled by solving Newton's equations of motions for each atom over the course of the simulation (**Figure 18**).[111] The number of equations used during these calculations, also based on the type of system (proteins, organic molecules, metals, etc.) and solvent are combined into a force field. For example, in the popular AMBER force field, there are different terms used for simulations with small organic molecules, protein-organic molecule docking, and proteins.[112] The simulation is divided into equal time steps, where between each step the atoms are allowed to move based on their previous energy, then stopped and recalculated, which provide new energies, velocities and trajectories for the atoms for the next step. This iterative process is usually repeated using millions of steps bringing the total simulation time to the ps – ms range.[113] The number of steps chosen and therefore the length of time for each simulation is dependent on the timescale of the phenomena of interest. For example, protein folding continues to be of great interest to the

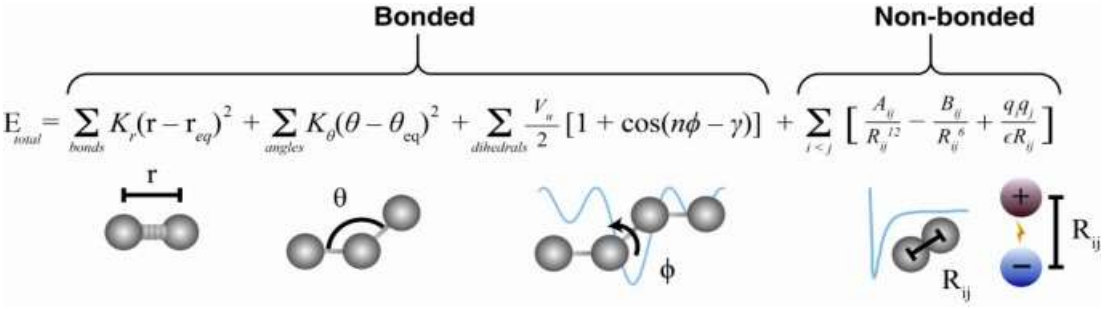



**Figure 18.** Equation used to calculate the total energy for a single atom in an MD simulations. Terms are added for each type of energy potential and include the oscillation of bonds, changes of angles between bonds, dihedral rotations, and Lennard-Jones type non-bonded potentials.[118]

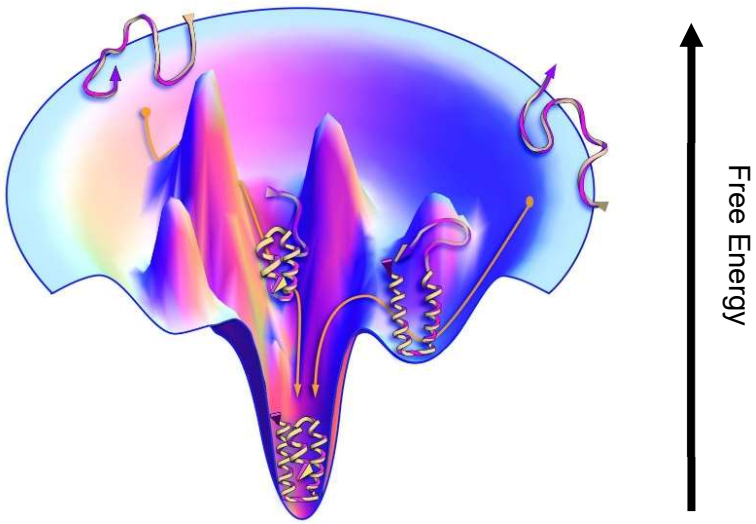



**Figure 19.** Example of a protein folding funnel showing the possible trajectories of a protein. For a well-folded protein, there is only one low-energy structure (shown here). For an aggregate-prone protein, there may be several low-energy structures (multiple wells).

scientific community since several well-known diseases arise from misfolded proteins.[114-117] MD has been utilized to understand energies associated with native and misfolded versions of the proteins, thus providing differences to the structures where a therapeutic weakness could potentially be exploited. MD has been utilized to study the folding and subsequent aggregation of amyloid-β-42 peptides which are found in plaques in the brain of patients with Alzheimer's dis ease.[118] In a recent work, the amyloid-β-42 peptide was put into a box of water and simulated for structural changes 10 times. The authors found that the appearance of a transient structure occurring periodically throughout the simulation was a trigger for fibril formation, in good agreement with a similar transient structure proposed from solution NMR.[118] The data provided throughout a trajectory can also be informative in the protein folding context, and is often used to generate energy funnels or contact maps for establishing the relationship between free energy and protein folding (**Figure 19**).

For larger systems as well as systems where the phenomena of interest occurs on a larger timescale, fully-atomistic models require exorbitant computational amounts of time to generate meaningful data. Coupled with the fact that these simulations also need to be repeated, the length

of time needed to run a simulation could take months. There are some time-saving measures that can be employed, however. For example, when dealing with a large system it's possible to use an implicit-solvent model where the dielectric of the solvent, rather than the solvent molecules themselves, are used.[119] Of course you can only use this if you are not interested in the solvent's interaction with your molecule. This dramatically reduces the time needed for simulations since the force field is only used for the molecule of interest and not the solvent molecules. Additionally, it is possible to use a coarse-grained (CG) MD approach where groups of atoms are abbreviated by a single point.[120, 121] Most CG systems are used as a "drop-in" replacement for all-atom systems, and work seamlessly with the most popular force fields, including AMBER. This approach has been used to study the nature of, for example, lipid-protein interactions as well as a number of other systems.[121, 122] In Chapter 3 of this dissertation, we simulate native-sized MaSp1 and MaSp2 proteins to understand their putative structure.

**Solid-phase Peptide Synthesis**

Peptide synthesis is utilized to make a number of significant commercial and pharmaceutical products such as the sugar substitute aspartame, the clinical hormone oxytocin, the adrenocorticotropic hormone, and stapled therapeutic peptides with enhanced specificity over small molecule drugs.[123, 124] Solid-phase peptide synthesis (SPPS) begun in the early 1960's after the synthesis of a tetrapeptide synthesized on the surface of a styrene bead and has since become a streamlined and robust way to generate small segments of proteins.[125] The next major advancement for SPPS occurred in the early 2000s with the addition of microwave-assisted

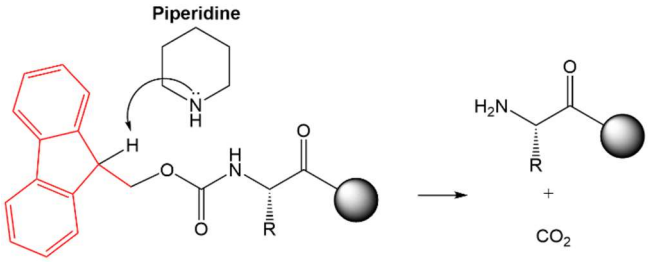



**Figure 20.** Mechanism of Fmoc deprotection.

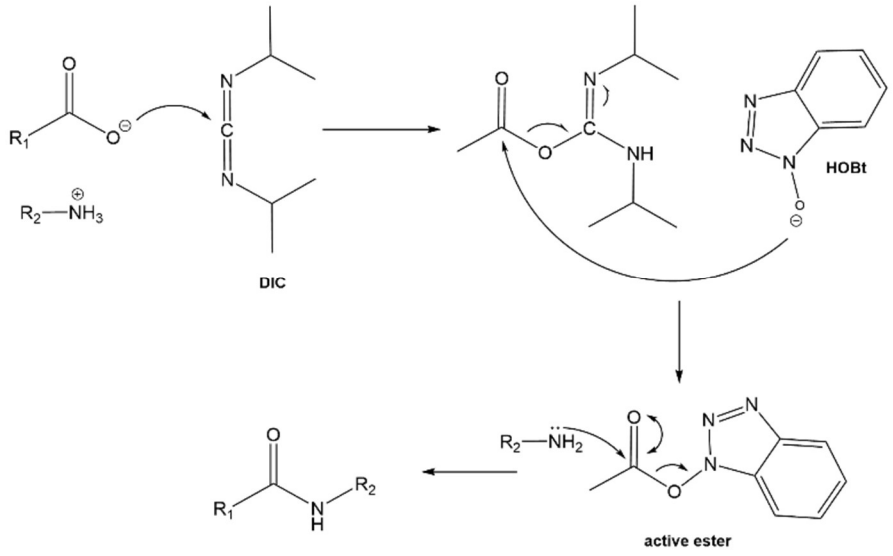



**Figure 21.** Carbonyl activation of the next amino acid and formation of the peptide bond. The incoming amino acid is reacted with diisopyopylcarbodiimide (DIC) and subsequently 1-hydroxy-benzotriazole (HOBt) leading to peptide bond formation.

synthesis which decreases time and helps to prevent peptide aggregation.[126] Current advancements include improved resin chemistries, efficient coupling schemes for growth, and optimization of activator/activator bases which work to improve purity and yield.[127] There are several competing peptide synthesizer manufacturers including CEM Corporation, Gyros Protein, AAPPTEC, and CSBIO to name a few. All utilize similar FMOC chemistry; adding amino acids, one by one, to a growing peptide chain. The general steps of SPPS have been relatively unchanged over the last decades, however modifications are made when problems arise that stem from the primary sequence, like resin choice and coupling schemes.[128, 129]

SPPS first begins with the selection of resin, or bead, from which the peptide will be grown. The resin is a small (100s μm) polystyrene sphere with linkers that decorate the surface.[123, 125] Peptides are synthesized in the C-to-N direction, where the type of linker bound to the resin determines what group will be present at the C-terminal of the final peptide. For example, if a carboxylic acid is desired at the C-terminus of the final peptide, a Wang resin would be used.[130] The linkers come protected and the first step in SPPS is to deprotect the linker in order to add the first amino acid. The protecting group chosen for the linker, as well as for the N-terminal amines

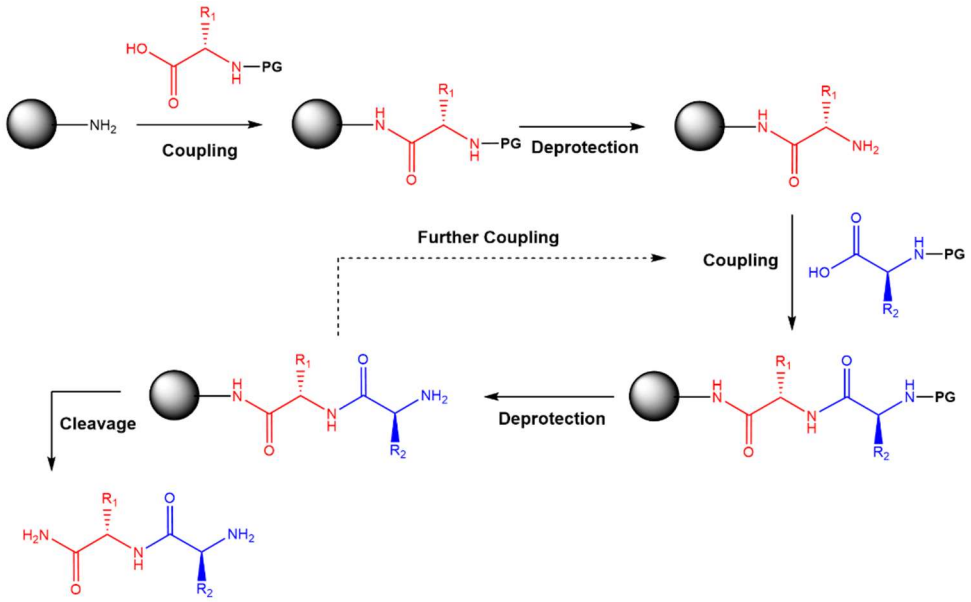



**Figure 22.** Overall scheme of solid-phase peptide synthesis coupling. The resin is deprotected allowing the first amino acid to be attached. The first amino acid is then deprotected to allow the incoming activated amino acid to form a peptide bond. The cycle is then repeated – deprotection, activation, coupling – until the peptide is complete. Cleavage is the last step in synthesis which removed the peptide from the solid support.

of the amino acids, is usually the base-labile molecule 9-fluorenylmethoxycarbonyl (Fmoc).[131] This protecting group Fmoc is removed by the addition of a strong organic base such as piperdine or piperazine.[128] The acidic proton on the five member ring of the Fmoc group (**Figure 20**) is remove by the base causing a bond rearrangement exposing the N-terminal of the linker/amino acid. The incoming amino acid is reacted with the activator base diisopyopylcarbodiimide (DIC), a good leaving group, which subsequently prepares it for the attack from the activator 1-hydroxy-benzotriazole (HOBt), creating a highly reactive species that is attacked by the exposed N-terminal (**Figure 21**), creating a peptide bond. The reaction vessel where this takes place is washed to remove excess reagent, and the process repeats; the Fmoc group of the recently added amino acid is removed followed by the addition of the next amino acid, activator, and activator base which covalently attach it to the growing peptide chain (**Figure 22**). After the last amino acid is added, a final cleavage step removes the peptide from the resin. The cleavage cocktail is a mixture of acid and scavengers whose additional role is to remove the protecting

groups covalently linked to the side chains of amino acids containing N or O atoms.[132, 133] Generally, cleavage cocktails contain > 80% trifluoroacetic acid with water and triisopropylsilane as scavengers.[127, 134] Depending on the primary sequence, it may be necessary to include reducing agents for Cys residues.

In practice, there are usually some modifications that are needed to optimize the synthesis. For example, our group uses SPPS for synthesizing silk-like peptides that are aggregate prone. One early modification in our SPPS work was to use a low-loading resin, which has fewer linker sites to prevent aggregation during synthesis. Additionally, some of the amino acids which contain large protecting groups, like Fmoc-Arg-Pbf, need to utilize a double-coupling scheme.[134] No deprotection occurs between each step; first the Arg is added and reacted, washed away, then a second aliquot of Arg is added without a deprotection step. The makeup of cleavage cocktail as well as the time necessary to cleave and deprotect the peptide is also dependent on the primary sequence. With large protecting groups, longer times are required for more than one or two instances of the amino acid. Together, SPPS has become an effective way to generate silk-like materials our group uses to study fundamental aspects of the spider silk spinning process. In Chapter 4, we synthesize silk peptide mimics to study the native pre-assembly structure of native silk proteins.

**References**

1. Hakimi, O.; Knight, D. P.; Vollrath, F.; Vadgama, P., Spider and mulberry silkworm silks as compatible biomaterials. *Composites Part B: Engineering* **2007,** *38* (3), 324-337.

2. Kuhbier, J. W.; Reimers, K.; Kasper, C.; Allmeling, C.; Hillmer, A.; Menger, B.; Vogt, P. M.; Radtke, C., First investigation of spider silk as a braided microsurgical suture. *J Biomed Mater Res B Appl Biomater* **2011,** *97* (2), 381-7.

3. Chaw, R. C.; Hayashi, C. Y., Dissection of silk glands in the Western black widow Latrodectus hesperus. *Journal of Arachnology* **2018,** *46* (1), 159-161.

4. Jeffery, F.; La Mattina, C.; Tuton-Blasingame, T.; Hsia, Y.; Gnesa, E.; Zhao, L.; Franz, A.; Vierra, C., Microdissection of black widow spider silk-producing glands. *J Vis Exp* **2011,** (47).

5. Ayoub, N. A.; Garb, J. E.; Kuelbs, A.; Hayashi, C. Y., Ancient properties of spider silks revealed by the complete gene sequence of the prey-wrapping silk protein (AcSp1). *Mol Biol Evol* **2013,** *30* (3), 589-601.

6. Tokareva, O.; Jacobsen, M.; Buehler, M.; Wong, J.; Kaplan, D. L., Structure-function-property-design interplay in biopolymers: spider silk. *Acta Biomater* **2014,** *10* (4), 1612-26.

7. Yarger, J. L.; Cherry, B. R.; van der Vaart, A., Uncovering the structure–function relationship in spider silk. *Nature Reviews Materials* **2018,** *3* (3), 18008.

8. Hayashi, C. Y.; Blackledge, T. A.; Lewis, R. V., Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. *Mol Biol Evol* **2004,** *21* (10), 1950-9.

9. Barbara A. Lawrence; Craig A. Vierra; Moore, A. M. F., Molecular and Mechanical Properties of Major Ampullate Silk of the Black Widow Spider, Latrodectus hesperus. *Biomacromolecules* **2004,** *5*.

10. Chaw, R. C.; Correa-Garhwal, S. M.; Clarke, T. H.; Ayoub, N. A.; Hayashi, C. Y., Proteomic Evidence for Components of Spider Silk Synthesis from Black Widow Silk Glands and Fibers. *J Proteome Res* **2015,** *14* (10), 4223-31.

11. Collin, M. A.; Clarke, T. H., 3rd; Ayoub, N. A.; Hayashi, C. Y., Genomic perspectives of spider silk genes through target capture sequencing: Conservation of stabilization mechanisms and homology-based structural models of spidroin terminal regions. *Int J Biol Macromol* **2018,** *113*, 829-840.

12. Kono, N.; Nakamura, H.; Ohtoshi, R.; Moran, D. A. P.; Shinohara, A.; Yoshida, Y.; Fujiwara, M.; Mori, M.; Tomita, M.; Arakawa, K., Orb-weaving spider Araneus ventricosus genome elucidates the spidroin gene catalogue. *Sci Rep* **2019,** *9* (1), 8380.

13. Ayoub, N. A.; Garb, J. E.; Tinghitella, R. M.; Collin, M. A.; Hayashi, C. Y., Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One* **2007,** *2* (6), e514.

14. Zhang, Y.; Zhao, A. C.; Sima, Y. H.; Lu, C.; Xiang, Z. H.; Nakagaki, M., The molecular structures of major ampullate silk proteins of the wasp spider, Argiope bruennichi: a second blueprint for synthesizing de novo silk. *Comp Biochem Physiol B Biochem Mol Biol* **2013,** *164* (3), 151-8.

15. Chaw, R. C.; Zhao, Y.; W, J.; Ayoub, N. A.; Allen, R.; Atrushi, K.; Hayashi, C. Y., Intragenic homogenization and multiple copies of prey-wrapping silk genes in Argiope garden spiders. *BMC Envolutionary Biology* **2014,** *14* (31).

16. Askarieh, G.; Hedhammar, M.; Nordling, K.; Saenz, A.; Casals, C.; Rising, A.; Johansson, J.; Knight, S. D., Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature* **2010,** *465* (7295), 236-8.

17. Jaudzems, K.; Askarieh, G.; Landreh, M.; Nordling, K.; Hedhammar, M.; Jornvall, H.; Rising, A.; Knight, S. D.; Johansson, J., pH-dependent dimerization of spider silk N-terminal domain requires relocation of a wedged tryptophan side chain. *J Mol Biol* **2012,** *422* (4), 477-87.

18. Andersson, M.; Chen, G.; Otikovs, M.; Landreh, M.; Nordling, K.; Kronqvist, N.; Westermark, P.; Jornvall, H.; Knight, S.; Ridderstrale, Y.; Holm, L.; Meng, Q.; Jaudzems, K.; Chesler, M.; Johansson, J.; Rising, A., Carbonic anhydrase generates CO2 and H+ that drive spider silk formation via opposite effects on the terminal domains. *PLoS Biol* **2014,** *12* (8), e1001921.

19. Kurut, A.; Dicko, C.; Lund, M., Dimerization of Terminal Domains in Spiders Silk Proteins Is Controlled by Electrostatic Anisotropy and Modulated by Hydrophobic Patches. *ACS Biomaterials Science & Engineering* **2015,** *1* (6), 363-371.

20. Xu, D.; Yarger, J. L.; Holland, G. P., Exploring the backbone dynamics of native spider silk proteins in Black Widow silk glands with solution-state NMR spectroscopy. *Polymer* **2014,** *55* (16), 3879-85.

21. Holland, G. P.; Jenkins, J. E.; Creager, M. S.; Lewis, R. V.; Yarger, J. L., Quantifying the fraction of glycine and alanine in beta-sheet and helical conformations in spider dragline silk using solid-state NMR. *Chem Commun (Camb)* **2008,** (43), 5568-70.

22. Gray, G. M.; van der Vaart, A.; Guo, C.; Jones, J.; Onofrei, D.; Cherry, B. R.; Lewis, R. V.; Yarger, J. L.; Holland, G. P., Secondary Structure Adopted by the Gly-Gly-X Repetitive Regions of Dragline Spider Silk. *Int J Mol Sci* **2016,** *17* (12).

23. Holland, G. P.; Lewis, R. V.; Yarger, J. L., WISE NMR Characterization of Nanoscale Heterogeneity and Mobility in Supercontracted Nephila clavipes Spider Dragline Silk. *JACS* **2004,** *126*.

24. Stengel, D.; Addison, J. B.; Onofrei, D.; Huynh, N. U.; Youssef, G.; Holland, G. P., Hydration-Induced β-Sheet Crosslinking of α-Helical-Rich Spider Prey-Wrapping Silk. *Advanced Functional Materials* **2021**.

25. Xu, L.; Tremblay, M. L.; Meng, Q.; Liu, X. Q.; Rainey, J. K., $^{1}H$, $^{13}C$ and $^{15}N$ NMR assignments of the aciniform spidroin (AcSp1) repetitive domain of Argiope trifasciata wrapping silk. *Biomol NMR Assign* **2012,** *6*, 147-51.

26. Tremblay, M. L.; Xu, L.; Lefevre, T.; Sarker, M.; Orrell, K. E.; Leclerc, J.; Meng, Q.; Pezolet, M.; Auger, M.; Liu, X. Q.; Rainey, J. K., Spider wrapping silk fibre architecture arising from its modular soluble protein precursor. *Sci Rep* **2015,** *5*.

27. Tremblay, M. L.; Xu, L.; Sarker, M.; Liu, X. Q.; Rainey, J. K., Characterizing Aciniform Silk Repetitive Domain Backbone Dynamics and Hydrodynamic Modularity. *Int J Mol Sci* **2016,** *17* (1305).

28. Addison, B.; Onofrei, D.; Stengel, D.; Blass, B.; Brenneman, B.; Ayon, J.; Holland, G. P., Spider prey-wrapping silk is an alpha-helical coiled-coil/beta-sheet hybrid nanofiber. *Chem Commun (Camb)* **2018,** *54* (76), 10746-10749.

29. Ruan, Q. X.; Zhou, P.; Hu, B. W.; Ji, D., An investigation into the effect of potassium ions on the folding of silk fibroin studied by generalized two-dimensional NMR-NMR correlation and Raman spectroscopy. *FEBS J* **2008,** *275* (2), 219-32.

30. Yang, Z.; Grubb, D. T.; Jelinski, L. W., Small-Angle X-ray Scattering of Spider Dragline Silk. *Macromolecules* **1997,** *30*, 8254-61.

31. Alexandra H. Simmons, C. A. M., Lynn W. Jelinski, Molecular Orientation and Two-Component Nature of the Crystalline Fraction of Spider Dragline Silk. *Science* **1996,** *271* (5245), 84-87.

32. Zhitong Yang; Oskar Liivak; Andreas Seidel; George LaVerde; David B. Zax; Jelinski, L. W., Supercontraction and Backbone Dynamics in Spider Silk: $^{13}C$ and $^{2}H$ NMR Studies. *Journal of the American Chemical Society* **2000,** *122*.

33. Porter, D.; Vollrath, F.; Shao, Z., Predicting the mechanical properties of spider silk as a model nanostructured polymer. *Eur Phys J E Soft Matter* **2005,** *16* (2), 199-206.

34. Papadopoulos, P.; Sölter, J.; Kremer, F., Hierarchies in the structural organization of spider silk—a quantitative model. *Colloid and Polymer Science* **2008,** *287* (2), 231-236.

35. Sponner, A.; Vater, W.; Monajembashi, S.; Unger, E.; Grosse, F.; Weisshart, K., Composition and hierarchical organisation of a spider silk. *PLoS One* **2007,** *2* (10), e998.

36. Jenkins, J. E.; Sampath, S.; Butler, E.; Kim, J.; Henning, R. W.; Holland, G. P.; Yarger, J. L., Characterizing the Secondary Protein Structure of Black Widow Dragline Silk Using Solid-State NMR and X-ray Diffraction. *Biomacromolecules* **2013,** *14*, 3472-83.

37. Jenkins, J. E.; Creager, M. S.; Butler, E. B.; Lewis, R. V.; Yarger, J. L.; Holland, G. P., Solid-state NMR evidence for elastin-like beta-turn structure in spider dragline silk. *Chem Commun (Camb)* **2010,** *46* (36), 6714-6.

38. Parnham, S.; Gaines, W. A.; Duggan, B. M.; Marcotte, W. R., Jr.; Hennig, M., NMR assignments of the N-terminal domain of Nephila clavipes spidroin 1. *Biomol NMR Assign* **2011,** *5* (2), 131-3.

39. Xu, L.; Lefevre, T.; Orrell, K. E.; Meng, Q.; Auger, M.; Liu, X. Q.; Rainey, J. K., Structural and Mechanical Roles for the C-Terminal Nonrepetitive Domain Become Apparent in Recombinant Spider Aciniform Silk. *Biomacromolecules* **2017,** *18* (11), 3678-86.

40. Lewis, R., Spider Silk: Ancient Ideas for New Biomaterials. *Chem. Rev.* **2006,** *106* (9), 3762−3774.

41. Hyoung-Joon Jin; Kaplan, D., Mechanism of silk processing in insects and spiders. *Nature* **2003,** *424* (28).

42. Lin, T. Y.; Masunaga, H.; Sato, R.; Malay, A. D.; Toyooka, K.; Hikima, T.; Numata, K., Liquid Crystalline Granules Align in a Hierarchical Structure To Produce Spider Dragline Microfibrils. *Biomacromolecules* **2017,** *18* (4), 1350-1355.

43. Vollrath, F.; Knight, D. P., Liquid crystalline spinning of spider silk. *Nature* **2001,** *410* (6828), 541-8.

44. Bauer, J.; Schaal, D.; Eisoldt, L.; Schweimer, K.; Schwarzinger, S.; Scheibel, T., Acidic Residues Control the Dimerization of the N-terminal Domain of Black Widow Spiders' Major Ampullate Spidroin 1. *Sci Rep* **2016,** *6*, 34442.

45. Asakura, T.; Matsuda, H.; Kataoka, N.; Imai, A., Changes in the Local Structure of Nephila clavipes Dragline Silk Model Peptides upon Trifluoroacetic Acid, Low pH, Freeze-Drying,

and Hydration Treatments Studied by (13)C Solid-State NMR. *Biomacromolecules* **2018,** *19* (11), 4396-4410.

46. Koeppel, A.; Stehling, N.; Rodenburg, C.; Holland, C., Spinning Beta Silks Requires Both pH Activation and Extensional Stress. *Advanced Functional Materials* **2021**.

47. Blackledge, T. A.; Swindeman, J. E.; Hayashi, C. Y., Quasistatic and continuous dynamic characterization of the mechanical properties of silk from the cobweb of the black widow spider Latrodectus hesperus. *J Exp Biol* **2005,** *208* (Pt 10), 1937-49.

48. Gosline, J. M.; Guerette, P. A.; Ortlepp, C. S.; Savage, K. N., The Mechanical Design of Spider Silks: From Fibroin Sequence to Mechanical Function. *J Exp Biol* **1999,** *202*, 3295–3303.

49. Blackledge, T. A.; Hayashi, C. Y., Silken toolkits: biomechanics of silk fibers spun by the orb web spider Argiope argentata (Fabricius 1775). *J Exp Biol* **2006,** *209* (Pt 13), 2452-61.

50. Gosline, J.; Lillie, M.; Carrington, E.; Guerette, P.; Ortlepp, C.; Savage, K., Elastic proteins: biological roles and mechanical properties. *Philos Trans R Soc Lond B Biol Sci* **2002,** *357* (1418), 121-32.

51. Blackledge, T. A., Spider silk: a brief review and prospectus on research linking biomechanics and ecology in draglines and orb webs. *Journal of Arachnology* **2012,** *40* (1), 1-12.

52. John M. Gosline, M. E. D., Mark W. Denny, The structure and properties of spider silk. *Endeavour* **1986,** *10* (1), 37-43.

53. Swanson, B. O.; Blackledge, T. A.; Beltrán, J.; Hayashi, C. Y., Variation in the material properties of spider dragline silk across species. *Applied Physics A* **2005,** *82* (2), 213-218.

54. Agnarsson, I.; Kuntner, M.; Blackledge, T. A., Bioprospecting finds the toughest biological material: extraordinary silk from a giant riverine orb spider. *PLoS One* **2010,** *5* (9), e11234.

55. Blackledge, T. A.; Perez-Rigueiro, J.; Plaza, G. R.; Perea, B.; Navarro, A.; Guinea, G. V.; Elices, M., Sequential origin in the high performance properties of orb spider dragline silk. *Sci Rep* **2012,** *2*, 782.

56. Mussig, J., *Industrial Applications of Natural Fibers: Structure, Properties and Technical Applications*. Wiley & Sons: West Sussex, United Kingdom, 2010.

57. Krasnov, I.; Seydel, T.; Greving, I.; Blankenburg, M.; Vollrath, F.; Muller, M., Strain-dependent fractional molecular diffusion in humid spider silk fibres. *J R Soc Interface* **2016,** *13* (122).

58. Liu, Y.; Shao, Z.; Vollrath, F., Relationships between supercontraction and mechanical properties of spider silk. *Nat Mater* **2005,** *4* (12), 901-5.

59. Blackledge, T. A.; Boutry, C.; Wong, S. C.; Baji, A.; Dhinojwala, A.; Sahni, V.; Agnarsson, I., How super is supercontraction? Persistent versus cyclic responses to humidity in spider dragline silk. *J Exp Biol* **2009,** *212* (Pt 13), 1981-9.

60. Work, R. W., VISCOELASTIC BEHAVIOUR AND WET SUPERCONTRACTION OF MAJOR AMPULLATE SILK FIBRES OF CERTAIN ORB-WEB-BUILDING SPIDERS (ARANEAE). *J Exp Biol* **1985,** *118*, 379-404.

61. Xia, X. X.; Qian, Z. G.; Ki, C. S.; Park, Y. H.; Kaplan, D. L.; Lee, S. Y., Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. *Proc Natl Acad Sci U S A* **2010,** *107* (32), 14059-63.

62. Widmaier, D. M.; Tullman-Ercek, D.; Mirsky, E. A.; Hill, R.; Govindarajan, S.; Minshull, J.; Voigt, C. A., Engineering the Salmonella type III secretion system to export spider silk monomers. *Mol Syst Biol* **2009,** *5*, 309.

63. Huemmerich, D.; Helsen, C. W.; Quedzuweit, S.; Oschmann, J.; Rudolph, R.; Scheibel, T., Primary Structure Elements of Spider Dragline Silks and Their Contribution to Protein Solubility. *Biochemistry* **2004,** *43*, 13604-13612.

64. Rising, A.; Widhe, M.; Johansson, J.; Hedhammar, M., Spider silk proteins: recent advances in recombinant production, structure-function relationships and biomedical applications. *Cell Mol Life Sci* **2011,** *68* (2), 169-84.

65. Xu, L.; Rainey, J. K.; Meng, Q.; Liu, X. Q., Recombinant minimalist spider wrapping silk proteins capable of native-like fiber formation. *PLoS One* **2012,** *7* (11), e50227.

66. Johansson, J.; Rising, A., Doing What Spiders Cannot-A Road Map to Supreme Artificial Silk Fibers. *ACS Nano* **2021,** *15* (2), 1952-1959.

67. Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D. L., Silk-based biomaterials. *Biomaterials* **2003,** *24*, 401–416.

68. Chao, P. H.; Yodmuang, S.; Wang, X.; Sun, L.; Kaplan, D. L.; Vunjak-Novakovic, G., Silk hydrogel for cartilage tissue engineering. *J Biomed Mater Res B Appl Biomater* **2010,** *95* (1), 84-90.

69. Vepari, C.; Kaplan, D. L., Silk as a Biomaterial. *Prog Polym Sci* **2007,** *32* (8-9), 991-1007.

70. Guo, C.; Li, C.; Vu, H. V.; Hanna, P.; Lechtig, A.; Qiu, Y.; Mu, X.; Ling, S.; Nazarian, A.; Lin, S. J.; Kaplan, D. L., Thermoplastic moulding of regenerated silk. *Nat Mater* **2020,** *19* (1), 102-108.

71. Bowen, C. H.; Dai, B.; Sargent, C. J.; Bai, W.; Ladiwala, P.; Feng, H.; Huang, W.; Kaplan, D. L.; Galazka, J. M.; Zhang, F., Recombinant Spidroins Fully Replicate Primary Mechanical Properties of Natural Spider Silk. *Biomacromolecules* **2018,** *19* (9), 3853-3860.

72. Mandal, B. B.; Grinberg, A.; Gil, E. S.; Panilaitis, B.; Kaplan, D. L., High-strength silk protein scaffolds for bone repair. *Proc Natl Acad Sci U S A* **2012,** *109* (20), 7699-704.

73. Ling, S.; Li, C.; Jin, K.; Kaplan, D. L.; Buehler, M. J., Liquid Exfoliated Natural Silk Nanofibrils: Applications in Optical and Electrical Devices. *Adv Mater* **2016,** *28* (35), 7783-90.

74. Rockwood, D. N.; Preda, R. C.; Yucel, T.; Wang, X.; Lovett, M. L.; Kaplan, D. L., Materials fabrication from Bombyx mori silk fibroin. *Nat Protoc* **2011,** *6* (10), 1612-31.

75. Ling, S.; Qin, Z.; Li, C.; Huang, W.; Kaplan, D. L.; Buehler, M. J., Polymorphic regenerated silk fibers assembled through bioinspired spinning. *Nat Commun* **2017,** *8* (1), 1387.

76. Scheibel, T., Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. *Microb Cell Fact* **2004,** *3* (1), 14.

77. Rammensee, S.; Slotta, U.; Scheibel, T.; Bausch, A. R., Assembly mechanism of recombinant spider silk proteins. *Proc Natl Acad Sci U S A* **2008,** *105* (18).

78. Heidebrecht, A.; Eisoldt, L.; Diehl, J.; Schmidt, A.; Geffers, M.; Lang, G.; Scheibel, T., Biomimetic fibers made of recombinant spidroins with the same toughness as natural spider silk. *Adv Mater* **2015,** *27* (13), 2189-94.

79. Andersson, M.; Jia, Q.; Abella, A.; Lee, X. Y.; Landreh, M.; Purhonen, P.; Hebert, H.; Tenje, M.; Robinson, C. V.; Meng, Q.; Plaza, G. R.; Johansson, J.; Rising, A., Biomimetic spinning of artificial spider silk from a chimeric minispidroin. *Nat Chem Biol* **2017,** *13* (3), 262-264.

80. Aigner, T. B.; DeSimone, E.; Scheibel, T., Biomedical Applications of Recombinant Silk-Based Materials. *Adv Mater* **2018,** *30* (19), e1704636.

81. Eisoldt, L.; Hardy, J. G.; Heim, M.; Scheibel, T. R., The role of salt and shear on the storage and assembly of spider silk proteins. *J Struct Biol* **2010,** *170* (2), 413-9.

82. Exler, J. H.; Hummerich, D.; Scheibel, T., The amphiphilic properties of spider silks are important for spinning. *Angew Chem Int Ed Engl* **2007,** *46* (19), 3559-62.

83. Malay, A. D.; Suzuki, T.; Katashima, T.; Kono, N.; Arakawa, K.; Numata, K., Spider silk self-assembly via modular liquid-liquid phase separation and nanofibrillation. *Science Advances* **2020,** *6*, 1-12.

84. Parent, L. R.; Onofrei, D.; Xu, D.; Stengel, D.; Roehling, J. D.; Addison, J. B.; Forman, C.; Amin, S. A.; Cherry, B. R.; Yarger, J. L.; Gianneschi, N. C.; Holland, G. P., Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks. *Proc Natl Acad Sci U S A* **2018,** *115* (45), 11507-11512.

85. Wang, Y.; Jardetzky, O., Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* **2002,** *11* (4), 852-61.

86. Bodenhausen, G.; Ruben, D. J., Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **1980,** *69* (1).

87. Meissner, A.; Sorensen, O. W., Sequential HNCACB and CBCANH protein NMR pulse sequences. *J Magn Reson* **2001,** *151* (2), 328-31.

88. Michael Wittekind; Mueller, L., High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *Journal of Magnetic Resonance* **1993,** *101*.

89. Muhandiram, D. R.; Kay, L. E., Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance* **1994,** *103*.

90. Stephan Grzesiekt; Bax, A., Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *Journal of the American Chemical Society* **1992,** *114*.

91. Stephan Grzesiek; Bax, A., Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *Journal of Biomolecular NMR* **1993,** *3*.

92. Kay, L. E.; Nicholson, L. K.; Delaglio, F.; Bax, A.; Torchia, D. A., Pulse Sequences for Removal of the Effects of Cross Correlation between Dipolar and Chemical-Shift Anisotropy Relaxation Mechanisms on the Measurement of Heteronuclear $T_1$ and $T_2$ Values in Proteins. *Journal of Magnetic Resonance* **1992,** *97*, 359-375.

93. Meiboom, S.; Gill, D., Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Review of Scientific Instruments* **1958,** *29* (8), 688-691.

94. N. Bloembergen; E. M. Purcell; Pound, R. V., Relaxation effects in nuclear magnetic resonance absorption. *Physical Review* **1948,** *73* (7).

95. Stejskal, E. O.; Tanner, J. E., Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *The Journal of Chemical Physics* **1965,** *42* (1), 288-292.

96. Miller, C. C., The Stokes-Einstein law for diffusion in solution. *Royal Society* **1924,** *106* (740).

97. Pines, A.; Gibby, M. G.; Waugh, J. S., Proton-enhanced Nuclear Induction Spectroscopy, $^{13}C$ Chemical Shielding Anisotropy in Some Organic Solids. *Chemical Physics Letters* **1972,** *15* (3), 373-376.

98. Pines, A.; Gibby, M. G.; Waugh, J. S., Proton-enhanced NMR of dilute spins in solids. *The Journal of Chemical Physics* **1973,** *59* (2), 569-590.

99. Ishii, Y., 13C–13C dipolar recoupling under very fast magic angle spinning in solid-state nuclear magnetic resonance: Applications to distance measurements, spectral assignments, and high-throughput secondary-structure determination. *The Journal of Chemical Physics* **2001,** *114* (19), 8473-8483.

100. Takegoshi, K.; Nakamura, S.; Terao, T., 13C–1H dipolar-driven 13C–13C recoupling without 13C rf irradiation in nuclear magnetic resonance of rotating solids. *The Journal of Chemical Physics* **2003,** *118* (5), 2325-2341.

101. Berne, B. J.; Pecora, R., *Dynamic Light Scattering*. Courier Dover Publications: 2000.

102. Goodman, J. W., Some fundamental properties of speckle. *J. Opt. Soc. Am.* **1976,** *66* (11).

103. Goldburg, W. I., Dynamic light scattering. *American Journal of Physics* **1999,** *67* (12), 1152-60.

104. Koppel, D. E., Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants. *The Journal of Chemical Physics* **1972,** *57* (11), 4814-4820.

105. Hijirida, D. H.; Do, K. G.; Michal, C.; Wong, S.; Zax, D.; Jelinski, L. W., $^{13}C$ NMR of Nephila clavipes Major Ampullate Silk Gland. *Biophysical Journal* **1996,** *71*, 3442-3447.

106. Hronska, M.; Beek, J. D. v.; Williamson, P. T. F.; Vollrath, F.; Meier, B. H., NMR Characterization of Native Liquid Spider Dragline Silk from Nephila edulis. *Biomacromolecules* **2004,** *5*, 834-839.

107. Jenkins, J. E.; Holland, G. P.; Yarger, J. L., High resolution magic angle spinning NMR investigation of silk protein structure within major ampullate glands of orb weaving spiders. *Soft Matter* **2012,** *8* (6), 1947-1954.

108. Onofrei, D.; Stengel, D.; Jia, D.; Johnson, H. R.; Trescott, S.; Soni, A.; Addison, B.; Muthukumar, M.; Holland, G. P., Investigating the Atomic and Mesoscale Interactions that Facilitate Spider Silk Protein Pre-Assembly. *Biomacromolecules* **2021**.

109. Luo, R.; David, L.; Gilson, M. K., Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J Comput Chem* **2002,** *23* (13), 1244-53.

110. Allen, M. P., Introduction to Molecular Dynamics Simulation. *Computational Soft Matter* **2004,** *23*, 1-23.

111. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biology* **2011,** *9* (71).

112. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J Comput Chem* **2005,** *26* (16), 1668-88.

113. Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K., Molecular dynamics simulations of large macromolecular complexes. *Curr Opin Struct Biol* **2015,** *31*, 64-74.

114. Murphy, M. P.; Harry LeVine, I., Alzheimer's Disease and the Amyloid-β Peptide. *Journal of Alzheimer's Disease* **2010,** *19* (311-323).

115. Stefanis, L., α-Synuclein in Parkinson's disease. *Cold Spring Harb Perspect Med* **2012,** *2* (2), a009399.

116. Zheng, Z.; Diamond, M. I., Huntington disease and the huntingtin protein. *Prog Mol Biol Transl Sci* **2012,** *107*, 189-214.

117. Lee, K. H.; Zhang, P.; Kim, H. J.; Mitrea, D. M.; Sarkar, M.; Freibaum, B. D.; Cika, J.; Coughlin, M.; Messing, J.; Molliex, A.; Maxwell, B. A.; Kim, N. C.; Temirov, J.; Moore, J.; Kolaitis, R. M.; Shaw, T. I.; Bai, B.; Peng, J.; Kriwacki, R. W.; Taylor, J. P., C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles. *Cell* **2016,** *167* (3), 774-788 e17.

118. Lee, C.; Ham, S., Characterizing amyloid-beta protein misfolding from molecular dynamics simulations with explicit water. *J Comput Chem* **2011,** *32* (2), 349-55.

119. Gallicchio, E.; Levy, R. M., AGBNP: An Analytic Implicit Solvent Model Suitable for Molecular Dynamics Simulations and High-Resolution Modeling. *J Comput Chem* **2003,** *25*, 479-499.

120. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J., The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput* **2008,** *4* (5), 819-34.

121. Shih, A. Y.; Arkhipov, A.; Freddolino, P. L.; Schulten, K., Coarse Grained Protein-Lipid Model with Application to Lipoprotein Particles. *J Phys Chem B* **2006,** *110*, 3674-3684.

122. Joshi, S. Y.; Deshmukh, S. A., A review of advancements in coarse-grained molecular dynamics simulations. *Molecular Simulation* **2020**, 1-18.

123. Fields, G. B., Introduction to Peptide Synthesis. *Current Protocols in Protein Science* **2001**, 18.1.1-18.1.9.

124. Moiola, M.; Memeo, M. G.; Quadrelli, P., Stapled Peptides-A Useful Improvement for Peptide-Based Drugs. *Molecules* **2019,** *24* (20).

125. Merrifield, R. B., Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *Journal of the American Chemical Society* **1963,** *85* (14), 2149–2154.

126. Erdélyi, M.; Gogoll, A., Rapid Microwave-Assisted Solid Phase Peptide Synthesis. *Synthesis* **2002,** *11* (22), 1592-1596.

127. Behrendt, R.; White, P.; Offer, J., Advances in Fmoc solid-phase peptide synthesis. *J Pept Sci* **2016,** *22* (1), 4-27.

128. Amblard, M.; Fehrentz, J. A.; Martinez, J.; Subra, G., Fundamentals of modern peptide synthesis. *Methods Mol Biol* **2005,** *298*, 3-24.

129. Coin, I.; Beyermann, M.; Bienert, M., Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. *Nat Protoc* **2007,** *2* (12), 3247-56.

130. Wang, S.-S., p-Alkoxybenzyl Alcohol Resin and p-Alkoxybenzyloxycarbonylhydrazide Resin for Solid Phase Synthesis of Protected Peptide Fragments. *Journal of the American Chemical Society* **1973,** *95* (4).

131. Carping, L. A.; Han, G. Y., The 9-Fluorenylmethoxycarbonyl Amino-Protecting Group. *J. Org. Chem.* **1972,** *37* (22), 3404-3409.

132. Albericio, F., Orthogonal Protecting Groups for N-Amino and C-Terminal Carboxyl Functions in Solid-Phase Peptide Synthesis. *Biopolymers* **2000,** *55*, 123–139.

133. Zervas, L.; Borovas, D.; Gazi, E., New Methods in Peptide Synthesis. I. Tritylsulfenyl and o-Nitrophenylsulfenyl Groups as N-Protecting Groups. *Journal of the American Chemical Society* **1963,** *85* (22), 3660-3666.

134. Hussein, W. M.; Skwarczynski, M.; Toth, I., *Peptide Synthesis*. Springer Nature: New York, New York USA, 2020; Vol. 2103.

**Chapter 2:** Investigating the Atomic and Mesoscale Interactions that Facilitate Spider Silk Protein Pre-assembly


## Abstract

Black widow spider dragline silk is one of nature's high-performance biological polymers, exceeding the strength and toughness of most man-made materials including high tensile steel and Kevlar®. Major ampullate (Ma), or dragline silk, is primarily comprised of two spidroin proteins (Sp) stored within the Ma gland. In the native gland environment, the MaSp1 and MaSp2 proteins self-associate to form hierarchical 200-300 nm superstructures despite being intrinsically disordered proteins (IDP's). Here, dynamic light scattering (DLS), three-dimensional (3D) triple resonance solution NMR and diffusion NMR is utilized to probe MaSp size, molecular structure and dynamics of these protein pre-assemblies diluted in 4M urea and identify specific regions of the proteins important for silk protein pre-assembly. 3D NMR indicates that the Gly-Ala-Ala and Ala-Ala-Gly motifs flanking the poly(Ala) runs, which comprise the β-sheet forming domains in fibers, are perturbed by urea, suggesting that these regions may be important for silk protein pre-assembly stabilization.


## Introduction

Spider silks have evolved to be one of Nature's most versatile biopolymers, exhibiting a broad range of mechanical properties – strength, toughness, and extensibility – that surpasses both steel and Kevlar®.[1-7] Each of these unique fibers is synthesized by a specialized gland within the spider's abdomen before being used to create webs, wrap prey and construct egg cases. This adaptable biomaterial has been sought after for a range of uses including defense, biomedical and textile applications.[8-11] Major ampullate (Ma), or dragline silk, is the strongest and most well-characterized of the spider silks and is primarily constructed from two Spidroin proteins (Sp) which are stored in the Ma gland, coined MaSp1 and MaSp2.[12] These high molecular weight proteins (>250 kDa, >2800 residues) exhibit a highly repetitive core containing alternating poly(Ala) runs

and either Gly-Gly-X for MaSp1 or Gly-Pro-Gly-X-X for MaSp2 (where X is Gln, Tyr, Ser or Arg) motifs flanked by non-repetitive N- and C-termini.[12, 13] Using a variety of protein solution NMR structural and dynamic methods, the repetitive core regions of these spidroins have been shown to be primarily intrinsically disordered random coils as stored in the Ma glands.[7, 14-16]

Despite knowing the full protein sequence of several spidroins, it is still unclear how these aggregation prone proteins are solubilized and stored in the Ma gland at very high concentrations (25-50 wt.%) before being converted into solid fibers.[14, 17] The native MaSp structure is particularly difficult to probe with solution NMR because of its highly repetitive sequence, random coil nature, and propensity for aggregation. To circumvent these challenges, many studies have focused on recombinant proteins that are more readily studied as they are often soluble in water and have a considerably lower molecular weight.[18-20] Previously we have avoided these issues altogether by analyzing intact glands with solution NMR and have found that the native MaSp proteins are stored predominantly in an unfolded, random coil state that is highly dynamic.[7] Further, these MaSp proteins are not individually dissolved, but form large nanoscale (~250 nm) hierarchical pre-assemblies comprised of 100's of proteins which are the precursors to fiber formation.[21] Using cryo-electron tomography (cryo-ET) as well as DLS on MaSp proteins solubilized in 4 M urea, we discovered that these structures continue to exist even under dilute conditions.[21] We believe these larger pre-assemblies are crucial for proper assembly of silk fibers since they bridge the mesoscale gap between the protein monomers (nm) and final fibers (micron). The formation of these pre-assemblies has drawn the attention of groups generating synthetic silk fibers due to their potential role in achieving native-like mechanical properties.[22-24]

A clear understanding of the atomic and molecular level details leading to protein pre-assemblies and the impact of the physiochemical spinning conditions on these superstructures is a significant knowledge gap. Native fiber formation is a highly controlled process involving changes in pH, the removal of salt and water, and the application of shear force along the narrowing spinning duct. [18, 25-28] Understanding the impact of these factors on the atomic and

molecular level structure of the MaSp proteins is key to devising a spinning system that accurately mimics the spiders' evolutionarily optimized spinning system. It is critical then, to understand how the individual amino acid residues interact with each other and assemble into these mesoscale protein superstructures, and furthermore, how these structures are altered when exposed to the changing physiochemical conditions in the duct prior to fiber extrusion at the spider's spinneret.

Here, we use dynamic light scattering (DLS) and solution NMR to further probe the impact of denaturant on spidroin pre-assembly size and the intermolecular interactions that are responsible for stabilizing these large, mesoscale complexes. Intentional denaturation is often used for purification of recombinantly produced spidroins.[29, 30] To date, only a few studies have looked at the structure and dynamics of spidroins under denaturing conditions.[15, 31] By intentionally disrupting the spidroin superstructures in denaturant, we are able to probe the subtle but crucial molecular interactions that facilitate their stabilization by comparing samples at near native conditions (intact glands) and following incubation in 4M urea. In the present work, we characterize the MaSp proteins by dissolving them in 4 M urea and analyzing their size using dynamic light scattering (DLS) and pulsed field gradient (PFG) diffusion NMR and atomic-level interactions and dynamics using two- and three-dimensional (3D) solution NMR of isotopically-enriched samples.

**Materials and Methods**

**Sample preparation**

Black widow spiders (*L. hesperus*) were fed a saturated solution of U-[$^{13}$C, $^{15}$N]-alanine once every other day for two weeks to label the spidroins for solution NMR studies. Spiders were forcibly silked to collect Ma silk during these feedings to drive metabolic turnover in the silk proteins and direct incorporation of isotope labeled amino acids into the Ma spidroins. Previous work has shown that this labeling scheme gives significant labeling for Ala, Gly and Gln residues.[32, 33] Ala and Gly make up over 60% of the MaSp1 and MaSp2 amino acid sequences. At the end of two weeks spiders were anesthetized with $CO_2$, sacrificed and Ma glands were

dissected and removed. The outer membrane of the Ma glands was carefully peeled away, and the dope was solubilized in a solution of 4 M urea and 90:10 $H_2O$:$D_2O$, with 0.5 mM DSS for use as an internal NMR reference. UV-vis measurements at 260 and 280 nm were used to determine the protein concentration. For diffusion NMR, samples were prepared at different protein concentrations. Solutions prepared for DLS were at lower concentrations as discussed below.

**Dynamic Light Scattering**

Samples were freshly prepared as described above and then sent overnight to U. Mass. for analysis. Samples arrived within 24 hrs. and were run immediately. DLS measures the intensity- time correlation function $g_2(q,t)$ by means of a multi-channel digital correlator, and relates it to the normalized electric field correlation function $g_1(q, t)$ through the Siegert relation,

$$g_2(q,t) - 1 = \frac{\langle J(0)I(t)\rangle}{\langle I\rangle^2} - 1 = \beta|g_1(q,t)|^2$$

where β is the instrument coherent factor, t is the decay time and *I*(t) is the scattered intensity. $g_1(q, t)$ can be expressed as,

$$g_1(q,t) = A_0 + \sum A_i e^{(-\Gamma_i t)}$$

where $A_0$ is the baseline, and $A_i$ is the intensity weighting fraction of ith mode, each having a characteristic decay rate $\Gamma_i$. From the slope of $\Gamma = Dq^2$ across all q, the diffusion coefficient D can be obtained. Both multiple exponential fitting method and CONTIN method were used to analyze the data and yielded similar results. Examples of fitting results can be found in the **Figure 30**.

**Solution NMR**

Solution NMR experiments were conducted with a Bruker 600 MHz spectrometer using a TXI probe with temperature control. All spectra were collected at pH ~9.0 in 4 M urea and 23° C. $^1H/^{15}N$-HSQC experiments were collected with non-uniform sampling out to 4096 points by 256 slices with 16 scans averaged per slice. Spectra were processed out to 8192 points with the balance used to zero-fill. Bruker Topspin 3.6.1 software package was used to baseline correct, phase and process 2D spectra. $^1H/^{13}C$-HSQC spectra were collected using 1024 points and 128

slices, averaging 32 scans per slice. 3-dimensional (3D) experiments were collected on individually prepared samples; separate samples were used for the structural experiments and dynamic measurements ($T_1$, $T_2$, and NOE). Sample concentrations were kept similar for all samples through the addition of 4 M urea solvent. HNCACB experiments were collected with 2048 points and non-uniformly sampled across 128 and 80 slices for the two indirect dimensions. Processing and NUS reconstruction was done in Topspin and peak picking and assignments were made using NMRFAM-Sparky.[34] Dynamics measurements were collected as pseudo-3D experiments using similar parameters for the spectroscopic dimensions ($^{1}$H and $^{15}$N) and 16 delay increments for the $T_1$ and $T_2$ measurements ranging from 10 ms to 5000 ms. Aqueous spider silk dope solutions are notoriously prone to aggregation. For this reason NMR data acquisition was completed within 48 hours of sacrificing spiders. Fitting of relaxation parameters was done in the Dynamics package of TopSpin using one and two-component exponential decay functions with error estimation done by Monte Carlo simulation.

**Diffusion NMR**

NMR diffusion experiments were conducted as previously described[21] using an 800 MHz Varian VNMR spectrometer equipped with a Doty diffusion probe using an stimulated echo (STE) pulse sequence with bipolar pulse-field gradient pulses optimized for each set of experiments lasting approximately 2-5 ms (δ) with a 50-400 ms (Δ) diffusion delay using 12 points for each decay. To reduce the effects of eddy currents in the probe, a 2 ms delay was added at the end of each phase of the bipolar gradient pulses. Experiments generally had a spectral width of 12 ppm, 32 scans, 10 s recycle delay, 2.45 s acquisition time and a gradient strength between 34-1363 Gauss/cm. The probe gradients were calibrated using an anhydrous glycerol sample at 20 °C where the diffusion value is known to be 1.37 x $10^{-8}$ $cm^2/s$. [35] Protein samples were equilibrated in 4 M urea for two hours before analysis, and temperature equilibrated in the probe for five minutes before the start of each experiment.

## Results and Discussion

### Dynamic Light Scattering

Dynamic light scattering measures the scattering of light by particles in solution and can be used to determine the size of individual protein monomers and larger structures such as protein assemblies or superstructures. Previously we have shown by routine DLS that dilute solutions of native silk dope dissolved in denaturant contain both a spidroin monomer with a hydrodynamic

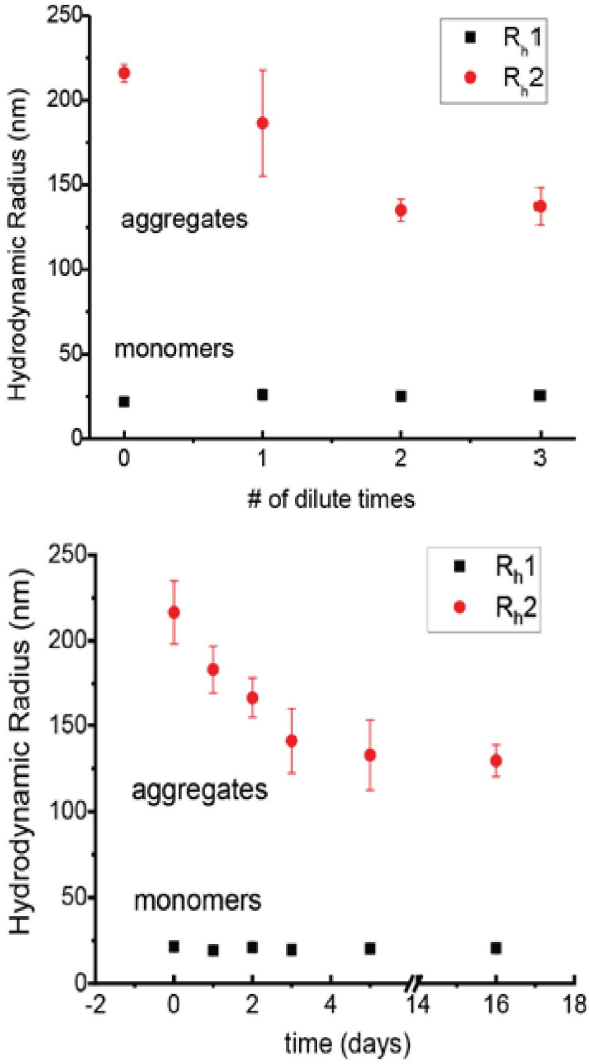



**Figure 23.** Hydrodynamic radius ($R_h$) determined from DLS data for native Ma silk proteins in 4 M urea as a function of dilution (top) and time (bottom). Both monomers and large protein assemblies (aggregates) are observed. A decrease in the size of the larger assemblies appears to stabilize after a few days. Monomer sizes remain roughly the same during the entire time. Dilution times correlate directly to the values of wt% reported in Table 1, namely 0.1, 0.05, 0.025, 0.0013. Samples measured over time had an initial concentration of 0.1 wt%.

radius ($R_h$) ~19 nm and larger protein assemblies on the order of 200-300 nm.[21] We now show that multi-angle DLS can detect these assemblies at different concentrations and in much higher detail as a function of concentration (**Table 2, Figure 23**). While DLS cannot probe the fine details of residue-specific structures, it provides an excellent measure of the dynamics and size of the larger-scale spidroin assembly. Furthermore, it is extremely sensitive at protein concentrations well below the levels that can be observed by NMR. We observe a decrease in the size of the larger assemblies as a function of both time and concentration (**Figure 23).** Stability is reached after a few days, which suggests that some fraction of the larger assemblies is long lived even in 4M denaturant albeit a smaller spidroin micelle superstructure that decreases in size from $R_h$ > 200 nm to a $R_h$ < 150 nm following a 4 day incubation period. These results indicate that the silk protein superstructures are disrupted by dissolution in the 4M urea as evidenced by a decrease in superstructure size as a function of both time and concentration in denaturant. Therefore, by comparing NMR spectra (below) at near native conditions (intact glands) and following incubation in denaturant, insight into the atomic and molecular interactions that stabilize the larger silk protein superstructures can be inferred.

**Table 2. DLS Results for Ma Spider Silk Proteins Diluted in 4M Urea**

| Dilution Factor | $D_1$($cm^2$/s) | $R_{h1}$ (nm) | $D_2$ ($cm^2$/s) | $R_{h2}$ (nm) |
|---|---|---|---|---|
| 0.100 %wt | (9.26 ± 0.88) E-8 | 22.0 ± 2.3 | (9.42 ± 0.22) E-9 | 216.0 ± 5.2 |
| 0.050 %wt | (7.81 ± 0.46) E-8 | 26.0 ± 3.1 | (1.09 ± 0.22) E-8 | 186.5 ± 31.3 |
| 0.025 %wt | (8.12 ± 0.13) E-8 | 25.1 ± 0.5 | (1.51 ± 0.07) E-8 | 135.0 ± 6.6 |
| 0.013 %wt | (8.02 ± 0.59) E-8 | 25.4 ± 2.1 | (1.48 ± 0.11) E-8 | 137.4 ± 11.0 |

**3D NMR Assignments**

Spiders fed U-$^{13}C/^{15}N$-Ala solutions displayed isotope enrichments for Ala, Gly and Gln similar to previous observations permitting 2D and 3D solution NMR experiments.[36] In the $^{1}H/^{15}N$-HSQC spectrum, we observe 12 resonances which correspond to three separate regions: Gly-rich, Gln-rich and Ala-rich (**Figure 24)**. These same 12 resonances were previously observed in solution NMR spectra of intact *L. hesperus* Ma glands and direct comparisons can be made.[7] Similar to intact glands, very narrow chemical shift (<0.5 ppm) dispersion in the $^{1}H$ dimension is observed, indicative of unstructured, random coil proteins, reminiscent of intrinsically disordered proteins (IDPs).[37-40] The very long chain length, sequence redundancy and poor chemical shift dispersion due to random coil protein structures make unique amino acid assignment for each of the thousands of amino acid residues challenging; nevertheless, by using 3D solution NMR experiments it is possible to resolve several short repetitive motifs. In order to partially assign the chemical shifts of all 12 resonances, we followed the standard assignment strategy of using HNCACB and CBCAcoNH triple-resonance experiments (**Figure 25**). This strategy initially permits two-residue assignments for each resonance, which agree very well with the correlated

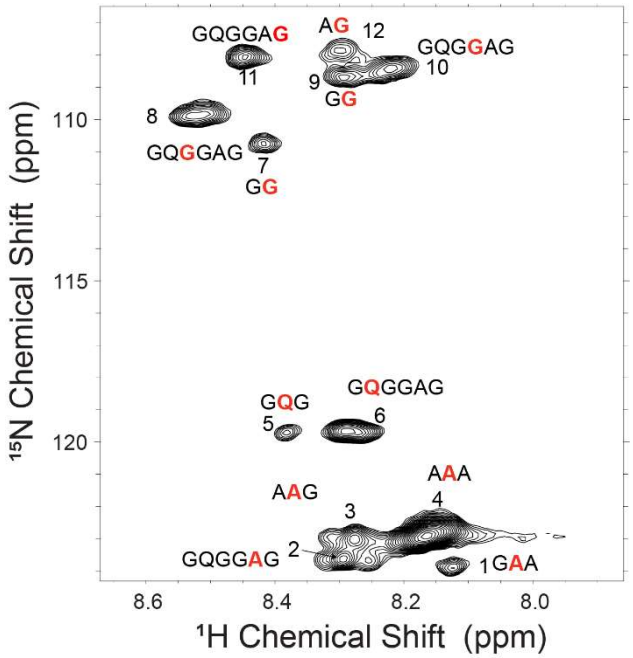



**Figure 24**. $^{1}H/^{15}N$-HSQC NMR spectrum collected at 600 MHz for native *L. hesperus* Ma silk proteins dissolved in 4 M urea (~1 wt%). The 3- and 6-residue assignments shown are possible with the use of 3D experiments shown in Figure 25.

assignment in the intact Ma glands from previous work (**Table 3**) indicating primarily random coil proteins.[7]

**Table 3. NMR Isotropic Chemical Shifts (ppm) for Ma Spider Silk Proteins in 4 M Urea[a]**

| * | Cα | Cß | NH | N | | |
|---|---|---|---|---|---|---|
| Ala | 52.67±1.28 | 19.03±1.27 | 8.11±0.68 | 132.52±3.51 | | |
| Gln | 55.94±1.83 | 28.67±1.73 | 8.25±0.75 | 119.73±3.85 | | |
| Gly | 45.34±1.17 | - | 8.34±0.83 | 109.94±4.09 | | |
| | **Cα** | **Lit.** | **difference** | **Cß** | **Lit.** | **difference** |
| G**A**A | 53.06 | 52.93 | 0.13 | 19.52 | 18.77 | 0.75 |
| A**A**A | 52.81 | 52.99 | 0.18 | 19.27 | 19.11 | 0.16 |
| A**A**G | 53.81 | 53.03 | 0.78 | 19.27 | 19.12 | 0.15 |
| GQGG**A**G | 52.81 | 52.95 | 0.14 | 19.52 | 19.36 | 0.16 |
| G**Q**G | 56.34 | 56.09 | 0.25 | 29.61 | 29.51 | 0.10 |
| G**Q**GGAG | 56.09 | 56.04 | 0.05 | 29.61 | 29.50 | 0.11 |
| G**G** | 45.50 | 45.50 | 0.00 | - | - | - |
| Q**G**G | 45.50 | 45.50 | 0.00 | - | - | - |
| G**G** | 45.50 | 0 | n.a. | - | - | - |
| GQG**G**AG | 45.50 | 45.41 | 0.09 | - | - | - |
| A**G** | 45.50 | 45.41 | 0.09 | - | - | - |
| A**G** | 45.50 | 45.37 | 0.13 | - | - | - |
| | **NH** | **Lit.** | **difference** | **N** | **Lit.** | |
| G**A**A | 8.11 | 8.10 | 0.01 | 123.9 | 123.99 | 0.1 |
| GQGG**A**G | 8.28 | 8.27 | 0.01 | 123.6 | 123.75 | 0.2 |
| A**A**G | 8.25 | 8.25 | 0.00 | 123.1 | 122.75 | -0.3 |
| A**A**A | 8.14 | 8.10 | 0.04 | 122.9 | 122.62 | -0.3 |
| G**Q**G | 8.36 | 8.35 | 0.01 | 119.8 | 119.72 | -0.1 |
| G**Q**GGAG | 8.27 | 8.26 | 0.01 | 119.8 | 119.68 | -0.1 |
| G**G** | 8.40 | 8.40 | 0.00 | 110.8 | 110.68 | -0.1 |
| QGG | 8.52 | 8.51 | 0.01 | 109.9 | 109.85 | -0.1 |
| G**G** | 8.28 | n.a. | n.a. | 108.7 | n.a. | |
| GQG**G**AG | 8.21 | 8.19 | 0.02 | 108.6 | 108.49 | -0.1 |
| A**G** | 8.43 | 8.44 | 0.01 | 108.2 | 108.08 | -0.1 |
| A**G** | 8.28 | 8.24 | 0.04 | 107.9 | 107.62 | -0.3 |
| | **Hα** | **Lit.** | **difference** | **Hß** | **Lit.** | |
| Ala | 4.26 | 4.23 | 0.03 | 1.41 | 1.38 | -0.03 |
| Ala | 4.33 | 4.30 | 0.03 | 1.39 | 1.38 | -0.01 |
| Gln | 4.36 | 4.35 | 0.01 | 1.98 | 1.97 | -0.01 |
| Gln | 3.97 | 3.95 | 0.02 | 2.16 | 2.14 | -0.02 |
| Gly | 3.89 | 3.95 | 0.06 | - | - | - |

Literature (Lit.) isotropic chemical shifts are for native *L. hesperus* Ma spider silk proteins from reference 1.

*The range in chemical shifts is shown for each amino acid observed in the solution NMR spectra.

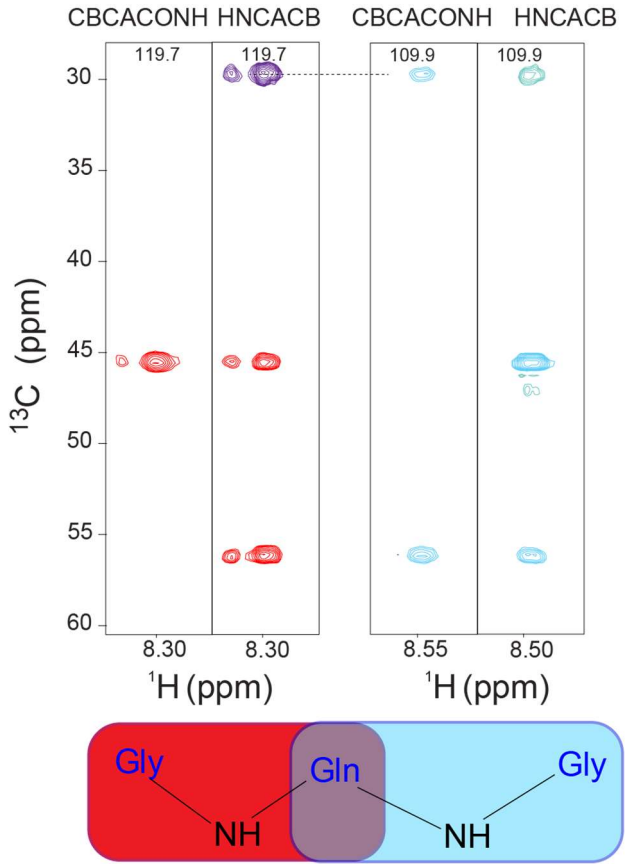



**Figure 25**. Strip plots from 3D NMR experiments collected at 600 MHz for $^{13}$C/$^{15}$N-Ala-labeled native Ma spider silk proteins from the glands of *L. hesperus* spiders, HNCACB and CBCAcoNH give 2-residue correlations (top). Each pair of strips shows the same plane of the $^{15}$N dimension for both experiments and gives a two-residue assignment. The majority of repeating motifs in the silk sequences are three-residues in length. Co-incident peaks between strips tie two amide planes together to a common set of carbon resonances giving 3- and in some cases 6- residue fragments represented by the bottom graphic for the G**Q**G motif. It is important to note that these are not unique residues, but rather a collection of repeating motifs that have the same 3- or 6- residue pattern (see **Figure 31**).

In the HNCACB experiment shown in **Figure 25**, green resonances correspond to the negatively phased $^{13}$ Cβ resonances while red resonances represent positively phased $^{13}$Cα. Strong resonances correspond to $^{13}$Cα/β resonances from those directly bound to the residue (i) and weak resonances correspond to the preceding (i-1) residue $^{13}$Cα/β. In the CBCAcoNH experiment, the $^{13}$Cα/β pairs correlate only to the i-1 residue. The use of both experiments allowed us to expand our assignments to three residues. Triple-residue assignments were made by correlating pairs of strips from both 3D experiments to each other (**Figure 25**). Each pair of strips from the CBCAcoNH and HNCACB experiments correlates to an identical $^{15}$N slice from the $^{1}$H/$^{15}$N

HSQC which provides the initial i-1 to i two-residue assignment. For several peaks in the Gly-rich region, considerable overlap exists making it challenging to confidently make more than these two-residue assignments. However, in a few cases, unique $^{13}$Cα/β pairs of resonances for a single $^{15}$N chemical shift can be found in both spectra, allowing for three residue assignments **(Figure 24).**

From the three residue assignments, tentative 6-residue assignments can be made by increasing the tolerances for matching $^{13}$C peaks between the two 3D spectra. The assignment GQGGAG is made based on the highest tolerance levels that result in unique matches. This 6-residue sequence occurs exclusively in MaSp1 and not in MaSp2 and provides an excellent point of reference for structural studies on future samples. Certain other motifs that include poorly or completely unlabeled amino acids (e.g. Tyr, Ser, Arg, etc.) may be further deconvoluted from these overlapping peaks by different isotope labeling strategies that are in process.

The NMR chemical shifts of all identified residues correlate closely to those of random coils, but in a few cases, we observed significant perturbations between the intact gland and 4 M urea NMR spectra.[7] The perturbations from the chemical shifts observed for native Ma silk glands are shown in **Table 3** and summarized in **Figure 26**. Because the most common motifs in the repetitive core of the spidroins are three residues long, this permits characterization of the backbone structure and dynamics more deeply than previous studies. For example, the GQG motif is very common in the sequence of MaSp1 and MaSp2. By applying typical $^{1}$H-detected 2D and 3D protein solution NMR strategies to silk protein solutions, assignments were made for up to three consecutive residues for all of the regions observed in the $^{1}$H/$^{15}$N HSQC spectrum.[41, 42] In fact, three assignments covers ~82% of the full sequence for MaSp1 and ~50% for MaSp2 (**Figure 31**). This is a tremendous step towards the complete NMR structural and dynamic characterization of the primary Ma spider silk proteins and opens the door to continuing investigations under various biochemical conditions (pH, salts, and denaturants).

**Chemical Shift Perturbations in 4M Urea from 2D and 3D NMR**

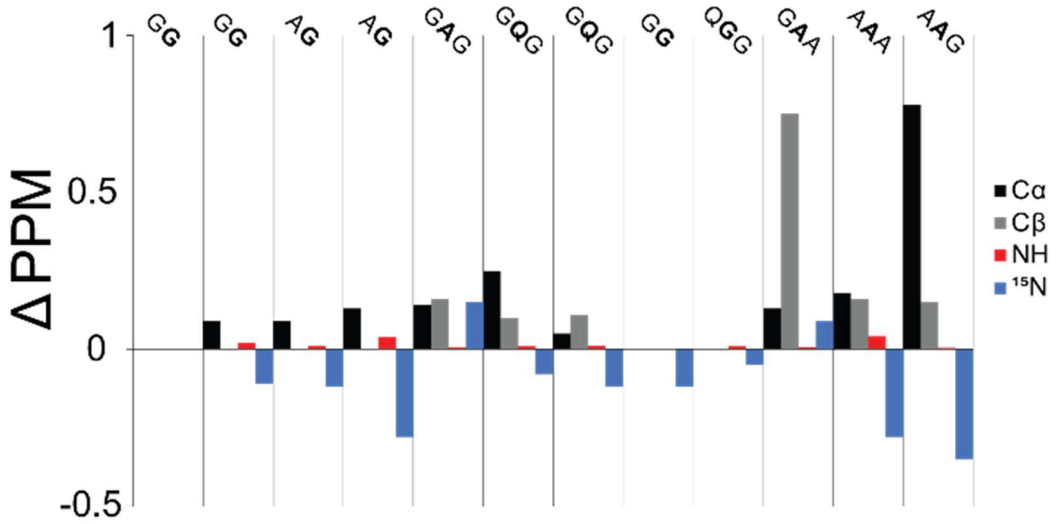



**Figure 26.** NMR chemical shift perturbations (Δ ppm) observed between the native intact Ma glands from *L. hesperus* spiders (~35 wt%) and diluted (~1 wt%) dope in 4M urea. The chart is arranged in the typical direction from the C- to N-terminus Gly-rich regions come before the poly(Ala) domains.

While most of the NMR chemical shifts are nearly identical to those of intact glands, we have identified regions of the proteins that do show significant (near 1 ppm) perturbations in the $^{13}C$ dimension following dilution in 4 M urea while other chemical shifts exhibit only minor or negligible perturbations (**Figure 26 and Figure 27**). These motifs that exhibit the largest chemical shift variability following dilution in 4M urea include the Gly-**Ala-**Ala and Ala-**Ala-**Gly that flank the poly(Ala) domains which form the nanocrystalline β-sheet core in Ma silk fibers.[43-45] Ma spider silk fibers are semi-crystalline biopolymers with small nanoscale β-sheet domains interspersed throughout an otherwise non-crystalline, disordered matrix.[46, 47] The β-sheet domains in solidified fibers consist of poly(Ala) runs of varying length and also include the flanking Gly-Ala regions.[43-45, 48] The poly(Ala) is thought to form the core of the β-sheet while, the flanking Gly-Ala are the interphase component that bridges the non- β -sheet, disordered Gly-Gly-X regions to the β-sheet component. We rationalize the large chemical shift changes at these Gly-containing β-sheet edge regions are due to some disruption of intermolecular contacts due to the presence of denaturant that could contribute to the shrinking superstructures over time in denaturant and further unfolding

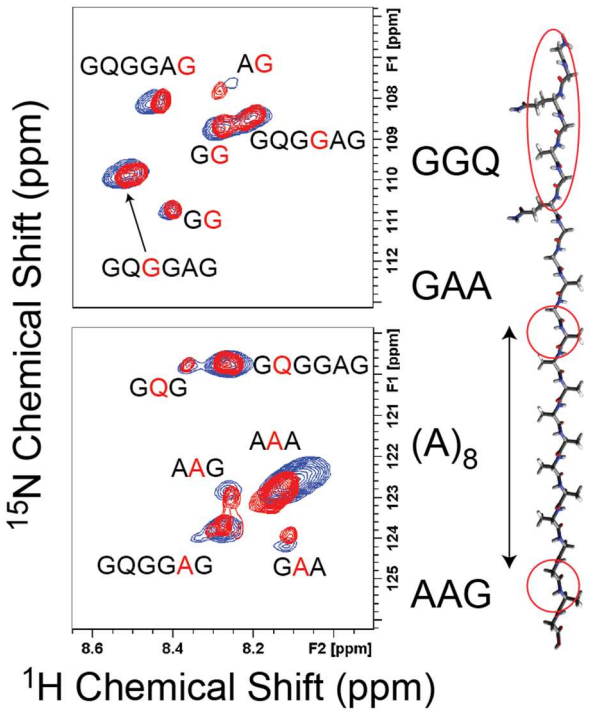



**Figure 27.** The $^{1}H/^{15}N$ HSQC NMR spectra collected at 600 MHz for $^{13}C/^{15}N$-Ala-labeled native Ma spider silk proteins from the glands of *L. hesperus* spiders. The Gly and Ala region of the spectrum are expanded in the top and bottom spectrum, respectively. The silk dope within intact Ma glands at native concentration (~35 wt%) is shown in red and Ma dope dissolved in 4 M urea (~1 wt% protein) in blue. MaSp1 backbone chain is shown to the right. The two samples yield very similar HSQC spectra with minor (AAA) or negligible chemical shift perturbations. Large $^{13}C$ chemical shift perturbations are observed in the 3D spectra for GAA and AAG (circled) with negligible shift changes for Q-containing regions (oval, see **Figure 26**).

of the monomer. This is further supported by the DLS measurements that exhibit a decreasing $R_h$ with time in denaturant (**Figure 23**). The current hypothesis is that the more hydrophilic Gln-rich Gly-Gly-X domains are solvent exposed at native concentration and exhibit no change in chemical environment due to denaturant (minimal chemical shift perturbation), while the flanking, β-sheet intermediate Gly-Ala-Ala and Ala-Ala-Gly regions that bridge the poly(Ala) and Gly-Gly-X (where is a large R-group amino acid) are perturbed. Due to the more hydrophobic nature of the poly(Ala) motif, we expect it to be preferentially buried and away from the protein surface and likely remains buried and stabilizes the smaller protein superstructures that persist for at least 2 weeks in 4 M urea (**Figure 23**).

Denaturation might cause the flanking Gly-Ala regions to be more exposed to solvent and therefore unable to form native contacts, which also helps to explain the solubility of these large aggregation-prone proteins where the superstructures are stabilized by the amphipathic nature of the MaSp proteins. It is interesting to note that the large chemical shift perturbations are observed for the Cβ of G**A**A and the Cα of A**A**G pointing to disruption of hydrophobic interactions compared to hydrogen-bonding interactions where perturbations would be expected for the $^{15}$N or NH group. This is not to say that the denaturant does not disrupt hydrogen bonds (this is likely) but, rather the chemical shift perturbations indicate the effects are larger for groups involved in hydrophobic interactions (Ala $CH_3$ and Ala CH) rather than hydrogen-bonding interactions (Ala NH). The poly(Ala) does exhibit non-negligible (~0.3 ppm) $^{15}$N perturbations indicating some degree of hydrogen-bond disruption in this region following dilution in 4M urea. The monomer also exhibits some degree of unfolding with a larger $R_h$ observed after incubation in 4M urea. Again lack of chemical shift perturbations to the poly(Ala) indicates that they may remain buried with possible minor hydrogen-bonding effects while, the flanking Ala-Ala-Gly and Gly-Gly-Ala become more solvent exposed leading to large (~0.8 ppm) chemical shift perturbations that could lead to the more unfolded nature and larger $R_h$ observed for the monomer in 4M urea.

**Dynamics from NMR Relaxation Measurements**

In order to get information about the motif-specific dynamics, we measured the Ma silk spidroin backbone relaxation parameters $T_1$, $T_2$, and heteronuclear NOEs for the amide MaSp proteins amides following dilution in 4 M urea. These values report on changes in local backbone dynamics and serve as an indicator of the level of order along the protein backbone. Not all 12 resonances in the $^1$H/$^{15}$N HSQC were sufficiently observed in the relaxation measurements to be adequately fit to a decay curve. Comparing published values of $T_1$ and $T_2$ for intact glands there is significant increase in the dynamics of the MaSp proteins suggesting they exhibit an extended disordered random coil state in 4 M urea that becomes more unfolded (more dynamic). The $T_1$ and $T_2$ parameters are extracted from a decaying exponential fitting function which can contain

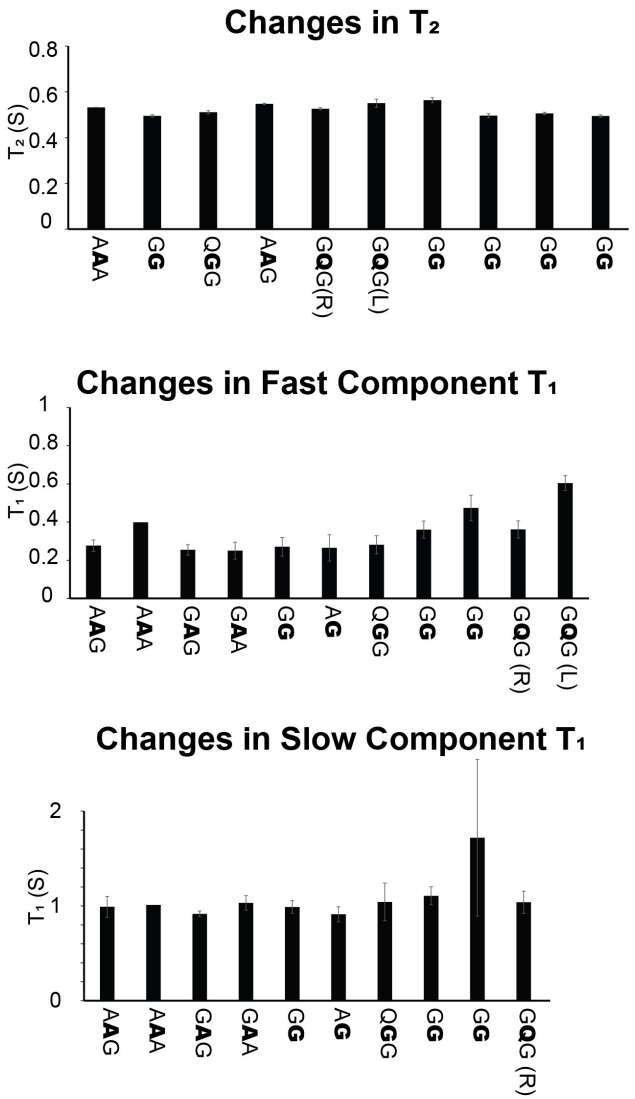



**Figure 28.** The $T_2$ (transverse) and $T_1$ (longitudinal) relaxation times measured by $^{1}H/^{15}N$ HSQC NMR at 600 MHz for $^{13}C/^{15}N$-Ala-labeled native Ma spider silk proteins from the glands of *L. hesperus* spiders following dilution in 4M urea (~1 wt%). $T_1$ relaxation times were fit to two components (**see Figure 32**.)

one or multiple components. The $T_2$ times fit to a single decaying exponential (as measured by Monte Carlo simulations to estimate error to a 95% confidence level). On the other hand, $T_1$ times fit more confidently to a two-component decaying curve (**Figure 32**). Thus, a fast and slow component for the $T_1$ relaxation times is reported in **Figure 28**. This data clearly shows that the bulk relaxation profiles of spidroins in 4 M urea are very different from their native state within intact Ma glands but, as expected, show an increase in dynamics at lower protein concentration

(1 wt%) in the presence of denaturant and exhibit two $T_1$ components that can ascribed to the monomer and protein pre-assembled environment.

Heteronuclear NOE values are another valuable indicator of protein dynamics (**Figure 33**). Values for all resolvable resonances are negative, including the AAA resonance, which is slightly below zero. Long relaxation and negative NOE values demonstrate that the spidroins are highly flexible at dilute protein concentration analogous to the Ma proteins at native concentration albeit with even faster dynamics. The AAA resonance is the least dynamic which is consistent with it being buried furthest from the solvent and presence in the core of the micelle-like protein pre-assemblies.

**Dynamics and Critical Protein Micelle Concentration from Diffusion NMR**

**Figure 29** summarizes the PFG NMR diffusion measurements for MaSp proteins diluted in 4 M urea. Diffusion values decrease with increasing concentration as expected **(Figure 29A)**. By calculating the MSD as a function of Δ time **(Figure 29B)** during the measurement and plotting the MSD slopes as a function of MaSp protein concentration **(Figure 29C**), we can determine the point at which the particles transition from freely-diffusing monomeric proteins to forming larger hierarchical protein assemblies. This approach is analogous to the approach done to routinely determine the critical micelle concentration for surfactants.[49] **Figure 29C** clearly shows an abrupt change in MaSp protein diffusion occurring between 4-5 wt%, indicating that hierarchical structures dominate above this concentration in 4 M urea.

In order to properly attribute changes in structure and dynamics to either the dissolved MaSp protein monomers or the pre-assemblies, we first needed to determine the protein concentration at which these pre-assemblies dissolve. Due to the high protein concentration DLS cannot be employed as the upper limit is ~1wt %. Instead by using diffusion NMR to monitor the disappearance of restricted diffusion as a function of concentration in 4 M urea the critical concentration for pre-assembled protein structures can be determined. Previously, we have used diffusion NMR to determine that the MaSp proteins are restricted in the gland to a volume space

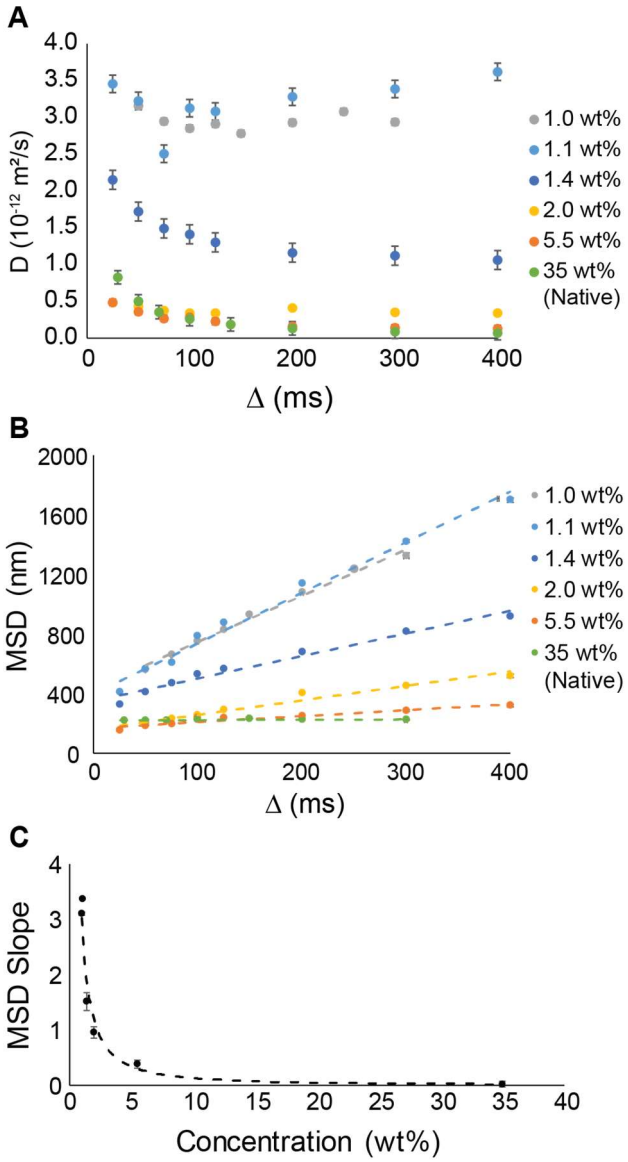



**Figure 29.** A, Diffusion coefficient measurements from PFG-STE NMR conducted at 800 MHz for native *L. hesperus* MaSp proteins as a function of concentration in 4M urea. B, MSD of MaSp proteins from A. C, Slopes of MSD from B plotted against concentration. Below ~5 wt%, MaSp proteins exist primarily as monomers in 4 M urea. Above 5 wt%, micelle-like protein pre-assemblies are present.

of ~ 250 nm after observing a diffusion dependence on Δ and subsequently taking the MSD to approximate the size of the pre-assembled protein superstructures. Additionally, we calculated the size of a single MaSp monomer using dilutions between 0.1-1 wt%, where there is no dependence on Δ, and thus free diffusion occurs, and extrapolated to infinite dilution. Modifying this approach slightly, here we measure the diffusion of MaSp proteins at concentrations between

native (35-50 wt%) and dilute (1 wt%) conditions, calculating the mean-squared displacement (MSD) and inspecting the MSDs for an abrupt change in slope corresponding to faster (unrestricted) diffusion. At native concentrations, MaSp proteins are restricted in the gland (**Figure 29A**) and the MSD is plateaued at ~300 nm (**Figure 29B**). This agrees well with our previously published data showing that in the native gland environment the MaSp proteins are found in pre-assembled, micelle-like structures in the ~300 nm regime.[21] As we begin to dilute the protein in 4 M urea, we see that the MSD increases slightly at 5.5 wt%, and abruptly between 1-2 wt% (**Figure 29B**). By plotting each slope from the calculated MSD at each concentration, we see a clear inflection between 4-5 wt% (**Figure 29C**), corresponding to the critical concentration of pre-assembled protein superstructures.

**Discussion**

Multi-angle DLS data confirms that a small population of the larger protein assemblies is maintained even in highly dilute, 4M urea protein solutions (**Figure 23**). Two peaks roughly matching the predicted sizes of both the monomeric proteins and larger assemblies were observed. The larger sized particles have previously been imaged by cryo-TEM.[21] The changes that occur over time imply that the larger assemblies are not stable upon dilution in denaturant, but may reach a stable equilibrium with the monomer population after several days. The disturbance from equilibrium may contribute to the increased dynamics observed from the NMR relaxation experiments. In order to better understand the structure of both larger assemblies and the monomers, further sequence specific information must be collected to add to the NMR assignment in **Figure 31**. Nonetheless, large chemical shift perturbations are observed for the Gly-Ala-Ala and Ala-Ala-Gly regions that flank the poly(Ala) runs. In the spun fibers the poly(Ala) comprises the core of the nanocrystalline β-sheet structures and the flanking Gly-Ala regions are also incorporated in the β-sheet but, viewed as the interphase that links the poly(Ala) to the non-β-sheet, disordered Gly-Gly-X regions. Because this observation is made following dilution and

incubation in 4M urea, it suggests that these poly(Ala) flanking regions that contain Gly may be important for protein pre-assembly stabilization into larger superstructures.

Dynamic experiments show that dilute solutions in 4 M urea have increased $T_1$, and $T_2$ times, corresponding to highly flexible unfolded random coils with increased chain dynamics. The increased dynamics can be readily explained by both the presence of denaturant and lower concentration of the sample, which agrees with diffusion data shown in **Figure 29**. Heteronuclear NOE values also indicate increased dynamics in samples dissolved in 4M urea. Based on the combination of results presented here in dilute (~1 wt%) solutions it appears to be primarily the monomer that is probed. However, the $T_1$ relaxation data clearly indicates that relaxation parameters are better fit to two-component exponential decay functions compared to one-component suggesting that the monomer and pre-assemblies could influence $T_1$. Further analysis of concentration dependence and variable temperature is required in order to truly understand this relationship and is the subject of future studies. Because the usable lifetime of these samples is roughly 48 hours, experimental parameters must be constrained to finish acquisition within this time limit. The goal of further investigation will build on this research to stabilize the protein structures and trap them in their intermediate, hierarchical states for longer periods of time to allow greater resolution and longer experiments on subtle changes to the protein dope structure and dynamics.

**Conclusions**

Understanding the nature of spider silk proteins in the dope solution, prior to fiber formation, is crucial to elucidating the spinning process, which in turn is a requirement for successfully producing synthetic spider silk fibers with native properties. Solution NMR is a powerful tool for probing the atomic- and molecular-level interactions both in terms of structure and dynamics. Recent work by our group has shown that spidroins assemble into hierarchical superstructures within the Ma glands at native concentration yet, they are surprisingly absent of virtually all secondary structure in the repetitive core. Here we have laid the groundwork for

beginning to understand the intermolecular interactions responsible for facilitating these large silk protein assemblies that point to the Gly-containing regions that terminate the poly(Ala) runs as potential sites for intermolecular interactions in pre-assembled superstructures. This was accomplished by intentionally diluting Ma silk protein dope in 4M urea, destabilizing the silk protein superstructures as evidenced by DLS and diffusion NMR, and observing NMR spectroscopic perturbations that are suggestive of interactions that stabilize these structures at native concentration. Diffusion NMR allowed for the determination of a critical concentration of ~5 wt% above which protein pre-assemblies occur, while monomers are expected to dominate at lower concentrations. Isotope labeling of the sparser amino acids such as Tyr, Ser and Arg will allow for further structural elucidation of MaSp proteins by solution NMR to expand the NMR assignment and delve deeper into the intermolecular interactions that stabilize these protein superstructures. This work is currently ongoing in our lab.

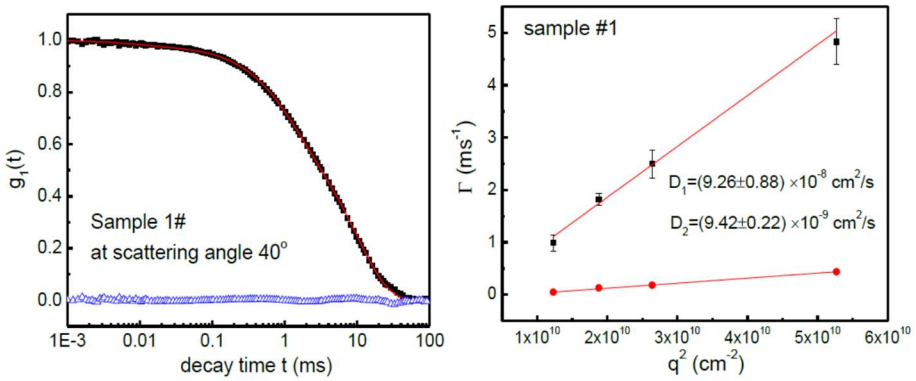



**Figure 30.** Representative example of fitting curves for DLS experiments. Results for Ma spider silk sample at a concentration 0.100 wt% in 4 M urea.

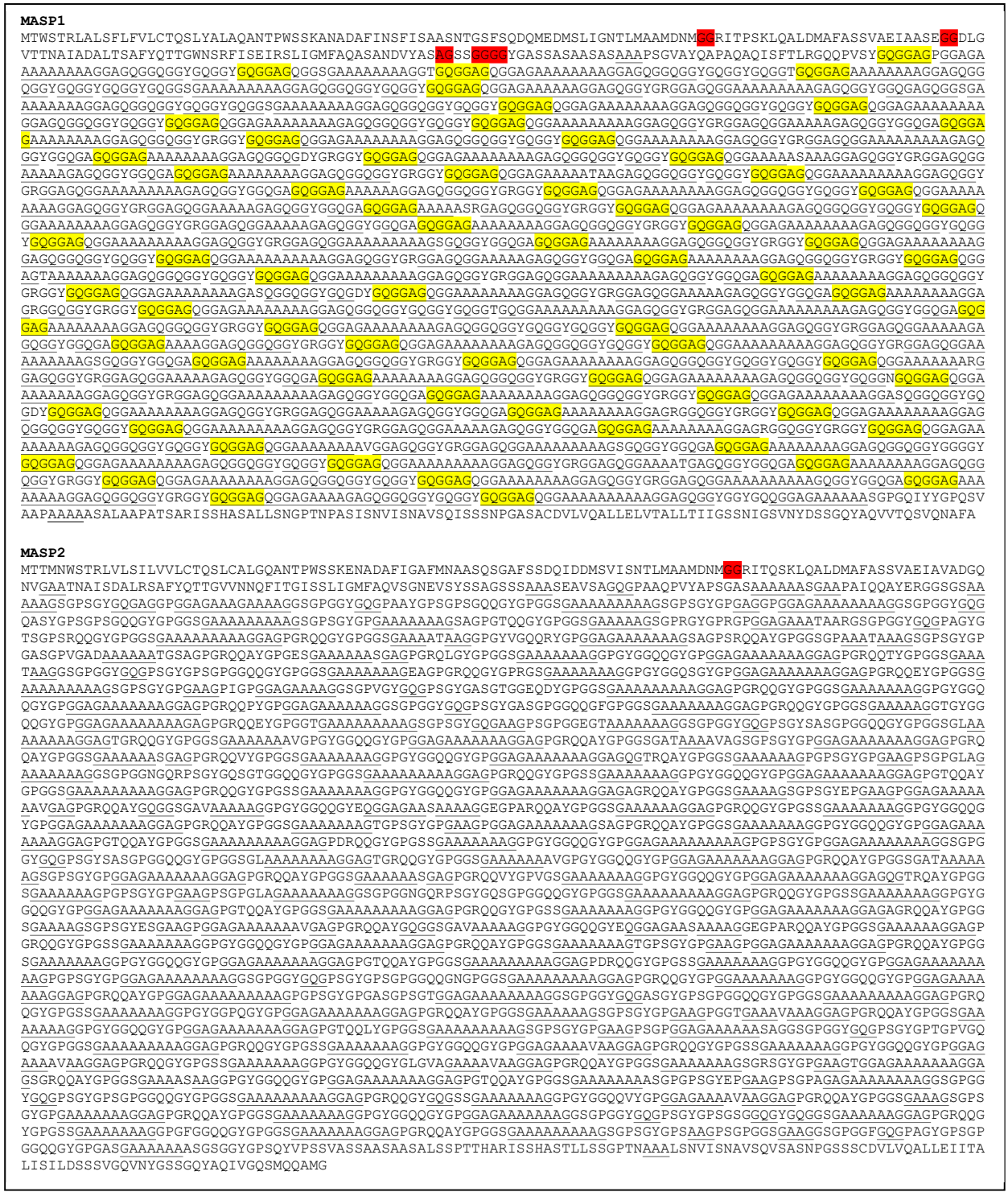



**Figure 31:** Underlined residues are three-letter motifs assigned by 3D solution NMR with six-letter motif assignments highlighted in yellow. The remaining letters highlighted in red are two-residue assignments from Figure 2 that could not be assigned to 3- or 6-residue repeat units.

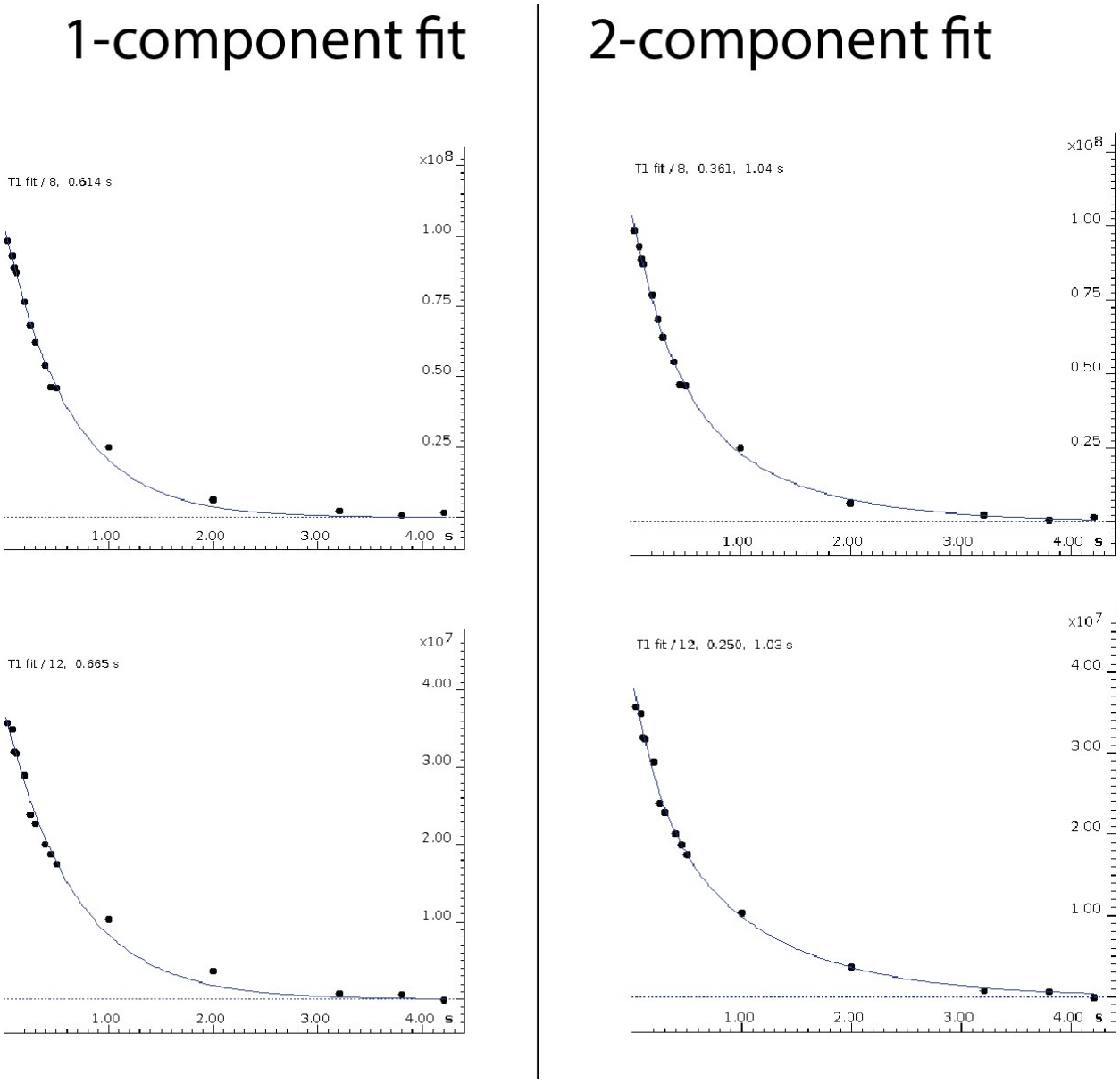



**Figure 32**. Comparison of one and two-component fit to the $T_1$ decay curves showing a statistically improved fit when two components are utilized. The fits are shown for resonances 8, GQ**G**GAG (top) and 12, A**G** (bottom) from the $^{1}H/^{15}N$ HSQC, Figure 2 in the manuscript.

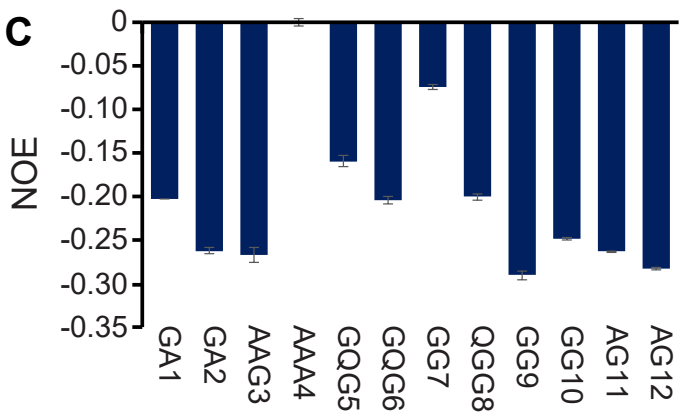



**Figure 33.** Heteronuclear NOE values are negative for all assigned resonances including AAA. However, AAA exhibits near zero NOE values indicating a lower degree of dynamics and flexibility compared to the other environments. This is consistent with AAA being the most buried and least solvent exposed.

**Acknowledgements**

D.O. is the primary author and wrote the manuscript. Co-contributors to this work are D.S., H.R.J., J.D., S.T., A.S., M.M., who all provided various data included in the figures and text and GPH who organized and led the project. All authors commented and edited the final version of the manuscript. M.M. thanks the National Science Foundation (DMR-1905730) and DOD-AFOSR Grant FA9550-20-1-0142 for financial support. G.P.H. acknowledges financial support from DOD-AFOSR under award number FA9550-17-1-0282 and FA9550-20-1-0103.

Chapter 2 is a full reprint of “Investigating the Atomic and Mesoscale Interactions That Facilitate Spider Silk Protein Pre-Assembly” published in Biomacromolecules 2021. Onofrei, D.; Stengel, D.; Jia, D.; Johnson, H. R.; Trescott, S.; Soni, A.; Addison, B.; Muthukumar, M.; Holland, G. P. David Onofrei was the primary author of this paper..

**References**


1. Ayoub, N. A.; Garb, J. E.; Kuelbs, A.; Hayashi, C. Y., Ancient Properties of Spider Silks Revealed by the Complete Gene Sequence of the Prey-Wrapping Silk Protein (AcSp1). *Mol. Biol. Evol.* **2013,** 30, (3), 589-601.
2. Agnarsson, I.; Kuntner, M.; Blackledge, T. A., Bioprospecting Finds the Toughest Biological Material: Extraordinary Silk from a Giant Riverine Orb Spider. *PLoS ONE* **2010,** 5, (9), e11234.

3. Tokareva, O.; Jacobsen, M.; Buehler, M.; Wong, J.; Kaplan, D. L., Structure–function–property–design interplay in biopolymers : Spider silk. *Acta Biomater.* **2014,** 10, (4), 1612-1626.

4. Lewis, R. V., Spider Silk: The Unraveling of a Mystery. *Acc. Chem. Res.* **1992,** 25, (9), 392-298.

5. Lewis, R. V., Spider Silk: Ancient Ideas for New Biomaterials. *Chem. Rev.* **2006,** 106, (9), 3762-74.

6. Eisoldt, L.; Smith, A.; Scheibel, T., Decoding the secrets of spider silk. *Mater. Today* **2011,** 14, (3), 80-86.

7. Xu, D.; Yarger, J. L.; Holland, G. P., Exploring the backbone dynamics of native spider silk proteins in Black Widow silk glands with solution-state NMR spectroscopy. *Polymer* **2014,** 55, (16), 3879-3885.

8. Lawrence, B. A.; Vierra, C. A.; Moore, A. M. F., Molecular and Mechanical Properties of Major Ampullate Silk of the Black Widow Spider, *Latrodectus hesperus. Biomacromolecules* **2004,** 5, (3), 689-695.

9. Zhao, A. C.; Zhao, T. F.; Nakagaki, K.; Zhang, Y. S.; Sima, Y. H.; Miao, Y. G.; Shiomi, K.; Kajiura, Z.; Nagata, Y.; Takadera, M.; Nakagaki, M., Novel molecular and mechanical properties of egg case silk from wasp spider, *Argiope bruennichi. Biochemistry* **2006,** 45, (10), 3348-56.

10. Lammel, A.; Schwab, M.; Hofer, M.; Winter, G.; Scheibel, T., Recombinant spider silk particles as drug delivery vehicles. *Biomaterials* **2011,** 32, (8), 2233-40.

11. Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D. L., Silk-based biomaterials. *Biomaterials* **2003,** 24, (3), 401-416.

12. Ayoub, N. A.; Garb, J. E.; Tinghitella, R. M.; Collin, M. A.; Hayashi, C. Y., Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One* **2007,** 2, (6), e514.

13. Motriuk-Smith, D.; Smith A.; Hayashi, C. Y.; Lewis, R. V., Analysis of the Conserved N-Terminal Domains in Major Ampullate Spider Silk Proteins. *Biomacromolecules* **2005**. *6,* (6), 3152-3159

14. Hijirida, D. H.; Do, K. G.; Michal, C.; Wong, S.; Zax, D.; Jelinski, L. W., $^{13}$C NMR of *Nephila clavipes* major ampullate silk gland. *Biophys. J.* **1996,** 71, (6), 3442-3447.

15. Hronska, M.; van Beek, J. D.; Williamson, P. T.; Vollrath, F.; Meier, B. H., NMR characterization of native liquid spider dragline silk from *Nephila edulis*. *Biomacromolecules* **2004,** 5, (3), 834-9.

16. Jenkins, J. E.; Holland, G. P.; Yarger, J. L., High resolution magic angle spinning NMR investigation of silk protein structure within major ampullate glands of orb weaving spiders. *Soft Matter* **2012,** 8, (6), 1947-1954.

17. Dicko, C.; Kenney, J. M.; Knight, D.; Vollrath, F., Transition to a β-Sheet-Rich Structure in Spidroin in Vitro: The Effects of pH and Cations. *Biochemistry* **2004,** 43, (44), 14080-14087.

18. Rising, A.; Johansson, J., Toward spinning artificial spider silk. *Nat. Chem. Biol.* **2015,** 11, (5), 309-315.

19. Heidebrecht, A.; Eisoldt, L.; Diehl, J.; Schmidt, A.; Geffers, M.; Lang, G.; Scheibel, T., Biomimetic Fibers Made of Recombinant Spidroins with the Same Toughness as Natural Spider Silk. *Adv. Mater.* **2015,** 27, (13), 2189-2194.

20. Oktaviani, N. A.; Matsugami, A.; Malay, A. D.; Hayashi, F.; Kaplan, D. L.; Numata, K., Conformation and dynamics of soluble repetitive domain elucidates the initial β-sheet formation of spider silk. *Nat. Commun.* **2018,** 9, (1), 2121.

21. Parent, L. R.; Onofrei, D.; Xu, D.; Stengel, D.; Roehling, J. D.; Addison, J. B.; Forman, C.; Amin, S. A.; Cherry, B. R.; Yarger, J. L.; Gianneschi, N. C.; Holland, G. P., Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks. *Proc. Natl. Acad. Sci. U S A* **2018,** 115, (45), 11507-11512.

22. Xu, L.; Weatherbee-Martin, N.; Liu, X.-Q.; Rainey, J. K., Recombinant Silk Fiber Properties Correlate to Prefibrillar Self-Assembly. *Small* **2019**, 15, (12), (1805294), 1-12.

23. Lang, G.; Neugirg, B. R.; Kluge, D.; Fery, A.; Scheibel, T., Mechanical Testing of Engineered Spider Silk Filaments Provides Insights into Molecular Features on a Mesoscale. *ACS Appl. Mater. Interfaces* **2017,** 9, (1), 892-900.

24. Andersson, M.; Jia, Q.; Abella, A.; Lee, X. Y.; Landreh, M.; Purhonen, P.; Hebert, H.; Tenje, M.; Robinson, C. V.; Meng, Q.; Plaza, G. R.; Johansson, J.; Rising, A., Biomimetic spinning of artificial spider silk from a chimeric minispidroin. *Nat. Chem. Biol.* **2017,** 13, (3), 262-264.

25. Römer, L.; Scheibel, T., The elaborate structure of spider silk: structure and function of a natural high performance fiber. *Prion* **2008,** 2, (4), 154-161.

26. Jin, H. J. J.; Kaplan, D. L. L., Mechanism of silk processing in insects and spiders. *Nature* **2003,** 424, 1057-1061.

27. Landreh, M.; Askarieh, G.; Nordling, K.; Hedhammar, M.; Rising, A.; Casals, C.; Astorga-Wells, J.; Alvelius, G.; Knight, S. D.; Johansson, J.; Jörnvall, H.; Bergman, T., A pH-Dependent Dimer Lock in Spider Silk Protein. *J. Mol. Biol.* **2010,** 404, (2), 328-336.

28. Giesa, T.; Perry, C. C.; Buehler, M. J., Secondary Structure Transition and Critical Stress for a Model of Spider Silk Assembly. *Biomacromolecules* **2016,** 17, (2), 427-36.

29. Foo, C. W.,; Patwardhan, S. V.; Belton, D. J.; Kitchel, B.; Anastasiades, D.; Huang, J.; Naik, R. R.; Perry, C. C.; Kaplan, D. L., Novel nanocomposites from spider silk-silica fusion (chimeric) proteins. *Proc. Natl. Acad. Sci. U S A* **2006,** 103, (25), 9428-33.

30. Rammensee, S.; Slotta, U.; Scheibel, T.; Bausch, A. R., Assembly mechanism of recombinant spider silk proteins. *Proc. Natl. Acad. Sci. U S A* **2008,** 105, (18), 6590-5.

31. Lefevre, T.; Leclerc, J.; Rioux-Dube, J. F.; Buffeteau, T.; Paquin, M. C.; Rousseau, M. E.; Cloutier, I.; Auger, M.; Gagne, S. M.; Boudreault, S.; Cloutier, C.; Pezolet, M., In situ conformation of spider silk proteins in the intact major ampullate gland and in solution. *Biomacromolecules* **2007,** 8, (8), 2342-4.

32. Alexandrescu, A. T.; Shortle, D., Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. *J. Mol. Biol.* **1994**. 242, (4), 527-546.

33. Shi, X.; Yarger, J. L.; Holland, G. P., Probing site-specific $^{13}C/^{15}N$-isotope enrichment of spider silk with liquid-state NMR spectroscopy. *Anal. Bioanal. Chem.* **2013,** 405, (12), 3997-4008.

34. Lee, W.; Tonelli, M.; Markley, J. L., NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015,** 31, (8), 1325-7.

35. Tomlinson, D. J., Temperature dependent self-diffusion coefficient measurements of glycerol by the pulsed N.M.R. technique. *Mol. Phys.* **1973,** 25, (3), 735.

36. Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M., Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* **2012,** 51, (11), 2224-2231.

37. Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z., Intrinsically disordered protein. *J. Mol. Graphics Modell.* **2001,** 19, (1), 26-59.

38. Kosol, S.; Contreras-Martos, S.; Cedeño, C.; Tompa, P., Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* **2013,** 18, (9), 10802-10828.

39. Brutscher B. et al. (2015) NMR Methods for the Study of Instrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. In: Felli I., Pierattelli R. (eds) Intrinsically Disordered Proteins Studied by NMR Spectroscopy. Advances in Experimental Medicine and Biology, vol 870. Springer, Cham. https://doi.org/10.1007/978-3-319-20164-1_3

40. Grzesiek, S.; Bax, A., Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J. Am. Chem. Soc.* **1992,** 114, (16), 6291–6293.

41. Grzesiek, S.; Bax, A., An Efficient Experiment for Sequential Backbone Assignment of Medium-sized Proteins. *J. Magn. Reson.* **1992,** 99, (1) 201-207.

42. Holland, G. P.; Creager, M. S.; Jenkins, J. E.; Lewis, R. V.; Yarger, J. L., Determining secondary structure in spider dragline silk by carbon-carbon correlation solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **2008,** 130, (30), 9871-9877.

43. Jenkins, J. E.; Creager, M. S.; Lewis, R. V.; Holland, G. P.; Yarger, J. L., Quantitative Correlation between the Protein Primary Sequences and Secondary Structures in Spider Dragline Silks. *Biomacromolecules* **2010,** 11, (1), 192-200.

44. Jenkins, J. E.; Sampath, S.; Butler, E.; Kim, J.; Henning, R. W.; Holland, G. P.; Yarger, J. L., Characterizing the Secondary Protein Structure of Black Widow Dragline Silk Using Solid-State NMR and X-ray Diffraction. *Biomacromolecules* **2013,** *14,* (10), 3472-3483

45. Marcotte, I.; van Beek, J. D.; Meier, B. H., Molecular Disorder and Structure of Spider Dragline Silk Investigated by Two-Dimensional Solid-State NMR Spectroscopy. *Macromolecules* **2007,** 40, (6), 1995-2001.

46. Holland, G. P.; Jenkins, J. E.; Creager, M. S.; Lewis, R. V.; Yarger, J. L., Quantifying the fraction of glycine and alanine in β-sheet and helical conformations in spider dragline silk using solid-state NMR. *Chem. Commun.* **2008,** 43, 5568-5570.

47. Sintya, E.; Alam, P., Self-assembled semi-crystallinity at parallel β-sheet nanocrystal interfaces in clustered MaSp1 (spider silk) proteins. *Mater. Sci. Eng. C* **2016,** 58, 366-371.

48. Denkova, P. S.; Lokeren, L. V.; Verbruggen, I.; Willem, R., Self-Aggregation and Supramolecular Structure Investigations of Triton X-100 and SDP2S by NOESY and Diffusion Ordered NMR Spectroscopy. *J. Phys. Chem. B* **2008,** 112, (35), 10935-10941.

**Chapter 3.** Molecular Dynamics Simulations on Native Sized Major Ampullate Spider Silk Proteins

## Introduction

Molecular dynamics (MD) simulations in recent years have helped to improve upon the silk fiber model as well as provide atomistic detail of mechanical characteristics of fibers under stress.[1-6] Most, if not all, of these simulations have been used to study the structure and mechanical properties of the amorphous Gly-Gly-X or poly(Ala) domains within dragline silk. In conjunction with other techniques such as NMR, these simulations can be utilized to study secondary, tertiary and higher order structure in order to improve upon the fiber architecture model. To date, there have been no investigations into the structure of the native glandular proteins using MD. In one paper from our group, coarse-grain (CG) modelling was used to illustrate the hierarchical assembly of several silk proteins as they exist in the gland.[7] We believe this pre-assembly of silk proteins in the gland is a crucial step in fiber formation, helping to establish protein-protein contacts which help to increase β-sheet character and generate a tougher silk. The first step in understanding this pre-assembly process is to first understand how a single MaSp protein is structured under native conditions. From this, we can identify peripheral residues on the surface which may be involved in facilitating protein-protein interactions, leading to the larger assemblies. Unfortunately, the secondary structure information we obtain from NMR does not directly provide a holistic illustration of a single protein since most of the chemical shifts are reported to be disordered, thus spanning a wide variety of phi-psi angles leading to a multitude of possible structures. Even with improved isotopic labeling and direct-detect experiments, the repetitiveness of the primary sequence and similarities in structure make it difficult to provide a consistent structure. In reality, multiple structures are most likely to occur for a single protein. To overcome this, we utilized molecular dynamics simulations to guide us in understanding how a protein with polymer-like characteristics could organize itself in the native state. For proteins, MD is usually implemented to understand a catalytic function, dynamics, small molecule binding, or

protein folding.[8-12] When the simulations become large, it is necessary to move to a coarse-grained approach, where groups of atoms are abbreviated by a single point which greatly reduces the time needed for simulations. This has been successfully implemented to study the dynamics and organization of large (>100s kDa) protein systems.[13] Here, we utilized MD for a similar purpose; to simulate the MaSp1 and MaSp2 proteins from extended to collapsed conformation to study their structure.

**Methods**

MD simulations were run in-house on a system assembled by Exxact Corporation. The VWS-264615-GRO mid-tower computer was built using two, NVIDIA GeForce GTX 1080 Ti, Pascal 11GB GDDR5X GPUs, with 8GB DDR4 of memory and a Core i7 processor. The Linux platform CentOS 7 operating system was used. Our general process of running a simulation with MaSp proteins was:

1. Generate a fully-atomistic model of the protein
2. Convert the model to coarse grain (CG)
3. Put it in a box, add water, counter ions, then equilibrate
4. Run the simulation until the radius of gyration plateaus
5. Export the final frame and convert it from CG to all-atom
6. Analyze the structure

In our process, we generated a fully-atomistic model of the repetitive core of MaSp1 and MaSp2, without termini, taken from previously published work.[14] The termini were added to the MaSp1 simulations after finishing. We chose to omit the termini because we know that 1) the termini are well-folded in the native environment and 2) although the N- and C-termini form dimers at certain stages of fiber formation, these are not on the timescale of our simulation.[15, 16] The randomly structured, fully-atomistic model was generated using PROFASI (protein folding and aggregation simulator) on the cluster at SDSU.[17] The run file contained only 1000 steps, a short

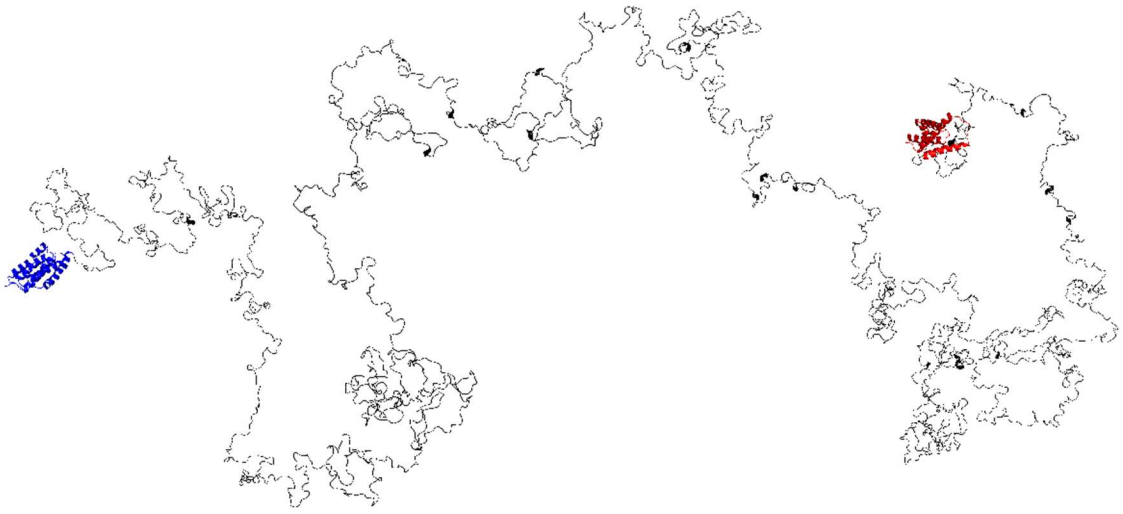


**Figure 34.** PROFASI-generated structure of MaSp1. The core of the protein (black) was randomly generated in PROFASI and the N- (blue) and C-termini (red) were added after with the Chimera program.

but ample amount of time to generate an initial extended structure. It is worth noting that these extended structures have an approximate radius of gyration between 18-26 nm (**Figure 34**), in excellent agreement with diffusion NMR data which reported the size to be 22 nm in 4 M urea.[7] The output pdb file was then converted to coarse grain (CG). The decision to utilize CG was necessary because of the exorbitant amount of time the simulation would require (5-6 months for a single run) if it was fully-atomistic (protein and water). Additionally, we anticipated running this simulation dozens of times in order to compare structures making the use of fully-atomistic models unnecessary for the level of detail we needed. Conversion to CG was done using the Martini force field in GROMACS.[18, 19] GROMACS also pilots the simulations, including adding water, counter ions, and equilibrating. The radius of gyration was monitored as a function of simulation time since this value reports on the size of the protein throughout the simulation. Generally, the simulations required 40 to 60 million steps with 0.025 nanoseconds for each step, bringing the total simulation time to 1-1.5 microseconds. Each run took approximately two weeks to complete.

**Results and Discussion**

Several structures of MaSp1 and MaSp2 were generated using this method and compared to identify patterns in their structure. We acknowledge that there was no energy minimization implemented throughout the simulations and that the structures generated are not the lowest energy conformation. These models generated are our first attempt at understanding how native, MaSp proteins are structured. The first thing we noted was the presence of a tubular-like structure throughout the entire length of the protein (**Figure 35**). Runs of 8-12 residues folded back and forth in a "switchback" like fashion and was present in all of the simulations (**Figure 35**, A-C). This seemed to prevent the full collapse of the protein in several of the simulations, leaving several structures stuck in an extended noodle-like conformation (**Figure X**, A and B). The short loop structures force the protein to interact with itself only in the immediate vicinity, an evolutionary trait possibly to prevent the protein from tangling on itself. During our analysis, we did not see a consistent pattern in these switchbacks nor consistent residues throughout the turns. Next, we compared the phi/psi angles of each residue in the simulations to published values for the

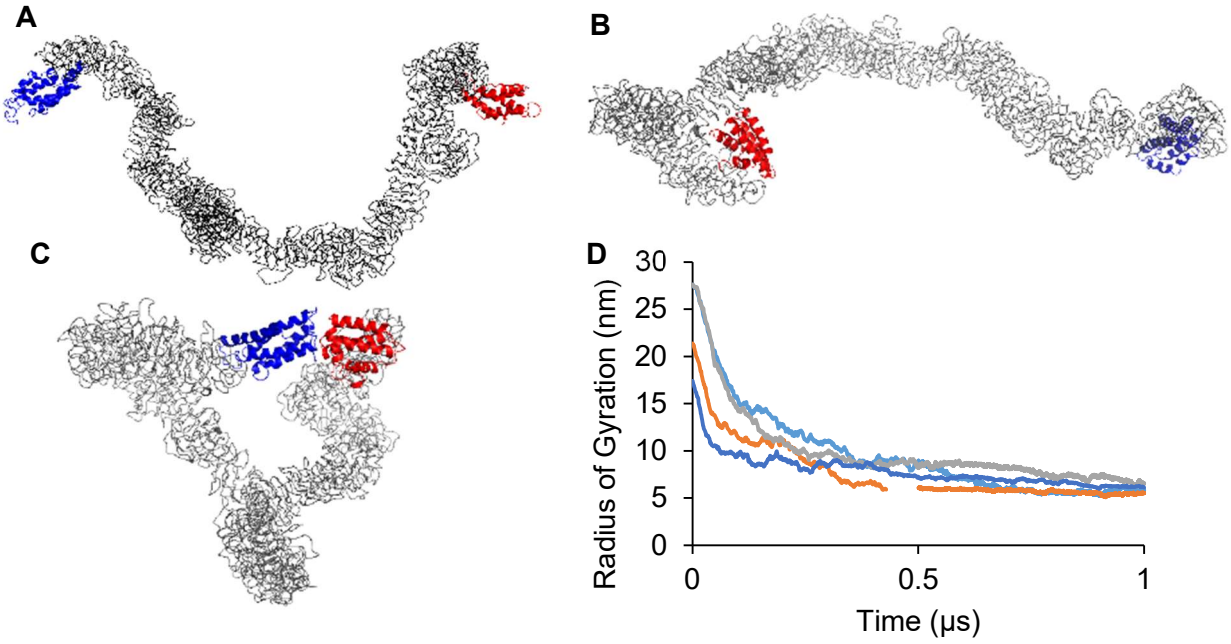
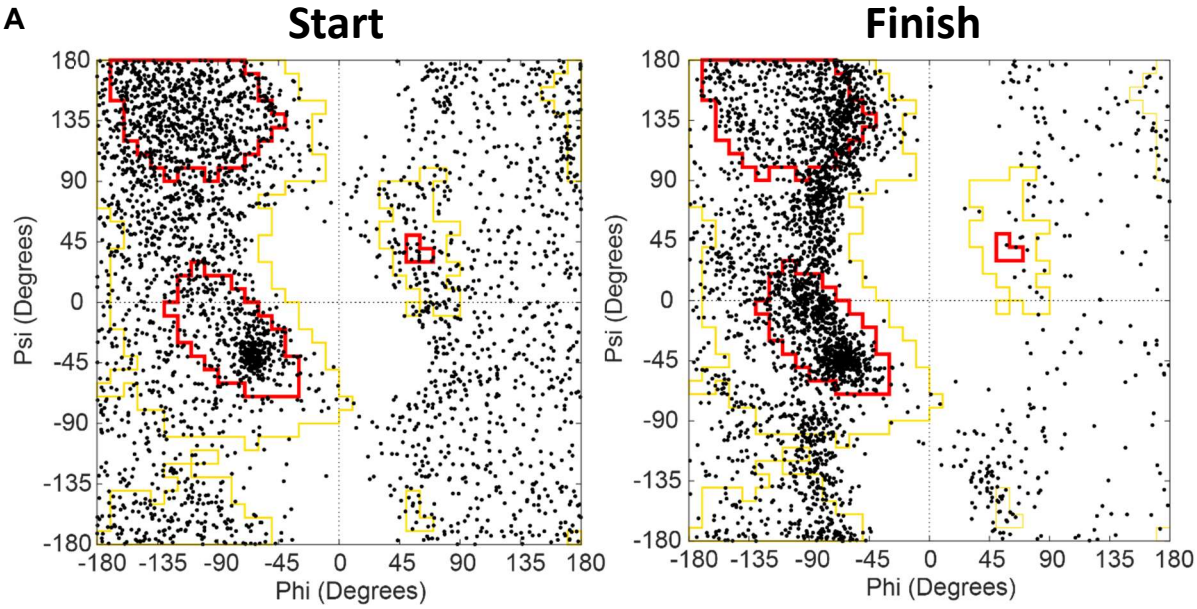



**Figure 35.** Final structures of MaSp1 simulations. The termini were added after the simulations were completed. In all structures, there is the presence of a tubular-like structure. In some cases (A and B) this prevented the complete collapse of the structure (C). The changes in radius of gyration as a function of time are plotted in D for four simulations.

structure of native silks (**Figure 36**).[20] Each phi/psi angle for Ala, Gly, Gln, Arg and Tyr were pulled from the final structures and used to determine the relative levels of α-helical, RC, and β-sheet secondary structure (**Figure 36**, B). Typical phi/psi angles for α-helical structures are near -60°/-50°. However, this was expanded to a practical range of possible α-helical phi/psi angles of phi = -89° to -39° and psi = -66 to -14. In the final structures, the total α-helical content is 15 ± 7% and

**A**

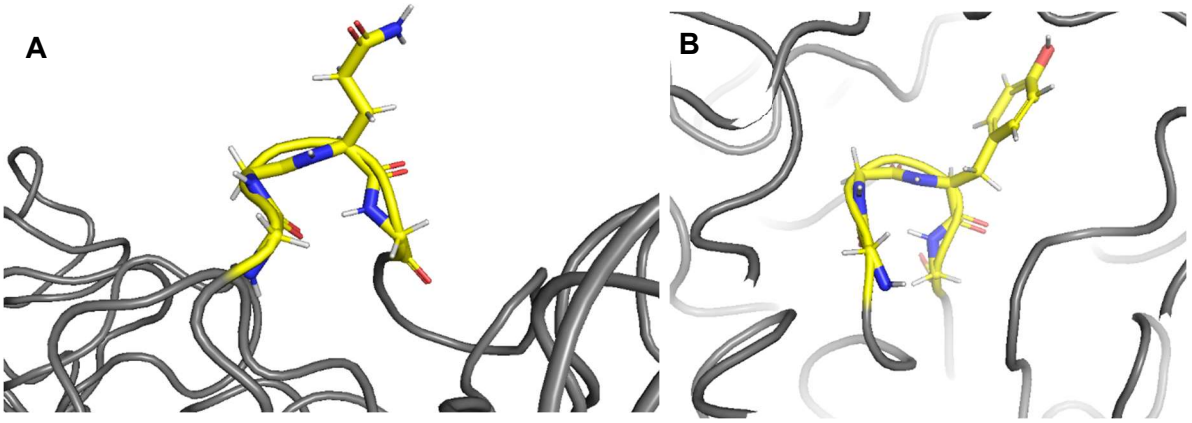



**B**

| MaSp1, N = 6 | Secondary Structure Prediction from φ/ψ Angles (%) | | |
|---|---|---|---|
| **Residue** | **α-helix** | **Random Coil** | **ß-sheet** |
| ALA | 16 | 46 | 38 |
| GLY | 11 | 51 | 38 |
| GLN | 12 | 47 | 41 |
| ARG | 15 | 48 | 38 |
| TYR | 18 | 41 | 41 |
| | | | |
| **MaSp2, N = 4** | | | |
| ALA | 13 | 45 | 42 |
| GLY | 8 | 52 | 40 |
| GLN | 13 | 44 | 43 |
| ARG | 17 | 46 | 36 |
| TYR | 13 | 40 | 47 |
| PRO | 16 | 54 | 30 |
| * MaSp2 did not have termini in this analysis | | | |

**Figure 36.** Secondary structure results for MaSp1 and MaSp2. A, Ramachandran plot comparing the starting and final simulation structures for MaSp1, which looks identical to MaSp2. B, relative secondary structure of several residues found in MaSp proteins. Overall, several key amino acids in both MaSp1 and MaSp2 are random coil, with the exception of Tyr, which agree with previous NMR structure work.

in good agreement with the proposed structure of spider silk proteins where the N- and C-termini are a well-folded α-helical bundle and contain approximately 10% of all residues in the primary structure.[14] It is possible that, upon the collapse of the structure, that some residues in the simulation fall within this phi/psi range for α-helices. Our analysis only considers residues with phi/psi angles that match an α-helix, and does not take into account multiple sequential residues (i.e. i – i+4) generally used in determining whether an amino acid truly adopts that structure. This is okay, given our routine NMR analysis of MaSp secondary structure also precludes this specificity. Similarly, the range for β-sheet structures which are typically near phi/psi angles of -140°/130° was expanded to phi = -180 to -45 and psi = 45 to 180. Random coil structures were




C

| Simulation | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| **MaSp1** | | | | |
| Type I | 1 | 5 | 3 | 2 |
| Type II | 0 | 0 | 0 | 0 |
| **MaSp2** | | | | |
| Type I | 2 | 3 | 0 | 3 |
| Type II | 0 | 0 | 0 | 0 |

D

| Simulation | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| **MaSp1** | | | | |
| Type I | 80 | 92 | 73 | 79 |
| Type II | 3 | 2 | 1 | 0 |
| **MaSp2** | | | | |
| Type I | 43 | 57 | 53 | 57 |
| Type II | 3 | 0 | 0 | 0 |

**Figure 37.** β-turn structures within MaSp1 and MaSp2 simulations. A, a type-I β-turn structure in MaSp1 (GGQG). B, a type-I β-turn structure in MaSp2 (GGYG). C, tabulated type I and type II β-turn structures in MD simulations using an angle of ±10° from phi/psi = -60/120. No type-II structures were found in any structure. D, tabulated type I and type II β-turn structures in MD simulations using an angle of ±30° from phi/psi = -60/120. After expanding the range, a few type II turns occur. Type I β-turns still appears to be the dominant turn type.

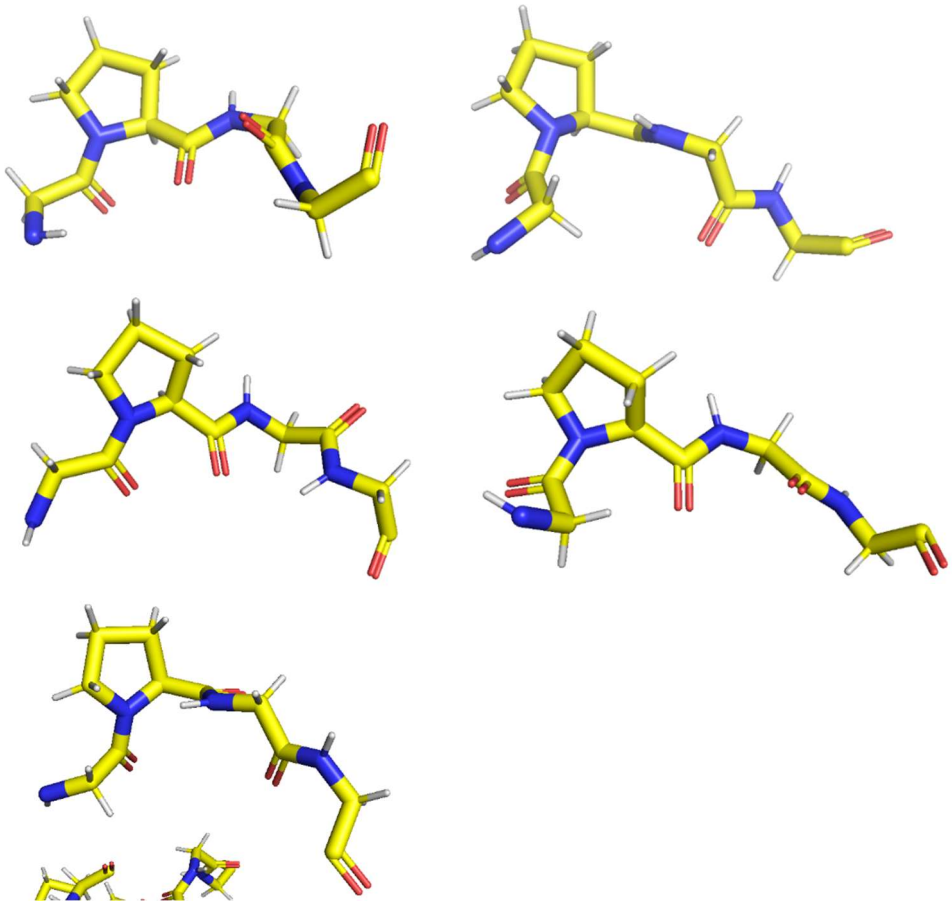


**Figure 38.** Representative images generated in Pymol for GPGG sequences taken from MaSp2 simulations. Although Pro facilitates turn-like behavior, there were few examples of this in the simulations. In most cases, the i+3 Gly was too far to form hydrogen bonds with the i Gly (> 7Å; a defining feature for both type I and type II β-turns.

assumed to be all amino acids that did not fall within this range (helical – sheet = RC). The secondary structure we report for the amino acids in our simulation agree with previously published data that shows Ala, Gly, and Gln are random coil in the native glandular environment (**Figure 36**, B).[20]

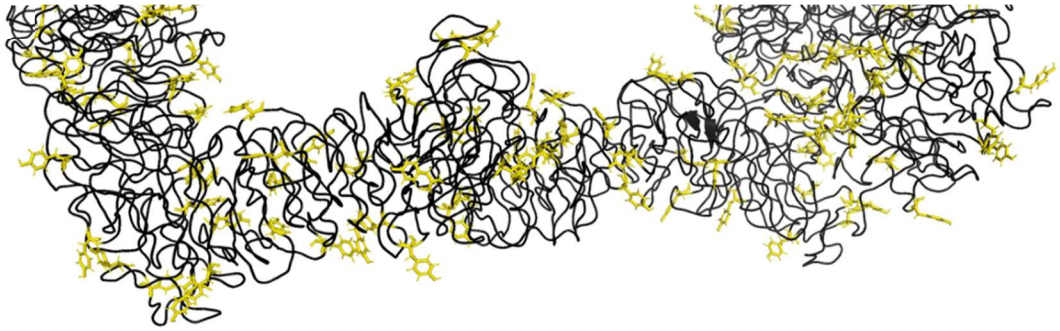
Additionally, we analyzed both MaSp1 and MaSp2 simulations for β-turn-like structures. Type I and type II β-turns, which are the most common types of turns, form hydrogen bonds between the i and i+3 residues and are differentiated by the phi and psi angles of the residues at the i+1 and i+2 positions.[21, 22] In type I β-turns, the i+1 and i+2 residues have phi/psi angles of approximately -64°/-27° and -90°/7°, respectively.[23] In type II β-turns, the i+1 and i+2 residues

have phi/psi angles of approximately -60°/131° and 84°/1°, respectively.[23] Verification of the i residue in a type II β-turn using only a single set of psi/psi angles alone is difficult, as was recently

**A**

| MaSp1 | Number Exposed | Total Residue | % Exposed (N=6) |
|---|---|---|---|
| Surface AA | 1527 | 3131 | 49 ± 7 |
| TYR | 129 | 151 | 85 ± 8 |
| GLN | 290 | 357 | 81 ± 5 |
| GLY | 535 | 1329 | 40 ± 8 |
| ALA | 369 | 1019 | 36 ± 7 |
| ARG | 48 | 51 | 95 ± 5 |
| | | | |
| **MaSp2** | | | **% Exposed (N=4)** |
| Surface AA | 1849 | 3777 | 49 ± 4 |
| TYR | 160 | 197 | 81 ± 1 |
| GLN | 221 | 261 | 85 ± 5 |
| GLY | 452 | 1266 | 36 ± 12 |
| ALA | 410 | 1176 | 35 ± 3 |
| ARG | 64 | 68 | 94 ± 0 |
| PRO | 264 | 324 | 81 ± 6 |

**B**



**C**

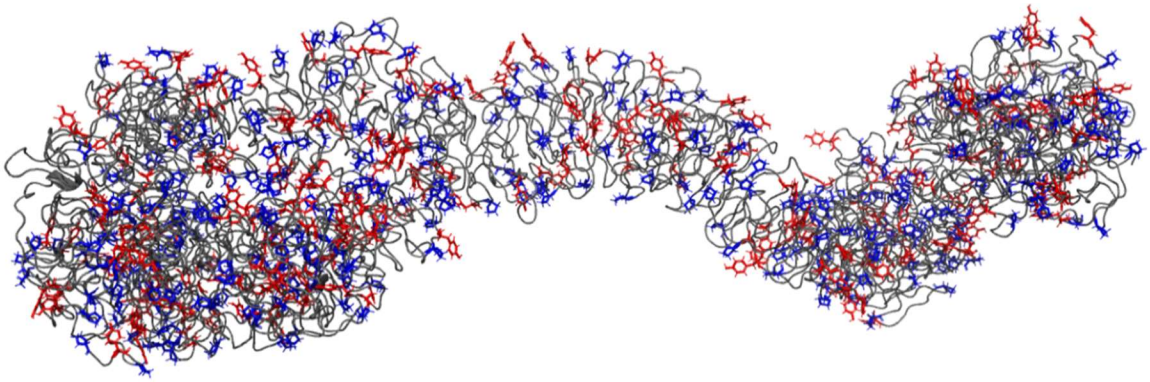


**Figure 39.** A, calculated solvent-exposed residues in MaSp1 and MaSp2 simulations. B, final partial frame of MaSp1 simulation highlighting the solvent-exposed Tyr residues (yellow). C, final frame of MaSp2 simulation highlighting Tyr (red) and Pro (blue) residues. MaSp2 simulations did not have termini.

discussed in a paper using spin-diffusion NMR experiments to calculate phi/psi angles for a residue known to be in the i position in a type II β-turn.[24] For both type I and type II analysis, we expanded this range to ±10°, and again to ±30°, from the listed values. For example, we searched for type I turns buy setting the i+1 residue psi = -74 to -54 as the first criteria, and so on. This range of angles was taken from three previous studies analyzing secondary structures from PDB deposits,[23] X-ray diffraction and solid-state NMR investigations into the peptide $(GA)_{15}$,[25] as well as solid-state NMR investigations calculating the torsion angles of $(GPGGA)_6$-G peptides.[26] Only residues which met both criteria (e.g. both residues i+1 and i+2 fall within this range) were tabulated.

Using a narrow range of ±10°, we were only able to identify a few instances of type I β-turns, and none for type II (**Figure 37**). Several of these type I β-turns occurred in the well-folded N- or C-termini. Only simulation two for MaSp1, which had five occurrences of type I β-turns, had three within the disordered region of the protein (**Figure 37, A and B**). Interestingly, there were no type-II β-turns found for MaSp2 which contains a number of GPGGX motifs, amenable to forming type-II β-turns since it contains Pro in the i+1 position and whose phi/psi angles are locked to approximately 60° and facilitates β-turn character.[23] Using a less stringent range of ±30° for each set of criteria, we were able to identify several type I turns, and a few type II (**Figure 37, D**). This value of ±30° was chosen based on the allowable range of phi/psi angles for β-turns.[27, 28] Even after expanding the range, type I β-turns appear to be the predominant turn type in our simulations. Upon closer inspection of the GPGG motifs, we see that the i+3 Gly residue is too far to form hydrogen bonds with the i Gly residue in most cases (**Figure 38**).[28] This prohibits the i+2 Gly from adopting the correct backbone conformation for a true turn, where the GPGG motifs are straight and not bent. These results differ from work investigating the structure of solid fibers, where type II β-turns were found throughout silkworm fibers,[25, 29, 30] in MD studies on *N. clavipes*,[5] and black widow dragline fibers.[31] One caveat to investigating this turn-like behavior is that these

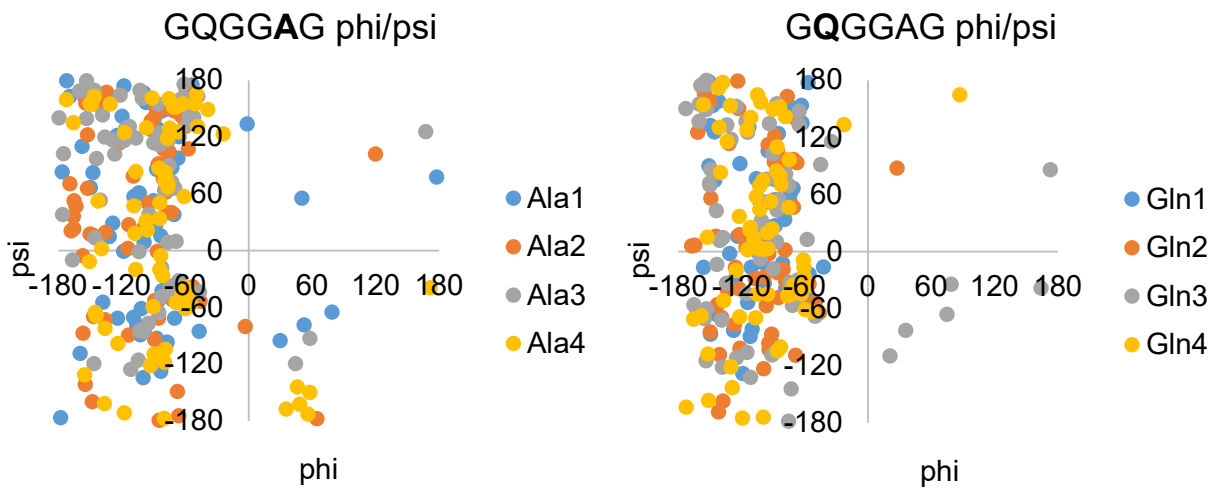



**Figure 40**. Ramachandran plot of Ala (left) and Gln (right) in GQGGAG for four simulations. Each structure used to generate the plot has 68 instances of the peptide GQGGAG, leading to 272 sets of phi/psi angles for each residue. Both Ala and Gln seem to adopt a wide range of phi/psi angles.

structures are not energy minimized and only stopped once the radius of gyration plateaus. It is possible that the structure has not reached a minimum, and that it is still possible to adopt turn-like character. Still, it is interesting that type I β-turns are more common than type II, considering the abundance of Pro in MaSp2.

As we continued our structural analysis, we noted that several residues were consistently located on the periphery of the final structures. Using a tool within the Pymol program to select solvent-exposed residues (findSurfaceResidues), we tabulated the number of residues that were on the surface and averaged them across all simulations (**Figure 39**). In both MaSp1 and MaSp2 simulations, >80% of Tyr, Gln, and Arg were found on the surface, in addition to Pro for MaSp2 (**Figure 39, A**). Ala and Gly were only 35-40% exposed in both structures. This makes sense give that Ala is commonly found in poly(Ala) motifs in the primary sequence, which is hydrophobic and would prefer not to be in contact with water. Interestingly, we also find Tyr residues residue on the surface of the protein, which are considered to be hydrophobic and would want to limit their interaction with water (**Figure 39, B and C**). However, Tyr sits somewhere in the middle of the hydropathy index among all amino acids.[32] Tyr decorations on the surface of the MaSp proteins

could function as sites for pi-pi or cation-pi interactions between proteins and may be involved in facilitating the larger pre-assemblies that have recently been observed and characterized.[7] The same could also be said for Gln and is currently the subject of our NMR investigations.

Recently, our group has investigated the atomic and mesoscale interactions that facilitate spider silk protein pre-assembly and discovered that peripheral regions near the ends of the poly(Ala) motifs show the greatest change in chemical shift between native and proteins dissolved in 4 M urea.[33] Additionally, using advanced 3D NMR assignment strategies we were able to sequence a hexapeptide repeat (GQGGAG) found exclusively within the MaSp1 sequence. Together, we utilized our MD computations and NMR analysis to guide us in enriching the model of glandular spider silk protein organization. First, we scanned the MaSp1 protein for the sequence GQGGAG and found 68 instances.[33] All of which are located in the repetitive core of the protein - not in either termini. We then tabulated the phi/psi angles for each instance across all four of our MaSp1 simulations giving us a total of 272 sequences to analyze. Ideally, our analysis would reveal that the amino acids in GQGGAG have a defined set of phi/psi angles thus providing clues to their structural role. Instead, we find that all residues occupy a wide range of phi/psi angles, including the unique residues Ala and Gln (**Figure 40**), which are predominantly found in a random coil conformation on average, in agreement with published work.[33]

**Conclusions**

Understanding the nature of spider silk proteins in the dope solution prior to fiber formation is crucial for understanding the spinning process, which in turn is a requirement for successfully producing synthetic spider silk fibers with native mechanical properties. One of the outstanding questions of this process is how the preliminary organization of these glandular proteins help to create this wonder material. Using both MD and NMR, we are currently laying the groundwork to identify key areas of the MaSp proteins that help to drive self-assembly. In this work, we utilized MD to understand how a single MaSp protein folds. Using several CG simulations for MaSp1 and MaSp2, we find that 1) both proteins have a tubular-like structure created from the looping of 8-

12 residues, 2) the secondary structure of the repetitive core of the protein is primarily random coil, 3) Arg, Tyr, Gln and Pro (MaSp2) were predominantly found on the surface of the structures. We suggest that the loops formed at the core of the structure are to prevent the protein from tangling with itself, allowing the protein to be easily extended as it travels down the duct. Because the tubular-like structure of the protein prevents some of the structures from fully collapsing, techniques that measure the anisotropy (SAXS, fluorescence anisotropy) can be utilized to corroborate what we see in MD. The exposed residues provide us clues as to what amino acids may be responsible for driving the pre-assemblies. The work to obtain atomic-level resolution of the structure of these proteins is currently ongoing in our group.

**References**


1. Keten, S.; Buehler, M. J., Nanostructure and molecular mechanics of spider dragline silk protein assemblies. *J R Soc Interface* **2010,** *7* (53), 1709-21.

2. Bratzel, G.; Buehler, M. J., Molecular mechanics of silk nanostructures under varied mechanical loading. *Biopolymers* **2012,** *97* (6), 408-17.

3. Giesa, T.; Perry, C. C.; Buehler, M. J., Secondary Structure Transition and Critical Stress for a Model of Spider Silk Assembly. *Biomacromolecules* **2016,** *17* (2), 427-36.

4. Asakura, T.; Matsuda, H.; Aoki, A.; Kataoka, N.; Imai, A., Conformational change of $^{13}$C-labeled 47-mer model peptides of Nephila clavipes dragline silk in poly(vinyl alcohol) film by stretching studied by $^{13}$C solid-state NMR and molecular dynamics simulation. *Int J Biol Macromol* **2019,** *131*, 654-665.

5. Gray, G. M.; van der Vaart, A.; Guo, C.; Jones, J.; Onofrei, D.; Cherry, B. R.; Lewis, R. V.; Yarger, J. L.; Holland, G. P., Secondary Structure Adopted by the Gly-Gly-X Repetitive Regions of Dragline Spider Silk. *Int J Mol Sci* **2016,** *17* (12).

6. Keten, S.; Xu, Z.; Ihle, B.; Buehler, M. J., Nanoconfinement controls stiffness, strength and mechanical toughness of beta-sheet crystals in silk. *Nat Mater* **2010,** *9* (4), 359-67.

7. Parent, L. R.; Onofrei, D.; Xu, D.; Stengel, D.; Roehling, J. D.; Addison, J. B.; Forman, C.; Amin, S. A.; Cherry, B. R.; Yarger, J. L.; Gianneschi, N. C.; Holland, G. P., Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks. *Proc Natl Acad Sci U S A* **2018,** *115* (45), 11507-11512.

8. Adcock, S. A.; McCammon, J. A., Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem Rev* **2006,** *106*, 1589−1615.

9. Karplus, M.; Kuriyan, J., Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* **2005,** *102* (19), 6679-6685.

10. Durrant, J. D.; McCammon, J. A., Molecular dynamics simulations and drug discovery. *BMC Biology* **2011,** *9* (71).

11. Scheraga, H. A.; Khalili, M.; Liwo, A., Protein-folding dynamics: overview of molecular simulation techniques. *Annu Rev Phys Chem* **2007,** *58*, 57-83.

12. Karplus, M.; McCammon, J. A., Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **2002,** *9* (9), 646-652.

13. Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Yu, H.; Wu, Z.; Schulten, K., Molecular dynamics simulations of large macromolecular complexes. *Curr Opin Struct Biol* **2015,** *31*, 64-74.

14. Ayoub, N. A.; Garb, J. E.; Tinghitella, R. M.; Collin, M. A.; Hayashi, C. Y., Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One* **2007,** *2* (6), e514.

15. Hagn, F.; Eisoldt, L.; Hardy, J. G.; Vendrely, C.; Coles, M.; Scheibel, T.; Kessler, H., A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature* **2010,** *465* (7295), 239-42.

16. Bauer, J.; Schaal, D.; Eisoldt, L.; Schweimer, K.; Schwarzinger, S.; Scheibel, T., Acidic Residues Control the Dimerization of the N-terminal Domain of Black Widow Spiders' Major Ampullate Spidroin 1. *Sci Rep* **2016,** *6*, 34442.

17. Irback, A.; Mohanty, S., PROFASI: A Monte Carlo simulation package for protein folding and aggregation. *J Comput Chem* **2006,** *27* (13), 1548-55.

18. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J., The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput* **2008,** *4* (5), 819-34.

19. Berendsen, H. J. C.; Spoel, D. v. d.; Drunen, R. v., GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995,** *91*.

20. Xu, D.; Yarger, J. L.; Holland, G. P., Exploring the backbone dynamics of native spider silk proteins in Black Widow silk glands with solution-state NMR spectroscopy. *Polymer* **2014,** *55* (16), 3879-85.

21. Venkatachalam, C. M., Stereochemical Criteria for Polypeptides and Proteins. V. Conformation of a System of Three Linked Peptide Units. *Biopolymers* **1968,** *6*.

22. Crawford, J. L.; Lipscomb, W. N.; Schellman, C. G., The Reverse Turn as a Polypeptide Conformation in Globular Proteins. *Proc Natl Acad Sci U S A* **1973,** *70* (2), 538-42.

23. Hutchinson, E. G.; Thornton, J. M., A revised set of potentials for ß-turn formation in proteins. *Protein Sci* **1994,** *3*.

24. Asakura, T., Structure of Silk I (Bombyx mori Silk Fibroin before Spinning) -Type II β-Turn, Not α-Helix-. *Molecules* **2021,** *6*.

25. Suzuki, Y.; Yamazaki, T.; Aoki, A.; Shindo, H.; Asakura, T., NMR study of the structures of repeated sequences, GAGXGA (X = S, Y, V), in Bombyx mori liquid silk. *Biomacromolecules* **2014,** *15* (1), 104-12.

26. Ohgo, K.; Kawase, T.; Ashida, J.; Asakura, T., Solid-state NMR analysis of a peptide (Gly-Pro-Gly-Gly-Ala)6-Gly derived from a flagelliform silk sequence of Nephi. *Biomacromolecules* **2006,** *7*.

27. Richardson, J. S., The Anatomy and Taxonomy of Protein Structure. In *Advances in Protein Chemistry Volume 34*, 1981; pp 167-339.

28. Brevern, A. G. d., Extension of the Classical Classification of b-turns. *Sci Rep* **2016,** *6*, 33191.

29. Asakura, T.; Ashida, J.; Yamane, T.; Kameda, T.; Nakazawa, Y.; Ohgo, K.; Komatsu, K., A repeated beta-turn structure in poly(Ala-Gly) as a model for silk I of Bombyx mori silk fibroin studied with two-dimensional spin-diffusion NMR under off magic angle spinning and rotational echo double resonance. *J Mol Biol* **2001,** *306* (2), 291-305.

30. Asakura, T.; Suita, K.; Kameda, T.; Afonin, S.; Ulrich, A. S., Structural role of tyrosine in Bombyx mori silk fibroin, studied by solid-state NMR and molecular mechanics on a model peptide prepared as silk I and II. *Magn Reson Chem* **2004,** *42* (2), 258-66.

31. Jenkins, J. E.; Creager, M. S.; Butler, E. B.; Lewis, R. V.; Yarger, J. L.; Holland, G. P., Solid-state NMR evidence for elastin-like beta-turn structure in spider dragline silk. *Chem Commun (Camb)* **2010,** *46* (36), 6714-6.

32. Kyte, J.; Doolittle, R. F., A Simple Method for Displaying the Hydropathic Character of a Protein. *Journal of Molecular Biology* **1982,** *157*, 105-132.

33. Onofrei, D.; Stengel, D.; Jia, D.; Johnson, H. R.; Trescott, S.; Soni, A.; Addison, B.; Muthukumar, M.; Holland, G. P., Investigating the Atomic and Mesoscale Interactions that Facilitate Spider Silk Protein Pre-Assembly. *Biomacromolecules* **2021**.

**Chapter 4.** Using Peptide Mimics to Study the Pre-assembly of Native Spider Silk Proteins

## Introduction

Due to the intrinsic disorder and repetitive sequence structure of native silk proteins, it remains challenging to determine their secondary structure and probe the silk protein hierarchical organization which exists in the gland prior to fiber formation. To solve this problem in part, we utilize high-efficiency solid-phase peptide synthesis (HE-SPPS) to generate short peptide sequences taken from the native primary silk sequence in order to study their structure under various native conditions.[1] Peptide mimics have been used ubiquitously to study the structure of liquid dopes and solid silk fibers.[2-8] However, none have utilized spider silk-like peptides to explore the mechanism of pre-assembly formation that native silks adopt which directly precede fiber formation. These native structures are 200-400 nanometers in diameter and comprise hundreds

…AAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAGGAG
QGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGDYGRGGYGQGGAGQGGAGAAAAAA
AAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAASAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAA
AAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAATAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGY
GRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAGGAG
QGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAASRGAGQ
GGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGA
AAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAGGAGQGGQGGYGQGGY
GQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGSGQGGYGGQGAGQGGAGAAAAAAAAGGAGGGGQGGY
GRGGYGQGGAGQGGAGAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAA
GAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGTAAAAAAAGGAGQGGQGGYGQGGYGQGG
AGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGGGGQGGYGRGG
YGQGGAGQGGAGAAAAAAAAGGASQGGQGGYGQGDYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGG
YGGQGAGQGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGTGQGGA
AAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGGGGQGGYGRGGYGQGGA
GQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGY
GGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAA
AAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGSGQGGYGGQGAGQGGAGAAAAAAAAGGAGGGGQGGYGRGGYGQGGAG
QGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGYGQGGAGQGGAAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGG
YGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGGGGQGGYGRGGAGQGGAGQGGA
AAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGGGGQGGYGRGGYGQGGAG
QGGAGAAAAAAAAGGASQGGQGGYGQGDYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAG
QGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAA
AGGAGQGGYGRGGAGQGGAAAAAGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGRGGQGGYGRGGYGQGGAGQGGAGAAAAA
AAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAVGGAGQGGYGRGGAGQGGAAAAAAAAAAGSGQGGYGGQGAGQGGA
GAAAAAAAAGGAGQGGQGGYGGGGYGQGGAGQGGAGAAAAAAAAGGAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGA
GQGGYGRGGAGQGGAAAATGAGQGGYGGQGAGQGGAGAAAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAG
GAGQGGQGGYGQGGYGQGGAGQGGAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAGQGGAGAAAAA
AAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAA…

**Figure 41.** Partial primary sequence of MaSp1 from *L. Hesperus* highlighting the location of our MaSp1 peptide chosen for study.

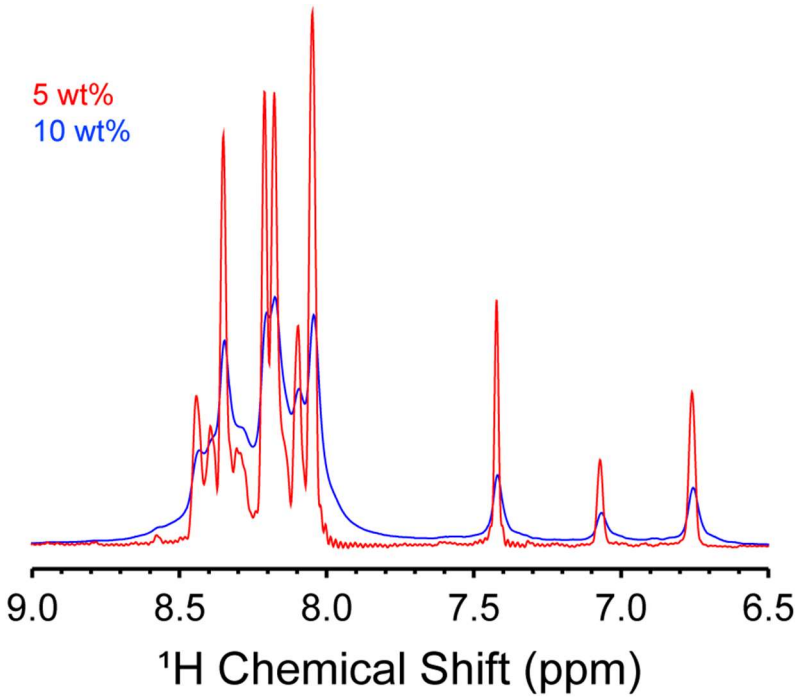



**Figure 42.** One-dimensional $^{1}H$-$^{15}N$ HSQC spectra for MaSp1 peptide in water at 5 wt% (red) and 10 wt% (blue) showing differences in signal intensity.

of silk protein monomers which serve as the fundamental precursors to silk formation.[9, 10] In the native glands, the balance of salts and pH help to drive their formation.[11-13] When these preassemblies begin to travel down the duct during the fiber spinning process, the exchange of monovalent ions for phosphate is part of the spinning process, where the kosmotropic nature of the phosphate ions helps to concentrate the native proteins in liquid dense droplets through a liquid-liquid phase separation (LLPS) mechanism .[12, 14-16] The protein-protein interactions that form during this pre-assembly process and how they change with the addition of phosphate ions are unknown. Here, we utilize HE-SPPS to generate silk-like peptides to understand the pre-assembly process that occurs in native silk dopes. Using NMR, TEM, and DLS we discover an interplay between pH, concentration, and multivalent ions that trigger pre-assembly formation.

**Methods**

**Peptide Synthesis**

The peptide, $H_3N^+$-GGAGQGGQGGYGRGGYGQGGAGQGGAG-$NH_2$, was synthesized on a Liberty Blue Peptide Synthesizer from CEM Corporation (CEM Matthews, North Carolina) using an Fmoc setup, low-loading rink amide resin and amino acids purchased from CEM Corp. Synthesis was improved (+5-10% yield) by double-coupling all Gln residues and cleaving for 15 min longer on the bench after the microwave step. After cleavage, the crude peptide was solubilized in water and injected onto a Biotage Sfar 100 g C18 column and each fraction subsequently tested with MS to identify the correct mass. The peptide was lyophilized and kept at room temperature. Isotopically-labeled amino acids were purchased from Cambridge Isotope Labs (Cambridge, MA). The peptide was sparely-labeled by mixing unlabeled and labeled amino acids in an approximate 1:1 ratio during synthesis. For example, if the recipe called for 1 g of Ala, 500 mg of labeled and 500 mg of unlabeled Ala were mixed. This was the case for all amino acids except for Gly, which was mixed in a ratio of 1:3, labeled: unlabeled.

**NMR Spectroscopy**

Solution NMR experiments were conducted with a Bruker 600 MHz Avance III spectrometer using a TXI probe with temperature control. All spectra were collected at 25 °C. $^{1}$H-$^{15}$N and $^{1}$H-$^{13}$C HSQC experiments were collected with non-uniform sampling out to 4096 points with 256 slices and 256 scans averaged per slice and a 2 s recycle delay. Spectra were processed out to 8192 points with the balance used to zero-fill. The Bruker Topspin 3.6.1 software package was used to baseline correct, phase, and process 2D spectra. Three-dimensional (3D) HNCACB experiments were collected with 2048 points, 96 scans, and were non-uniformly sampled across 64 slices for the two indirect dimensions. Processing and NUS reconstruction was done in Topspin. Peak picking and assignments were made using NMRFAM-Sparky.[17]

**Dynamic Light Scattering**

Dynamic light scattering (DLS) was done using a Malvern Instruments Zetasizer Nano-ZS. All measurements were conducted at 25 °C with a backscatter angle of 173°. The

manufacturer's suggested refractive index of protein, 1.450, was used. The data were fit using the general processing model.

**Transmission Electron Miscroscopy**

Negative staining transmission electron microscopy (NS-TEM) images were collected using a Tecnai T12 operating at 120 kV, spot size, 1 with a side-mounted CCD camera. For 1 wt% no phosphate samples, 20 μL was spotted onto a 400 mesh lacey carbon grid that was not glow discharged using a boroscillate Pasteur pipette to minimize shearing. The samples sat on the grids for five minutes and then stained with 2% uranyl acetate twice for 15-20 seconds each. Samples were dried on filter paper face-down for at least 15 minutes before imaging. The 5 wt% no phosphate, 5 wt% phosphate supernatant, 5 wt% phosphate precipitate, and 1 wt% no phosphate samples were spotted onto a 400 mesh pure carbon grid which was not glow discharged and then stained with 2% ammonium molybdate for 15 seconds each.

**Results and Discussion**

The sequence chosen, $H_3N^+$-GGAGQGGQGGYGRGGYGQGGAGQGGAG-$NH_2$, is from the *L. Hesperus* MaSp1 sequence where it appears 14 times (**Figure 41**). A sequence without a poly(Ala) run was chosen because preliminary work with peptides having this sequence were insoluble and difficult to recover. This specific sequence was chosen because it contained a similar proportion of all amino acids found throughout the disordered Gly-Gly-X sequence of the MaSp1 protein, namely X = Tyr and Arg. The peptide is readily soluble in water up to approximately 20 wt%, where gentle heating is necessary near the saturation limit. Preliminary NMR experiments indicated that highly concentrated (10-20 wt%) solutions produced lower signal intensity than less-concentration (<5 wt%) samples (**Figure 42**). All experiments thereafter were done with 5 wt% unless otherwise stated.

$^1$H experiments on unlabeled peptide show a significant difference in amide signal intensity between water and added sodium phosphate buffer at pH 7.2 (**Figure 43**). The most likely

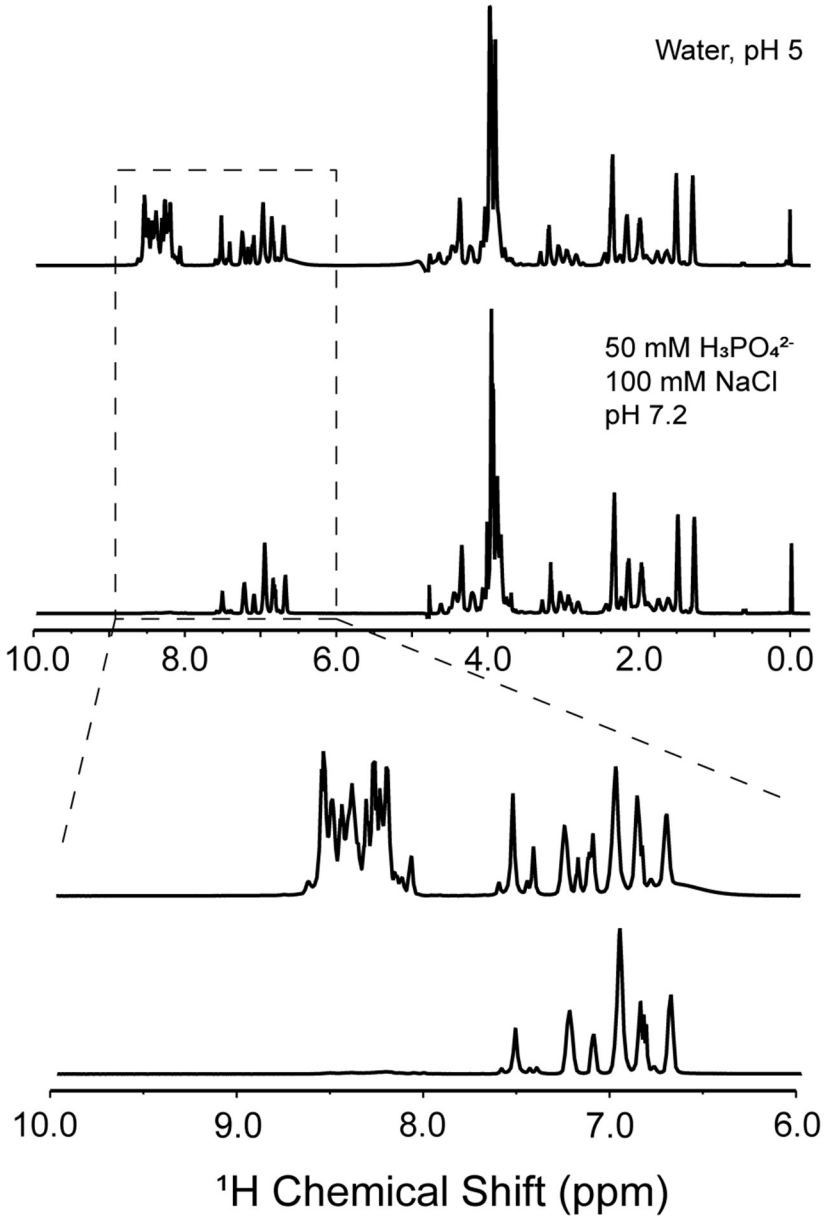



**Figure 43.** $^1$H spectra (NOESYPR1D) comparing MaSp1 peptide in water and with added sodium phosphate buffer. Several amide peaks can be resolved in the water spectra which disappear once buffer is added.

explanation is that the higher pH buffer allows $^1$H-$^{15}$N exchange on the backbone. For the peptide in water sample, the $^1$H-$^{15}$N HSQC had 23 resolvable peaks, nearly as many as the primary sequence of 27. It was therefore possible to assign several of the peaks to specific residues in the primary sequence using HNCACB experiments. The $^{15}$N HSQC spectral assignments look similar to others collected on both intact glands and silk proteins dissolved in 4 M urea where

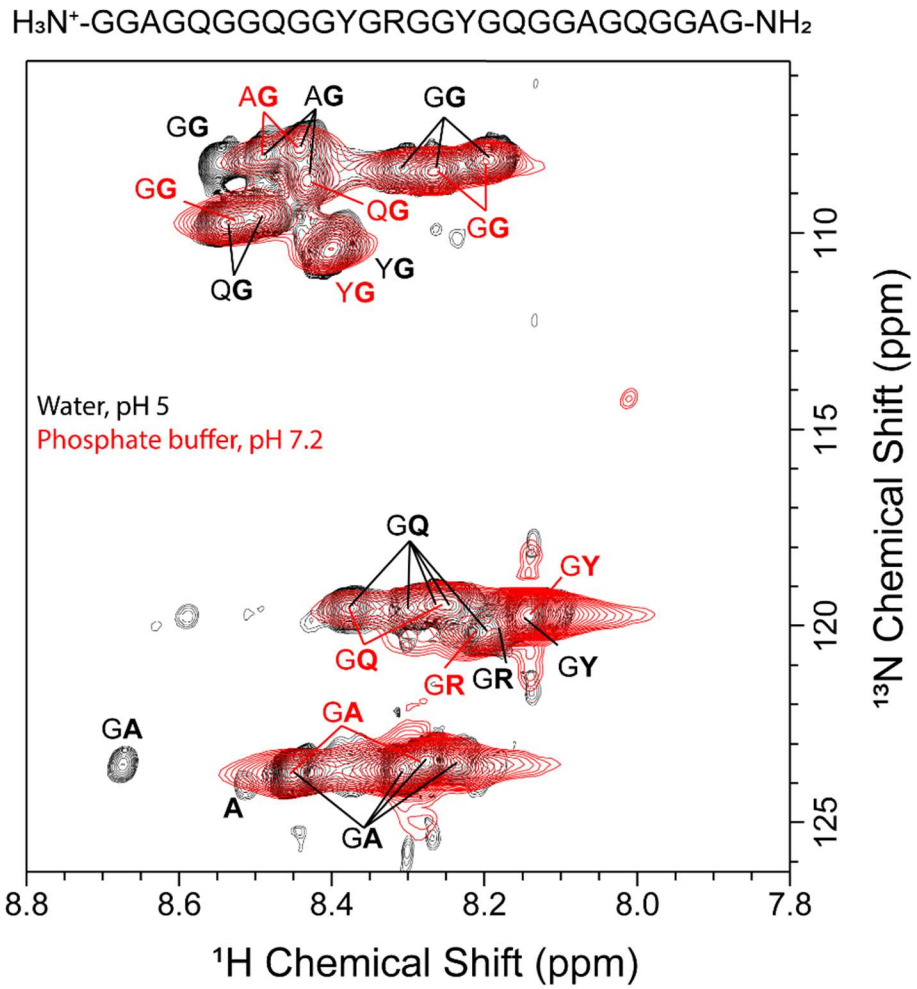



**Figure 44.** $^{1}$H-$^{15}$N HSQC of $^{13}$C/$^{15}$N-labeled MaSp1-like peptide in water, pH 5 (black) or in phosphate buffer, pH 7.2 (red). There are 22 resolvable peaks in the black spectra, and only 13 in the red. Several of the peaks shift or disappear after phosphate buffer is added.

there are three regions near 110 ppm, 120 ppm, and 123 ppm that roughly correlate to Gly, Tyr/Gln/Arg and Ala residues, respectively (**Figure 44, black**).[18, 19] The i amino acid is bolded for emphasis in the dipeptide assignments with the HNCACB experiment. For the first time, we are able to assign Arg and Tyr which occur at approximately 3 and 7 %, respectively, in the native sequence.[20] Tyr is well resolved in both the Y**G** and G**Y** regions, being furthest downfield in the Gly region's $^{15}$N dimension, and furthest up field in the Tyr/Gln/Arg region's $^{1}$H dimension. The $^{15}$N, $^{13}$C and $^{1}$H chemical shifts for Tyr assign it to a random coil (RC) structure in water (**Table 4**). Arg is buried under a G**Q** peak in the 120 ppm region (**Figure 44**). Although Arg has a similar

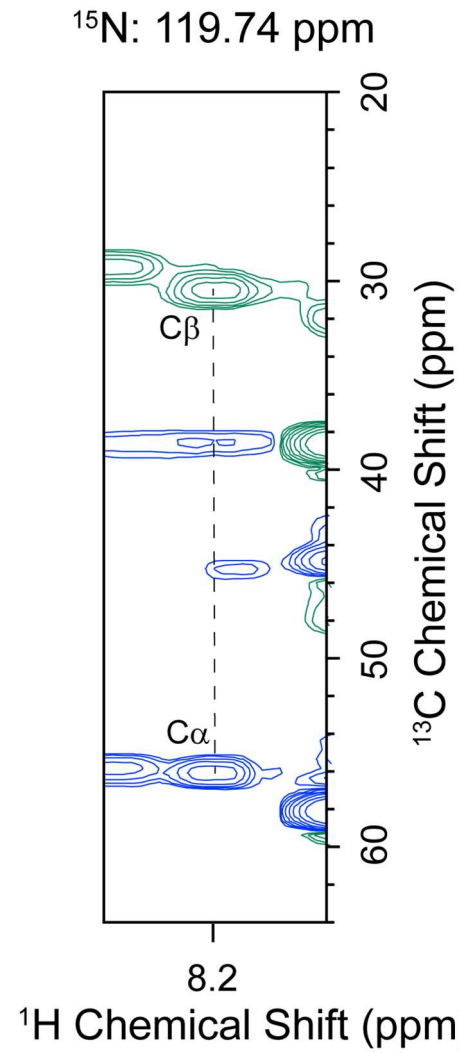



**Figure 45.** Slice taken from the NHCACB experiment at the $^{15}$N chemical shift of Arg. In this experiment, the Cα are shown as blue and Cβ green. Arg shows a slight downfield chemical shift in the carbon dimension, and a slightly upfield chemical shift in the $^{1}$H dimension. The preceding amino acid is Gly at ~ 45 ppm.

$^{15}$N chemical shift to Gln which appears several times throughout the sequence, it was possible to identify Arg by the differences in Cα, Cβ, and $^{1}$H amide chemical shifts (**Figure 45**). Arg Cα and Cβ have chemical shifts downfield from Gln, in addition the $^{1}$H amide peak is slightly more up field than Gln, in agreement with published chemical shifts for Gln and Arg in a random coil.[21]

When the peptide was dissolved in phosphate buffer, this correlated to a shift and/or complete disappearance of a number of peaks (**Figure 44, red**). In the Gly region near 108 ppm, the G**G** dipeptide pair near 8.55 ppm ($^{1}$H) appears to move downfield in the $^{15}$N dimension, taking

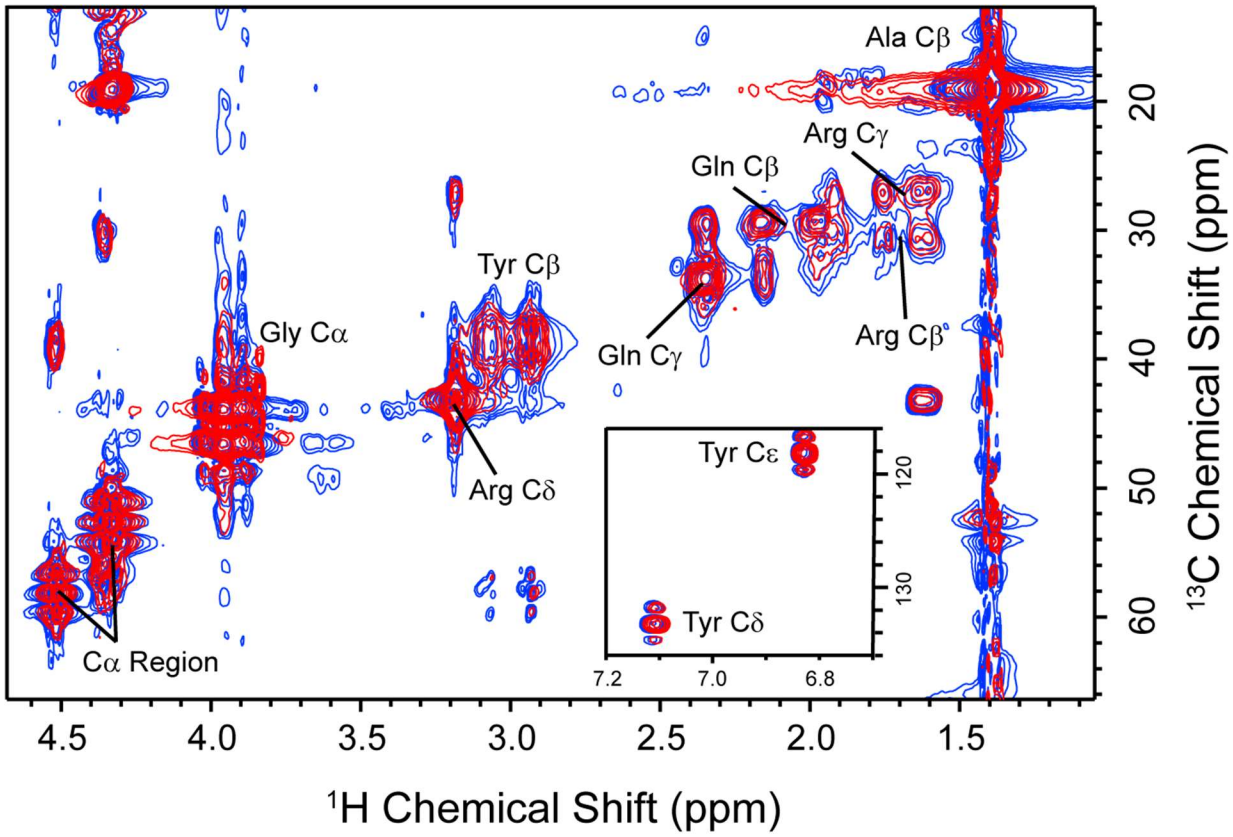



**Figure 46.** $^{1}$H-$^{13}$C HSQC of $^{13}$C/$^{15}$N-labeled MaSp1-like peptide in water, pH 5 (blue) or in phosphate buffer, pH 7.2 (red). Overall, the peaks look similar between each sample including in the aromatic region where Tyr is.

the place of the Q**G** peak previously assigned in water. This Q**G** peak appears to have shifted up field in both the $^{1}$H and $^{15}$N dimension, taking the place of a previous A**G** assignment. In the up field GG region, two of the three peaks appear to coalesce to form only two peaks. In the G**X** region, several of the G**Q** pairs disappear in favor of only two G**Q** peaks in the phosphate sample (**Figure 44, ~ 120 ppm**). This is the same in the Ala region, where several G**A** peaks disappear in favor of only two G**A** peaks (**Figure 44, ~ 123 ppm**). For the first time, Arg becomes well-resolved and has its own peak, with a slight downfield shift in both the $^{1}$H and $^{15}$N dimensions from water, giving it a more α-helical character in phosphate buffer (**Tables 4 and 5**). Both the G**Y** and Y**G** assignments look identical between each sample, suggesting that both GYG sequences in the peptide are similar, and do not change after phosphate is added. Overall, the

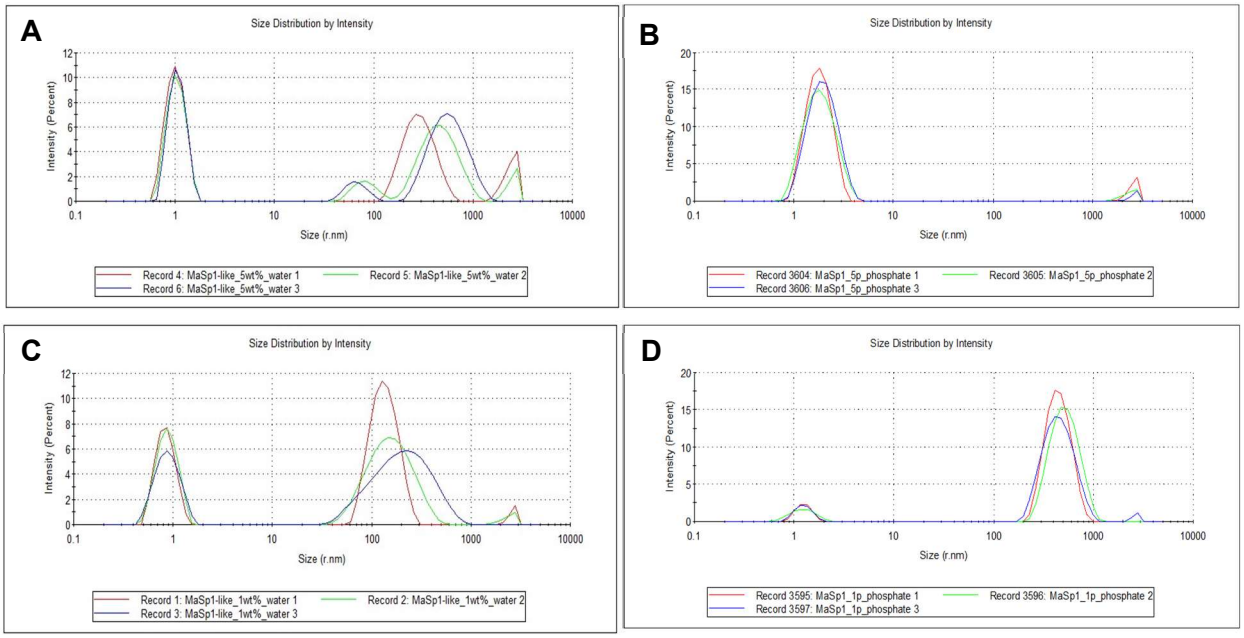



**Figure 47.** Dynamic light scattering (DLS) analysis of MaSp1-like peptide in various solutions. The concentration is 5 wt% for A and B, 1 wt% for C and D. In A and C, the sample is dissolved in water, pH 5. In B and D, the sample is dissolved in phosphate buffer pH 7.2. All samples were centrifuged prior to DLS. Each experiment was repeated three times and overlaid in green, blue and red.

peaks appear broader for the phosphate sample than in water. The $^{1}H$-$^{13}C$ HSQC shows little difference between each sample, including in the aromatic region (**Figure 46**).

These changes in chemical shift are correlated with a change in phase properties of the material to a more gel-like state. Using dynamic light scattering (DLS), we compared our MaSp1-like peptide at 1 and 5 wt%, with and without added phosphate buffer (**Figure 47**). All samples were centrifuged before analysis. In all samples was the presence of a peak near 1 nm which roughly corresponds to our monomeric MaSp1 peptide. In the 5 wt% water sample, there are structures in the 300 – 1000 nm diameter range that disappear when phosphate is added. Upon closer inspection, the phosphate buffer induced a phase change to the material resulting in a gel after centrifugation (not shown). This result is in line that phosphate buffer may help to facilitate protein-protein interactions that are not present in water even at low pH, and that centrifuging the sample removed them from solution. In the 1 wt% sample, the assembly structures are somewhat smaller, and between 100-300 nm in diameter. The addition of phosphate to this sample does not

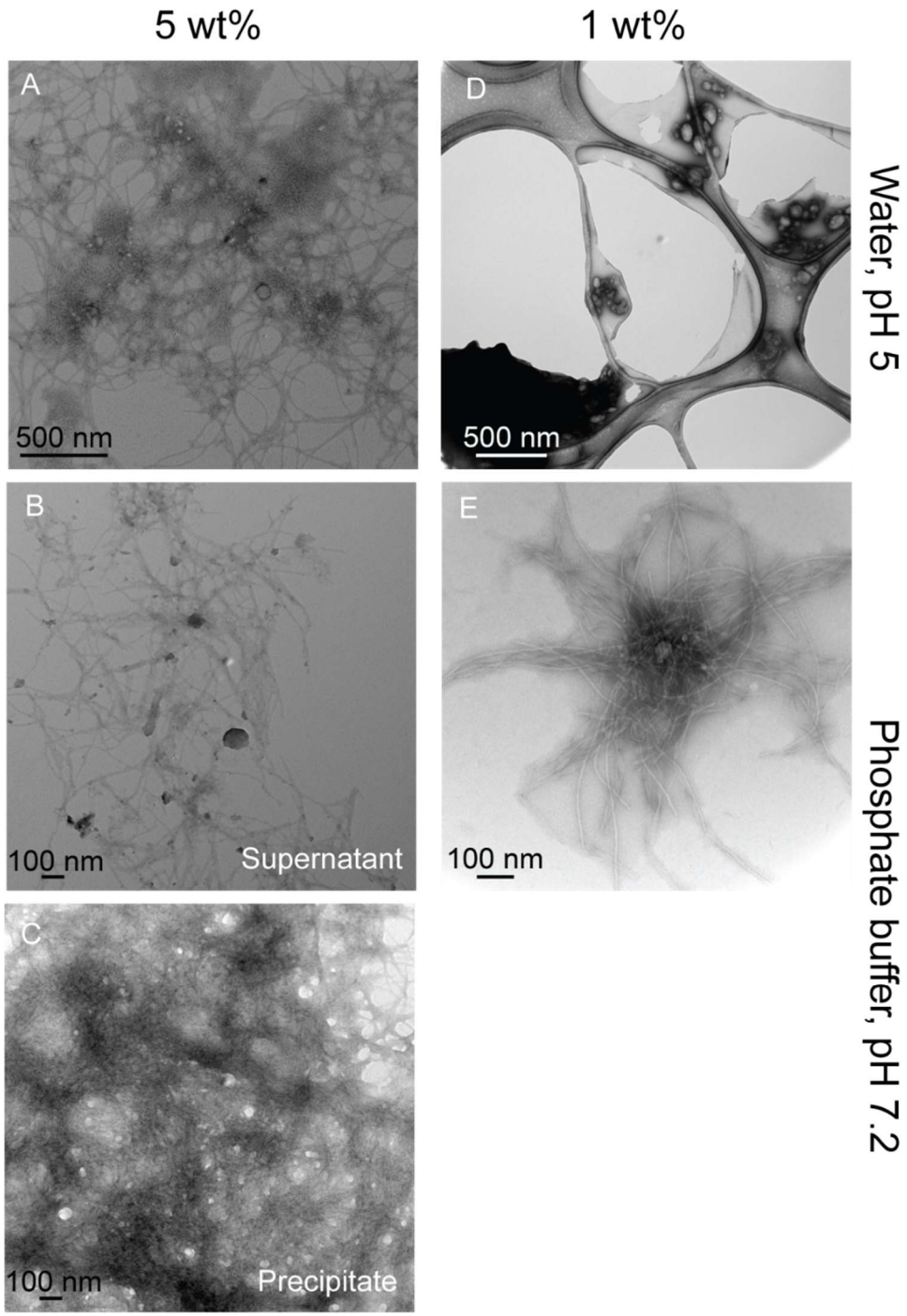



**Figure 48.** Negative staining transmission electron microscopy (NS-TEM) images of MaSp1-like peptide in different solutions. Fibrils appear to form at 5 wt% (A-C) and when the peptide is in phosphate buffer at 1 wt%, E. Pre-assemblies appear in all images and are between 50 – 200 nm.

remove the assemblies, but instead creates a homogenous population that are between 300-600 nm in diameter that persists even with centrifugation. The mostly likely explanation for the

**Table 4.** Chemical shifts of MaSp1-like peptide in water pH 5 referenced to DSS.

| | $^{15}$N | | | | $^{13}$Cα | | | | $^{13}$Cβ | | | | $^{1}$H (amide) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Peptide | α-helix | RC | β-sheet | Peptide | α-helix | RC | β-sheet | Peptide | α-helix | RC | β-sheet | Peptide | α-helix | RC | β-sheet |
| GG, 1 | 108.11 | 107.34 | 109.94 | 110.19 | 45.20 | 45.08 | 45.34 | 47.02 | | | | | 8.20 | 8.23 | 8.34 | 8.27 |
| GG, 2 | 108.40 | | | | 45.31 | | | | | | | | 8.27 | | | |
| GG, 3 | 108.30 | | | | 45.13 | | | | | | | | 8.31 | | | |
| GG, 4 | 108.21 | | | | 44.99 | | | | | | | | 8.54 | | | |
| AG, 1 | 108.62 | | | | 45.24 | | | | | | | | 8.43 | | | |
| AG, 2 | 107.77 | | | | 45.20 | | | | | | | | 8.44 | | | |
| AG, 3 | 108.11 | | | | 45.27 | | | | | | | | 8.49 | | | |
| QG, 1 | 109.50 | | | | 45.32 | | | | | | | | 8.49 | | | |
| QG, 2 | 109.72 | | | | 45.01 | | | | | | | | 8.53 | | | |
| YG | 110.47 | | | | 45.33 | | | | | | | | 8.40 | | | |
| GQ, 1 | 120.15 | 118.59 | 119.73 | 123.14 | 55.95 | 54.33 | 55.94 | 58.61 | 29.21 | 28.33 | 28.67 | 31.92 | 8.20 | 8.11 | 8.25 | 8.51 |
| GQ, 2 | 119.43 | | | | 55.95 | | | | 29.37 | | | | 8.24 | | | |
| GQ, 3 | 119.43 | | | | 55.95 | | | | 29.27 | | | | 8.26 | | | |
| GQ, 4 | 119.55 | | | | 55.81 | | | | 29.33 | | | | 8.30 | | | |
| GQ, 5 | 119.52 | | | | 55.95 | | | | 29.27 | | | | 8.38 | | | |
| GR | 120.14 | 118.99 | 120.59 | 122.60 | 56.03 | 59.05 | 55.96 | 54.63 | 30.48 | 30.00 | 30.53 | 32.36 | 8.20 | 8.03 | 8.17 | 8.57 |
| GY | 119.74 | 119.67 | 120.05 | 122.55 | 58.03 | 56.56 | 57.72 | 61.07 | 38.69 | 40.79 | 38.71 | 38.38 | 8.14 | 8.10 | 7.90 | 8.69 |
| GA, 1 | 123.47 | 121.65 | 123.52 | 125.57 | 52.40 | 50.86 | 52.67 | 54.86 | 19.15 | 21.72 | 19.03 | 18.27 | 8.24 | 7.99 | 8.11 | 8.59 |
| GA, 2 | 123.42 | | | | 52.55 | | | | 19.02 | | | | 8.28 | | | |
| GA, 3 | 123.71 | | | | 52.55 | | | | 19.02 | | | | 8.31 | | | |
| GA, 4 | 123.73 | | | | 52.63 | | | | 19.16 | | | | 8.45 | | | |
| GA, 5 | 123.50 | | | | 52.66 | | | | 19.21 | | | | 8.67 | | | |
| A | 124.05 | | | | 52.49 | | | | 19.10 | | | | 8.51 | | | |

**Table 5.** Chemical shifts of MaSp1-like peptide in phosphate buffer pH 7.2 referenced to DSS.

| | $^{15}N$ | | | | $^{13}C\alpha$ | | | | $^{13}C\beta$ | | | | $^{1}H$ (amide) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Peptide | α-helix | RC | β-sheet | Peptide | α-helix | RC | β-sheet | Peptide | α-helix | RC | β-sheet | Peptide | α-helix | RC | β-sheet |
| GG, 1 | 108.20 | 107.34 | 109.94 | 110.19 | 45.02 | 45.08 | 45.34 | 47.02 | | | | | 8.19 | 8.23 | 8.34 | 8.27 |
| GG, 2 | 108.40 | | | | 45.03 | | | | | | | | 8.27 | | | |
| GG, 3 | - | | | | - | | | | | | | | - | | | |
| GG, 4 | - | | | | - | | | | | | | | - | | | |
| GG, * | 109.66 | | | | 45.01 | | | | | | | | 8.52 | | | |
| AG, 1 | - | | | | - | | | | | | | | - | | | |
| AG, 2 | 107.80 | | | | 45.19 | | | | | | | | 8.44 | | | |
| AG, 3 | 108.02 | | | | 45.50 | | | | | | | | 8.50 | | | |
| QG, 1 | - | | | | - | | | | | | | | - | | | |
| QG, 2 | - | | | | - | | | | | | | | - | | | |
| GQ, * | 108.65 | | | | 45.12 | | | | | | | | 8.43 | | | |
| YG | 110.46 | | | | 45.22 | | | | | | | | 8.40 | | | |
| | | | | | | | | | | | | | | | | |
| GQ, 1 | - | 118.59 | 119.73 | 123.14 | - | 54.33 | 55.94 | 58.61 | - | 28.33 | 28.67 | 31.92 | - | 8.11 | 8.25 | 8.51 |
| GQ, 2 | 119.48 | | | | 55.89 | | | | 29.20 | | | | 8.26 | | | |
| GQ, 3 | - | | | | - | | | | - | | | | - | | | |
| GQ, 4 | - | | | | - | | | | - | | | | - | | | |
| GQ, 5 | 119.54 | | | | 55.94 | | | | 29.21 | | | | 8.08 | | | |
| | | | | | | | | | | | | | | | | |
| GR | 120.48 | 118.99 | 120.59 | 122.60 | 56.18 | 59.05 | 55.96 | 54.63 | 30.14 | 30.00 | 30.53 | 32.36 | 8.21 | 8.03 | 8.17 | 8.57 |
| | | | | | | | | | | | | | | | | |
| GY | 119.72 | 119.67 | 120.05 | 122.55 | 56.16 | 56.56 | 57.72 | 61.07 | 38.52 | 40.79 | 38.71 | 38.38 | 8.14 | 8.10 | 7.90 | 8.69 |
| | | | | | | | | | | | | | | | | |
| GA, 1 | - | 121.65 | 123.52 | 125.57 | | 50.86 | 52.67 | 54.86 | - | 21.72 | 19.03 | 18.27 | - | 7.99 | 8.11 | 8.59 |
| GA, 2 | 123.48 | | | | 52.50 | | | | 18.92 | | | | 8.28 | | | |
| GA, 3 | - | | | | - | | | | - | | | | - | | | |
| GA, 4 | 123.75 | | | | 52.56 | | | | 18.97 | | | | 8.46 | | | |
| GA, 5 | - | | | | - | | | | - | | | | - | | | |
| A | - | | | | - | | | | - | | | | - | | | |

*New peak not in Table 4.

differences between samples are that we are at the critical concentration in the 5 wt% sample with the addition of phosphate, whereas the protein concentration is too low in the 1 wt% sample and the addition of phosphate does not encourage larger assemblies. This result is in line with our recent findings that the critical aggregation concentration of native MaSp silk proteins in 4 M urea is approximately 4 wt%.[19] This is the first evidence that the Gly-Gly-X regions of spider silk proteins may be responsible for facilitating the large pre-assemblies found within the native gland. NS-TEM images of 1 wt% MaSp1 peptide samples with and without phosphate buffer help to corroborate what was seen with DLS. At 5 wt%, there are spherical structures and fibrils in both the water and phosphate added samples (**Figure 48, A-C**). In the 5 wt% water sample, the pre-assemblies appear to be between 50-150 nm. The fibrils are joined and form a web-like structure found throughout the TEM grids. When phosphate buffer is used and the sample is centrifuged, the precipitate again shows both these fibril structures as well as the spherical pre-assemblies. In the 1 wt% sample, the fibers are absent but appear after phosphate is added. The occurrence of both fibrils and assemblies have been seen previously in samples of bovine insulin which coalesce to form spherulites and subsequently sonicated producing fibrils.[22] This demonstrates that the fibrils can be produced either with phosphate, or high concentration of protein, and may reflect a critical aggregation limit for this peptide.

**Conclusions**

Understanding the nature of the hierarchical assembly of silk proteins in the gland is critical in making synthetic versions of this impressive material. The transitions in silk protein structure that occur at this mesoscale regime (100's of nanometers) are poorly understood, due in part to the multitude of rapid physiological changes that occur to convert an intrinsically-disordered, water-soluble protein to fiber. For the first time, we are able to use silk peptide mimics to study the assembly of these proteins and found that they behaved similarly to native-sized proteins. DLS and NS-TEM revealed the presence of large assemblies at 5 wt%, which form a gel network when phosphate is added. NMR at low pH reveals almost twice as many peaks than at high pH,

indicative of backbone hydrogen bonding. Together, the data show that when high pH phosphate buffer is added, aggregates occur but remain dynamic. Without phosphate and at lower pH, the proteins form assemblies but are smaller than samples with phosphate. Our group will continue to explore using peptides to understand the structural changes that may occur in native silks during the spinning process, and is the current focus of our work.

**References**


1. Collins, J. M.; Porter, K. A.; Singh, S. K.; Vanier, G. S., High-efficiency solid phase peptide synthesis (HE-SPPS). *Org Lett* **2014,** *16* (3), 940-3.

2. Suzuki, Y.; Yamazaki, T.; Aoki, A.; Shindo, H.; Asakura, T., NMR study of the structures of repeated sequences, GAGXGA (X = S, Y, V), in Bombyx mori liquid silk. *Biomacromolecules* **2014,** *15* (1), 104-12.

3. Suzuki, Y.; Aoki, A.; Nakazawa, Y.; Knight, D. P.; Asakura, T., Structural Analysis of the Synthetic Peptide (Ala-Gly-Ser-Gly-Ala-Gly)5, a Model for the Crystalline Domain of Bombyx mori Silk Fibroin, Studied with13C CP/MAS NMR, REDOR, and Statistical Mechanical Calculations. *Macromolecules* **2010,** *43* (22), 9434-9440.

4. Yazawa, K.; Yamaguchi, E.; Knight, D.; Asakura, T., 13 C solid-state NMR study of the 13 C-labeled peptide, (E)8 GGLGGQGAG(A)6 GGAGQGGYGG as a model for the local structure of Nephila clavipes dragline silk (MaSp1) before and after spinning. *Biopolymers* **2012,** *97* (6), 347-54.

5. Asakura, T.; Matsuda, H.; Aoki, A.; Kataoka, N.; Imai, A., Conformational change of $^{13}$C-labeled 47-mer model peptides of Nephila clavipes dragline silk in poly(vinyl alcohol) film by stretching studied by $^{13}$C solid-state NMR and molecular dynamics simulation. *Int J Biol Macromol* **2019,** *131*, 654-665.

6. Asakura, T.; Ohgo, K.; Ishida, T.; Taddei, P.; Monti, P.; Kishore, R., Possible Implications of Serine and Tyrosine Residues and Intermolecular Interactions on the Appearance of Silk I Structure of Bombyx mori Silk Fibroin-Derived Synthetic Peptides: High-Resolution 13C Cross-Polarization/Magic-Angle Spinning NMR Study. *Biomacromolecules* **2005,** *6*, 468-474.

7. Ohgo, K.; Kawase, T.; Ashida, J.; Asakura, T., Solid-state NMR analysis of a peptide (Gly-Pro-Gly-Gly-Ala)6-Gly derived from a flagelliform silk sequence of Nephi. *Biomacromolecules* **2006,** *7*.

8. Tetsuo Asakura, M. O., Yasumoto Nakazawa, Kazuo Yamauchi, Structural Analysis of Alanine Tripeptide with Antiparallel and Parallel ß-Sheet Structures in Relation to the Analysis of Mixed ß-Sheet Structures in Samia cynthia ricini Silk Protein Fiber Using Solid-State NMR Spectroscopy. *JACS* **2006,** *128*, 6231-6238.

9. Parent, L. R.; Onofrei, D.; Xu, D.; Stengel, D.; Roehling, J. D.; Addison, J. B.; Forman, C.; Amin, S. A.; Cherry, B. R.; Yarger, J. L.; Gianneschi, N. C.; Holland, G. P., Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks. *Proc Natl Acad Sci U S A* **2018,** *115* (45), 11507-11512.

10. Lin, T. Y.; Masunaga, H.; Sato, R.; Malay, A. D.; Toyooka, K.; Hikima, T.; Numata, K., Liquid Crystalline Granules Align in a Hierarchical Structure To Produce Spider Dragline Microfibrils. *Biomacromolecules* **2017,** *18* (4), 1350-1355.

11. Eisoldt, L.; Smith, A.; Scheibel, T., Decoding the secrets of spider silk. *Materials Today* **2011,** *14* (3), 80-86.

12. Braun, F., Modelling self assembly of natural silk solutions. *International Journal of Biological Macromolecules* **2003,** *32* (3-5), 59-65.

13. Knight, D. P.; Vollrath, F., Spinning an elastic ribbon of spider silk. *Philos Trans R Soc Lond B Biol Sci* **2002,** *357* (1418), 219-27.

14. Knight, D. P.; Vollrath, F., Changes in element composition along the spinning duct in a Nephila spider. *Naturwissenschaften* **2001,** *88*.

15. Rammensee, S.; Slotta, U.; Scheibel, T.; Bausch, A. R., Assembly mechanism of recombinant spider silk proteins. *Proc Natl Acad Sci U S A* **2008,** *105* (18).

16. Malay, A. D.; Suzuki, T.; Katashima, T.; Kono, N.; Arakawa, K.; Numata, K., Spider silk self-assembly via modular liquid-liquid phase separation and nanofibrillation. *Science Advances* **2020,** *6*, 1-12.

17. Lee, W.; Tonelli, M.; Markley, J. L., NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **2015,** *31* (8), 1325-7.

18. Xu, D.; Yarger, J. L.; Holland, G. P., Exploring the backbone dynamics of native spider silk proteins in Black Widow silk glands with solution-state NMR spectroscopy. *Polymer* **2014,** *55* (16), 3879-85.

19. Onofrei, D.; Stengel, D.; Jia, D.; Johnson, H. R.; Trescott, S.; Soni, A.; Addison, B.; Muthukumar, M.; Holland, G. P., Investigating the Atomic and Mesoscale Interactions that Facilitate Spider Silk Protein Pre-Assembly. *Biomacromolecules* **2021**.

20. Ayoub, N. A.; Garb, J. E.; Tinghitella, R. M.; Collin, M. A.; Hayashi, C. Y., Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One* **2007,** *2* (6), e514.

21. Wang, Y.; Jardetzky, O., Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* **2002,** *11* (4), 852-61.

22. Krebs, M. R. H.; MacPhee, C. E.; Miller, A. F.; Dunlop, I. E.; Dobson, C. M.; Donald, A. M., The formation of spherulites by amyloid fibrils of bovine insulin. *Proc Natl Acad Sci U S A* **2004,** *101* (40).

**Chapter 5.** Structure and Hierarchical Organization of Recombinant Spider Silk Proteins and Fibers

## Introduction

Spider dragline silk exhibits extraordinary mechanical properties combining a moderate strength with good extensibility resulting in a toughness exceeding that of all other natural or synthetic fibers.[1-3] Although spider silk has been the focus of research for decades, the mechanical properties of reconstituted man-made fibers have never reached those of natural spider silk. These properties are based on the underlying spider silk proteins, their self-assembly within the gland, and the highly-controlled subsequent spinning process. Currently, there are several groups trying to understand the spinning process from soluble proteins to fibers using recombinant proteins. However, none have provided a detailed and holistic view of the structure of the soluble proteins and the true hierarchical structure of the fibers that result. Here, we utilize both liquid- and solid-state NMR, dynamic light scattering (DLS) and light microscopy to study the differences in a recombinant Masp2-like spider silk dope reconstituted under native and non-native conditions as well as the subsequent fibers produced after extrusion. Our goal is to provide

GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ
GAASAAVSVGGYGPQSSSAPVASAAASRLSSPAASSRVSSAVSSLVS
SGPTNQAALSNTISSVVSQVSASNPGLSGCDVLVQALLEVVSALVSILG
SSSIGQINYGASAQYTQMetVGQSVAQALAG

**Figure 49**. Full sequence of $(AQ)_{12}$NR3 used throughout this chapter to study the structure and hierarchical organization of silk proteins. The last three lines correspond to the C-terminus.

a full picture, from starting protein dope to final fiber, of how the differences in silk protein preparation dictate the downstream mechanical properties of fibers.

Producing recombinant spider silk that has the same mechanical properties as its natural counterpart suffers from several key knowledge gaps. These gaps partially lie in understanding the atomic-to-mesoscale regime (10s-100s nm) of protein structure and how these translate to macroscale material properties. These interactions occur while the protein is still in the soluble state and are influenced by the salt content, pH, and concentration which change as the proteins travel down the duct during the spinning process.[4] Several groups have worked toward a complete understanding of this process although none have provided a dramatic improvement to the silk community.[5] Again, the mesoscale regime where silk pre-assemblies are converted to fibers within the duct is poorly understood. This is in part due to the lack of resolution afforded by techniques common to studying transitions in protein structure, the nature of silk proteins to aggregate, and the repetitiveness of the protein making it difficult to isolate a region that could be responsible for this transition. To overcome this limitation, we use NMR to investigate the structural changes that occur on a recombinant silk protein in dope form which has been previously shown to produce fibers that rival native materials.[6]

In the ideal case, we would want to mimic what Mother Nature has already done; use an aqueous-based native-like spinning apparatus to form synthetic fibers. Fortunately, there is another group working to solve the spider silk problem in Bayreuth, Germany who has optimized a process for generating and processing recombinant silk materials. The Scheibel group in Germany has worked for decades to understand the silk spinning process and uses recombinant proteins in order to make biomimetic spider silk.[7] They have published several papers detailing the silk assembly process and have investigated the differences in buffer, pH and various salts on the mechanical properties of spun recombinant fibers.[6] Together, we have built on previous work using MaSp2-inspired protein to study the fiber-spinning process. The recombinant protein, $(AQ)_{12}$NR3, has been studied in a variety of iterations since it can be readily modified to include

an N-terminus, additional (AQ) repeats, as well as no C-terminus.[6, 7] $(AQ)_{12}$NR3 is comprised of 44 residue blocks repeated 12 times, followed by a non-repetitive C-terminus bringing the molecular weight to ~ 59 kDa (**Figure 49**).[7] It has been previously shown that the C-terminus facilitates pH-dependent dimerization during the dope reconstitution protocol, creating a protein with a molecular weight of approximately 120 kDa.[7] This sequence is inspired from MaSp2, having shorter poly$(Ala)_6$ runs compared to lengths of… found in MaSp1, juxtaposed Gln-Gln dipeptide repeats, and a number of Pro residues, the hallmark of MaSp2.[7]

During our collaboration, we would receive lyophilized $(AQ)_{12}$NR3 protein and reconstitute it into either a biomimetic spinning dope (BSD) or classical spinning dope (CSD) following previous published protocols.[6] The difference being the type of buffer, where the BSD is spun from a phosphate-containing buffer and the CSD a tris-containing buffer. We then investigated the differences in the prepared dope morphology and structure, as well as the final fibers that were spun in Bayreuth and sent to us.

**Methods**

**(AQ)12NR3 Dope Preparation and Labeling**

Two types of labeled protein were used throughout the NMR experiments; proteins sparingly labeled with 1-$^{13}$C-glucose/$^{15}NH_4Cl$, and those fully-labeled with $^{13}$C-glucose/$^{15}NH_4Cl$. The $^{13}$C enrichment content was variable in the sparingly-labeled material, reaching a maximum of approximately 30% for specific sites as anticipated.[8] Fully-labeled materials were $^{13}$C-enriched to approximately 85% for several amino acids including Gly, Ala, and Gln. BSD and CSD were created from the SOP “Preparation of Biomimetic Spinning Dope” developed by the biomaterials group at the University of Bayreuth (Scheibel Lab). Briefly, the protein was first dissolved in 6 M GdmSCN for several hours at a concentration between 10 – 25 mg/mL, then loaded into a dialysis bag and dialyzed against a 50 mM Tris/HCl + 100 mM NaCl pH 8.0 buffer overnight, then again for four hours the following day using fresh buffer which creates classical spinning dope (CSD).

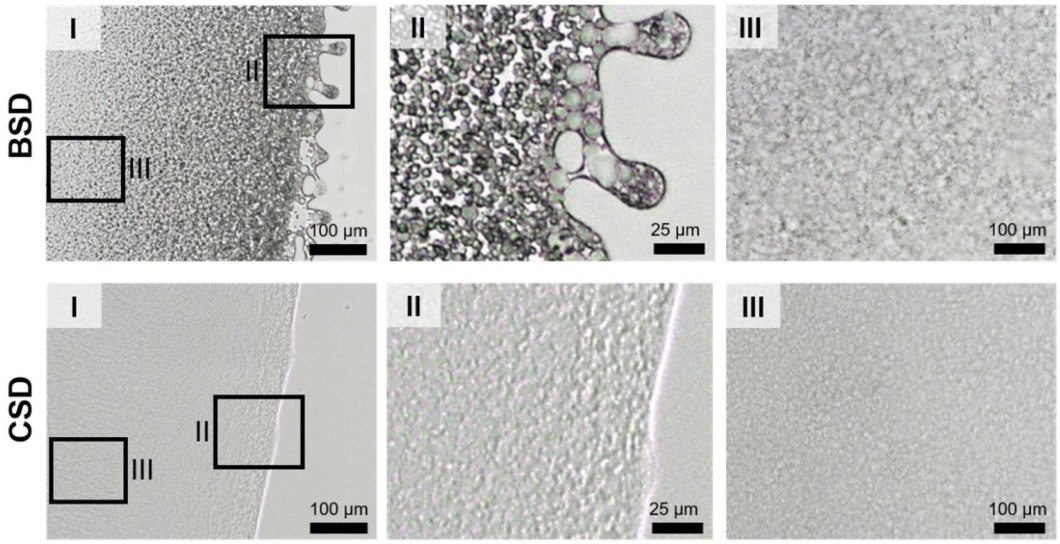



**Figure 50**. Light microscopy pictures of BSD and CSD dopes. Regions from II and III are enlarged to see the differences between the materials, where the BSD contains large droplets indicative of liquid-liquid phase separation.

After, the dialysis bag is finally dialyzed against a 50 mM sodium phosphate buffer pH 7.2 overnight. The final solution is then centrifuged to pellet the dope, creating biomimetic spinning dope (BSD).

**Liquids NMR Experiments**

$^{1}$H, preliminary HSQC, and relaxation experiments were conducted at SDSU on a 600 MHz Bruker AVANCE III spectrometer with a 5 mm TXI HCN solution probe. The temperature was 25°C for all experiments. High-resolution $^{1}$H-$^{15}$N HSQC experiments were conducted at Arizona State University in the Magnetic Resonance Research Center using a Bruker AVANCE III 850 MHz spectrometer with a cryogenic HCN probe. 16 scans with 384 points in the indirect dimension were collected with 16 dummy scans and recycle delay of 1.3 s for HSQC experiments. The $^{1}$H and $^{15}$N sweep widths were 10 and 22 ppm, respectively. For 3D HNCACB experiments,[9] 64 scans were collected with 64 points in the $^{13}$C dimension and 112 points in the $^{15}$N dimension with 32 dummy scans and a recycle delay of 1 s. The sweep width for the $^{1}$H, $^{13}$C, and $^{15}$N dimensions were 10, 22, and 74 ppm, respectively. Both experiments used 50 % non-uniform sampling (NUS).

The CSD and BSD materials were only used for a maximum of four days as judged by $^{1}H$-$^{15}N$ HSQC which yielded consistent spectra up to that time. For $T_2$ experiments, the CPMG loops were 2, 48, 4, 16, 8, 24, 12, 32 and staggered this way to account for changes in the material of the course of the experiment. 128 scans were collected with 32 points in the indirect dimension using a recycle delay of 3 s. For $T_1$ experiments, 256 points were collected with a single slice in the indirect dimension using a recycle delay of 3 s. Chemical shifts were referenced to DSS for $^{1}H$ and $^{13}C$, and externally referenced to formamide for $^{15}N$. Secondary structure assignments were based on work from Jardetsky.[10] Approximately 300 μL of material was used for all experiments. The CSD and BSD concentrations were 11.5 and 13 wt% protein, respectively.

**Solid-state NMR Experiments**

$^{1}H$-$^{13}C$ CP-MAS solid-state NMR experiments were done at 30 kHz with 16384 scans and a recycle delay of 3 s. DARR experiments[11, 12] were done at 14 kHz MAS with 256 scans, 256 points in the indirect dimension and a recycle delay of 2.5 s with continuous wave decoupling under a rotary-resonance ($\omega_r = \omega_{rf}$) of 14 kHz. DARR difference experiments[13] were collected at 8-10 kHz MAS with either 4096 or 8192 scans using a recycle delay of 4 s. $^{1}H$-$^{13}C$ HETCOR-FSLG experiments were conducted at 30 kHz MAS with 256 scans and 256 points in the t1 dimension with a recycle delay of 2.5 s. The contact time for all experiments was 1 ms. Approximately 6-8 mg of 1-$^{13}C$-glucose/$^{15}N$-labeled $(AQ)_{12}$NR3 biomimetic and classical fibers were loaded separately into 1.9 mm rotors.

**Results and Discussion**

Light microscopy reveals dramatic morphological differences between the CSD and BSD dope. With the BSD, there are droplets between 5 – 25 μm which are absent in the CSD sample. In the CSD sample, the droplets appear to bleed together creating a single phase. The key

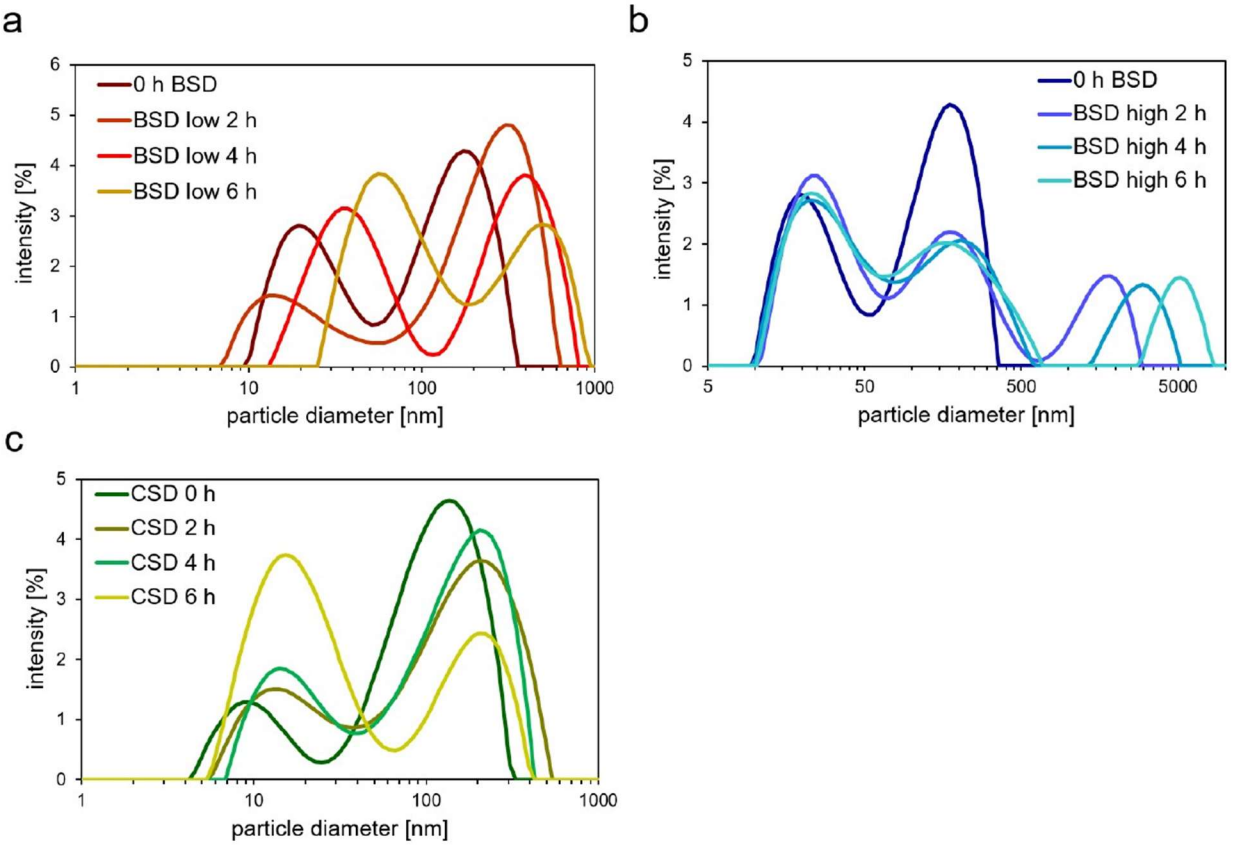



**Figure 51.** Dynamic light scattering of CSD and BSD dopes. a, BSD low concentration phase (supernatant). Concentrations at 0, 2, 4 and 6 hours were 27, 25, 20 and 16 mg/mL, respectively. b, BSD dope after diluting to 50 mg/mL. c, CSD. Concentrations at 0, 2 4 and 6 hrs were 25, 41 106 and 197 mg/mL, respectively.

difference is the choice of buffer, where BSD contains phosphate buffer and CSD tris. Protein droplets arising from liquid-liquid phase separation (LLPS) have been shown to be present in the native gland of the spider and may help to keep the proteins soluble until ready to be spun and could play a critical role in the pre-organization of silk proteins for spinning.[14, 15] This type of phase separation with phosphate has also been seen recently with recombinant MaSp2 systems where the addition of phosphate to recombinant dopes produce a phase separation and subsequent fibrillization when the pH is dropped.[16]

Although we see evidence of pre-assembly droplets with light microscopy, we know from previous work that this behavior also occurs at the mesoscale (100's of nm) for native systems where hierarchical silk protein micelle-like superstructures are observed by cryoEM.[14] We

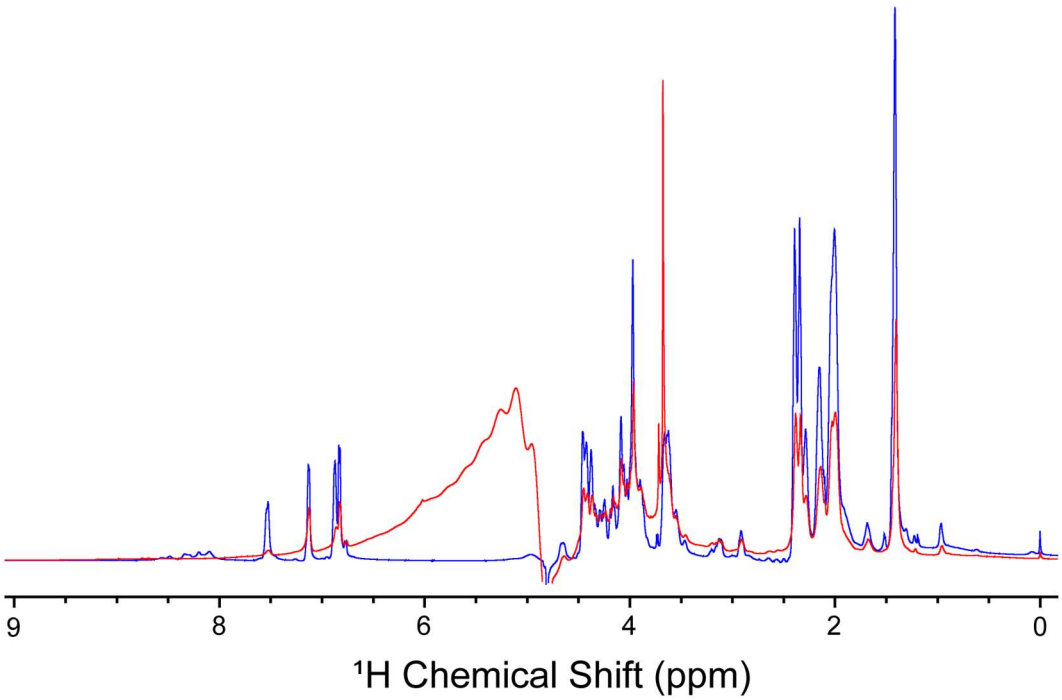



**Figure 52**. $^{1}$H spectra of BSD at 13-15 wt% (blue) and CSD at 12-14 wt% (red). In the amide region (7.8-8.6 ppm) it is possible to see peaks in the BSD spectra which are absent in the CSD spectra. The large peak in the CSD spectra at ~ 3.65 ppm is tris.

therefore explored the size and population distributions of these proteins in solution at this scale. Using DLS (**Figure 51**), we monitored particle size over time and found that the size and population distribution of these pre-as semblies changes depending on buffer type. Typically, there are two major populations reported with DLS in the early dope formation process (**Figure 51, A-C**). For a protein of this size (~120 kDa) the predicted diameter based on the empirical relationship between the number of residues and diameter, would be between approximately 8-26 nm depending on the degree of denaturation.[17] The smaller peak reported from DLS is in line with this approximation for CSD (**Figure 51, C**) and BSD at 0 h (**Figure 51, A**), but not for the other BSD samples (**Figure 51, A, B**). The larger peak between 100-1000 nm is due to the formation of large, hierarchical assemblies and varies with sample and time. Since there are only two peaks in the CSD (**Figure 51, C**) this suggests that during the concentration step of creating the dope, the monomers coalesce to form pre-assemblies in one step, whereas in the BSD this occurs gradually over time (**Figure 51, A**). In the final BSD dope (**Figure 51, B**), there are three

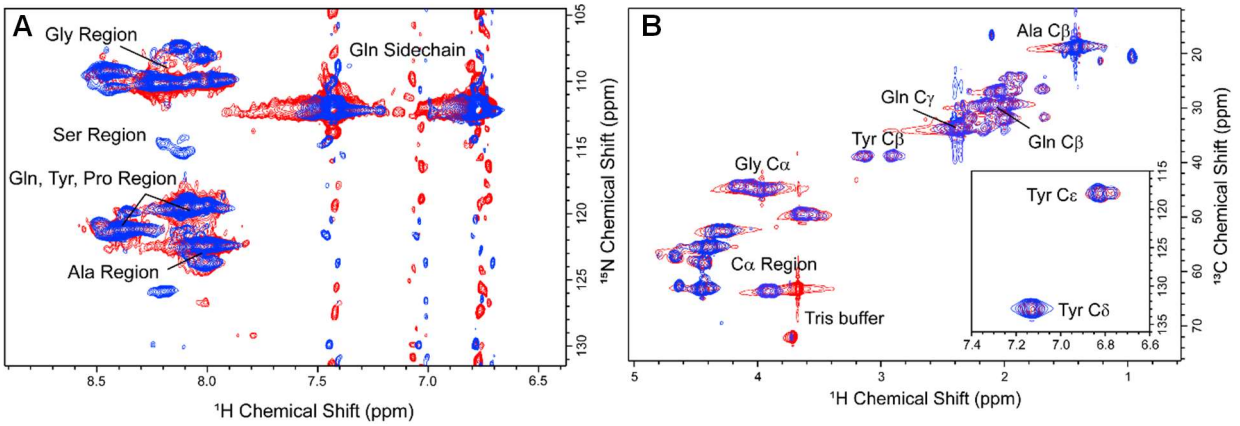



**Figure 53**. HSQC spectra of CSD (red) and BSD (blue). A, $^{1}$H-$^{15}$N HSQC. B, $^{1}$H-$^{13}$C HSQC. Peaks in the $^{1}$H-$^{15}$N spectra are sharper for BSD than CSD, with some peaks in the CSD missing entirely. In the $^{1}$H-$^{13}$C spectra, the chemical shifts are similar including in the aromatic region for Tyr (inset). The protein here was sparingly labeled with $^{13}$C glucose.

sizes present. The peaks at 20 and 200 nm appear consistent, suggesting that when the dope is formed the proteins interact and form a consistent size, driven by the conditions of the buffer. These results are in line with what we saw previously with light microscopy (**Figure 50**), that dopes prepared using a phosphate buffer create droplets larger in size than those created in CSD. To understand the role of phosphate in creating these pre-assemblies, we use NMR to probe the atomic-level interactions that facilitate this behavior.

Using NMR, we see clear differences in the quality of the BSD and CSD spectra that reveal high backbone H-N exchange. In the $^{1}$H spectra (**Figure 52**), resonances corresponding to the H-N amide re (7.8-8.6 ppm) are present in the BSD sample but absent in the CSD. This suggests that, although the CSD can be concentrated to the same wt% as the BSD, $^{1}$H exchange is greater for proteins in CSD. This has a direct impact on the ability to form backbone hydrogen bonding interactions which are necessary to form the pre-assemblies in the dope and the β-sheet structures in the final fibers that are prevalent and responsible for the toughness in silk fibers.[18] We note that these spectra were collected with water suppression, where fast exchange is not recorded during the experiment. Despite this, it was possible to collect HSQC spectra on both materials (**Figure 53**). In the $^{1}$H-$^{15}$N HSQC, we see sharper peaks in the $^{15}$N dimension for the

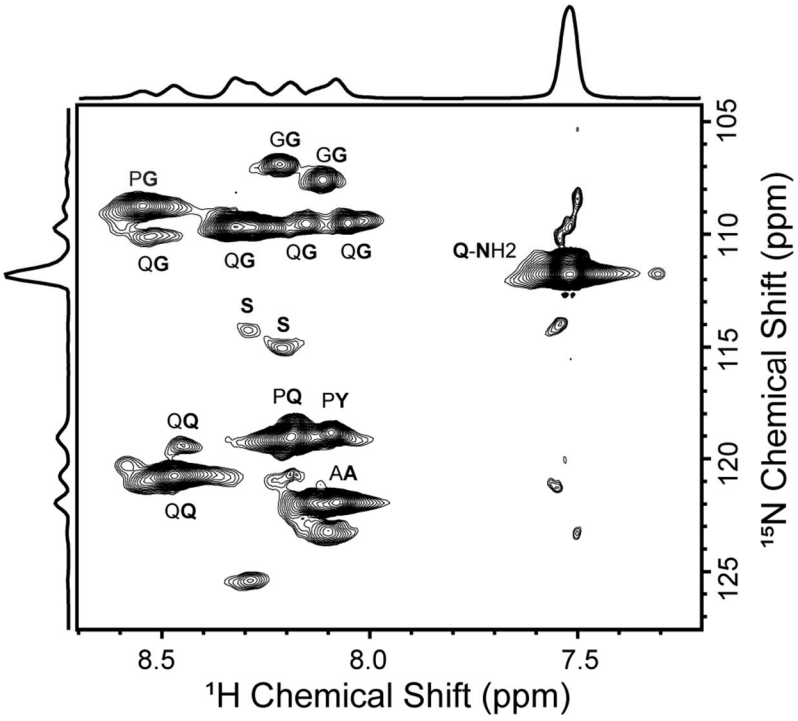



**Figure 54**. $^{1}$H-$^{15}$N HSQC spectra from Figure 53 labeled according to 3D HNCACB assignments. This HSQC was assigned using sparingly labeled glucose.

BSD sample as well as greater intensity (**Figure 53, A blue**). In the CSD spectra, most peaks are broad, disappearing entirely for some resonances (**Figure 53, A red**). This behavior is commonly seen when backbone H-N exchange is too high, preventing efficient INEPT signal transfer from $^{1}$H to $^{15}$N due to chemical exchange occurring on the timescale of the INEPT delays. In the $^{1}$H-$^{13}$C HSQC, the chemical shifts appear to be the same including those in the aromatic region (**Figure 53, B**). This was also the case for the material at low (~ 1wt%) concentration – the BSD and CSD appear identical in the carbon HSQC (data not shown) which again points to the line broadening and disappearance of resonances in the 1H-15N HSQC of CSD to be due to fast proton exchange.

In order to understand what regions of the protein are changing between dope preparations, we utilized 3D HNCACB experiments to assign peaks in the $^{1}$H-$^{15}$N HSQC spectra

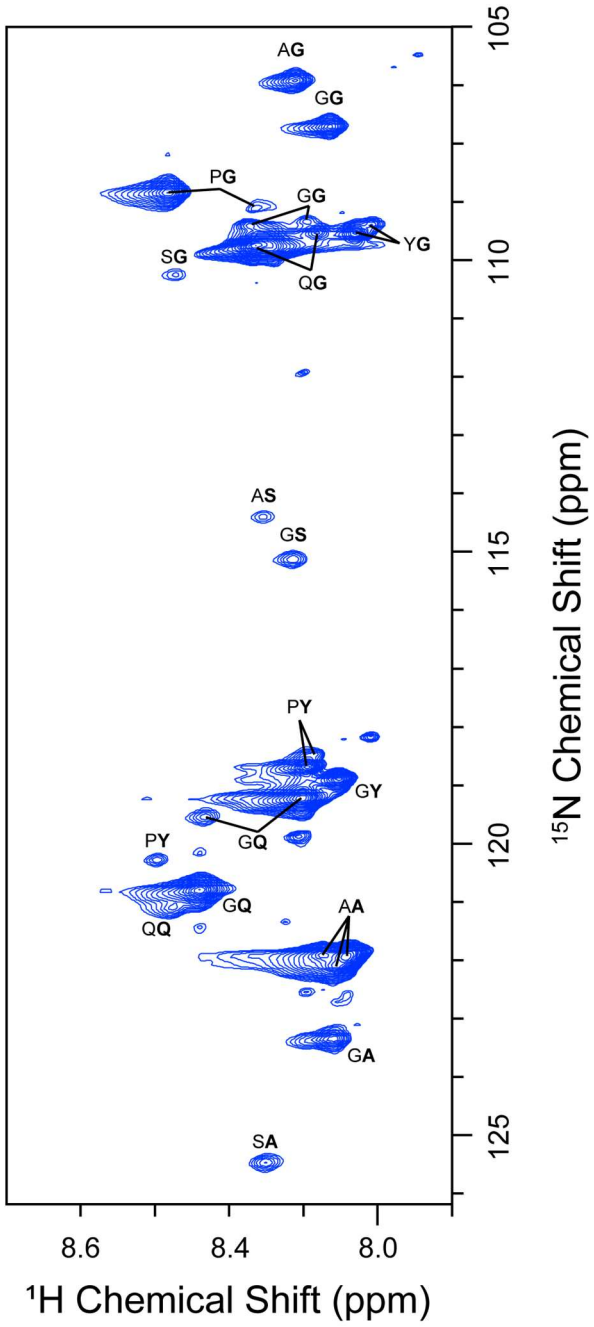



**Figure 55**. $^{1}$H-$^{15}$N HSQC spectra on fully-labeled $^{13}$C/$^{15}$N BSD. Several more peaks are assigned with higher fidelity from Figure 54.

(**Figure 54**). In this experiment, we find the $^{15}$N chemical shift of the "i" amino and correlate that to the Cα and Cβ resonances of the same amino acid. This experiment also has the benefit of revealing the preceding i-1 residue, which helps to further assign longer sequences of the protein. Unfortunately, we could not obtain signals for the CSD sample likely due to the long J-transfer

times needed during a 3D experiment which is negatively impacted by the high H-N exchange rate previously revealed in the $^{1}$H spectra (**Figure 52**). Using the primary sequence for $(AQ)_{12}$NR3 (**Figure 49**) we were able to assign several dipeptide motifs in the $^{1}$H-$^{15}$N HSQC (**Figure 54**). The bold letter indicates the "i" amino acid which is assigned without ambiguity with corresponding Cα and Cβ shifts when applicable. Overall, this spectra is similar to other $^{1}$H-$^{15}$N HSQCs we have collected previously on native MaSp silk dopes, where there are generally four regions residing near $^{15}$N = 110, 115, 120, and 123,[19, 20] corresponding to areas where the "i" amino acid is Gly, Ser, X (X = Gln, Tyr, Pro) and Ala, respectively. Shortly after we collected these initial spectra, we received fully-labeled $^{13}$C material from our collaborators and secured time on a higher-field magnet in an effort to obtain more resolution and better assignment quality (**Figure 55**).

Chemical shifts are tabulated in Table 6 and assigned several of the amino acids as random coil (RC), with a few exceptions of α-helices (I would discuss which ones, why??)_. Since the poly(Ala) motif has been shown to play a critical role in determining the toughness of spider silk fibers,[5, 18, 21] we approach our assignment discussion by framing them with respect to this motif. GG sequences are found exclusively after the $poly(Ala)_6$ motif and exhibit a random coil (RC) secondary structure, with the exception of the assigned G**G** furthest up field (**Table 6**). P**G** sequences are found throughout the entire repeated sequence primarily located in the Gln-rich region (**Figure 49**, 4/5 per repeat). Both report an RC environment for Gly. Also present are two peaks corresponding to Y**G** sequences, where the Gly residue is reported as RC. Y**G** appears in two places throughout the sequence; as PYG preceding the poly(Ala) motif, or GYG following the $poly(Ala)_6$ motif. Both report an RC type environment for Gly. On the other hand, there are several P**Y** motifs and one G**Y** motif, which again are found either preceding or following the poly(Ala) motif. Two of the three P**Y** sequences assign the $^{15}$N chemical shift to an α-helix, although looking further into the chemical shifts of the Cα and Cβ we see that they

**Table 6.** Tabulated chemical shifts from HNCACB strip plots as well as their secondary structure assignments.

| | $^{15}$N | | | | $^{13}$Cα | | | | $^{13}$Cβ | | | | $^{1}$H (amide) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Protein | α-helix | RC | β-sheet | Protein | α-helix | RC | β-sheet | Protein | α-helix | RC | β-sheet | Protein | α-helix | RC | β-sheet |
| G**G**, 1 | 107.72 | 107.34 | 109.94 | 110.19 | 45.37 | 45.08 | 45.34 | 47.02 | | | | | 8.12 | 8.23 | 8.34 | 8.27 |
| G**G**, 2 | 109.41 | | | | 44.90 | | | | | | | | 8.34 | | | |
| G**G**, 3 | 109.33 | | | | 43.90 | | | | | | | | 8.19 | | | |
| P**G**, 1 | 108.84 | | | | 45.58 | | | | | | | | 8.56 | | | |
| P**G**, 2 | 109.08 | | | | 45.21 | | | | | | | | 8.36 | | | |
| A**G** | 106.95 | | | | 45.72 | | | | | | | | 8.22 | | | |
| S**G** | 110.26 | | | | 45.85 | | | | | | | | 8.54 | | | |
| Y**G**, 1 | 109.38 | | | | 45.07 | | | | | | | | 8.07 | | | |
| Y**G**, 2 | 109.53 | | | | 45.06 | | | | | | | | 8.06 | | | |
| Q**G**, 1 | 109.53 | | | | 43.92 | | | | | | | | 8.17 | | | |
| Q**G**, 2 | 109.74 | | | | 44.94 | | | | | | | | 8.33 | | | |
| | | | | | | | | | | | | | | | | |
| A**S** | 114.40 | 114.78 | 115.94 | 117.44 | 59.29 | 60.86 | 58.35 | 57.14 | 64.32 | 62.81 | 63.88 | 65.39 | 8.32 | 8.11 | 8.26 | 8.57 |
| G**S** | 115.12 | | | | 59.38 | | | | 64.41 | | | | 8.23 | | | |
| | | | | | | | | | | | | | | | | |
| G**Q**, 1 | 119.21 | 118.59 | 119.73 | 123.14 | 56.38 | 58.61 | 55.95 | 54.33 | 29.55 | 28.33 | 28.67 | 31.92 | 8.20 | 8.11 | 8.25 | 8.51 |
| G**Q**, 2 | 119.54 | | | | 56.38 | | | | 29.55 | | | | 8.48 | | | |
| G**Q**, 3 | 120.80 | | | | 56.27 | | | | 29.85 | | | | 8.48 | | | |
| Q**Q** | 121.09 | | | | 56.40 | | | | 29.80 | | | | 8.56 | | | |
| | | | | | | | | | | | | | | | | |
| P**Y**, 1 | 118.47 | 119.67 | 120.05 | 122.55 | 57.92 | 61.07 | 57.72 | 56.56 | 38.89 | 38.38 | 38.71 | 40.79 | 8.19 | 8.10 | 7.90 | 8.69 |
| P**Y**, 2 | 118.63 | | | | 57.86 | | | | 38.95 | | | | 8.19 | | | |
| P**Y**, 3 | 120.27 | | | | 57.98 | | | | 39.68 | | | | 8.59 | | | |
| G**Y** | 118.18 | | | | 58.11 | | | | 39.25 | | | | 8.11 | | | |
| | | | | | | | | | | | | | | | | |
| A**A**, 1 | 121.93 | 121.65 | 123.52 | 125.57 | 53.13 | 54.86 | 52.67 | 50.86 | 18.98 | 18.27 | 19.03 | 21.72 | 8.08 | 7.99 | 8.11 | 8.59 |
| A**A**, 2 | 121.90 | | | | 53.37 | | | | 18.97 | | | | 8.14 | | | |
| A**A**, 3 | 122.07 | | | | 53.10 | | | | 19.01 | | | | 8.09 | | | |
| G**A** | 123.35 | | | | 53.24 | | | | 19.37 | | | | 8.12 | | | |
| S**A** | 125.46 | | | | 53.46 | | | | 19.05 | | | | 8.31 | | | |

assign the Tyr to RC. This is likely an effect of the preceding Pro residue which is known to form turn-like structures with neighboring amino acids causing deviations in “normal” secondary structure shifts. Together, this reveals that Tyr residues are most likely RC in $(AQ)_{12}$NR3 prepared as BSD. Gln residues in the repetitive core region are found exclusively as QQ pairs allowing either QG or QQ sequences to be assigned using this experiment. Three G**Q** peaks were assigned as RC. Interestingly, only a single Q**Q** peak was found and assigned as RC / β-sheet. This suggests that overall Gln residues are found as RC in the repetitive region of the protein. Ala residues were found to occupy a range of secondary structures. All A**A** sequences and the G**A** sequence were found to be a-helical / RC. In the Ser region near 115 ppm, we see two peaks corresponding to either an A**S** or G**S** sequence. A**S** sequences are predominantly found in the repetitive core of the protein with a few in the C-terminal (**Figure 49**). These are found at the beginning of the poly(Ala) motif and are assigned as α-helical / RC. A**A** sequences occur primarily in the core repetitive region of the protein within $poly(Ala)_6$ and are assigned as α-helical / RC. The S**A** sequence is reported to be β-sheet in character. Together, this suggests that the three residues of ASA found at the beginning of the $poly(Ala)_6$ motif transition from RC, to a-helical, to β-sheet as we move into the $poly(Ala)_6$.

Ser/Ala interactions have been implicated in the transition from soluble protein to solid fiber for a number of silks including caddisfly, spider prey-wrapping, and silkworm silks.[22-24] Recently our group showed that SA and AS sequences in spider prey-wrapping silks may play a critical role in the β-sheet crosslinking of individual fibers upon the addition of water.[25] These SA and AS pairs are found predominantly in the disordered regions of the protein, or flanking the well-folded regions.[22, 25] Upon the addition of water, we see a transition from RC to β-sheet structure from Ala and Ser with a further increase upon the addition of shear force.[25] These motifs may act as a trigger for the liquid dope, where their structure in the dope may be vital to the pre-

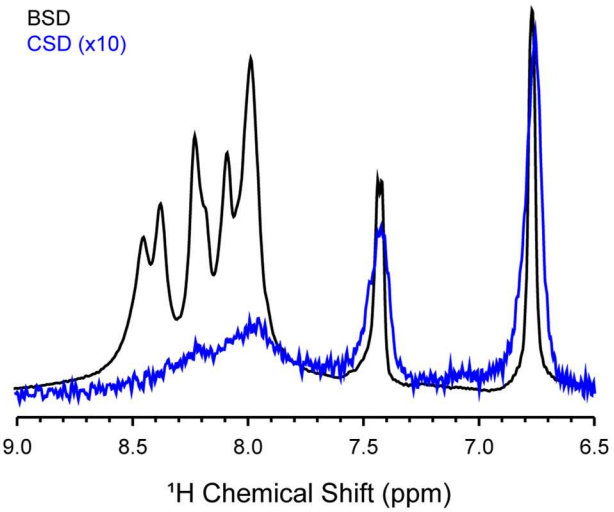



**Figure 56.** First slice from the $T_1$ experiment comparing the amide region in BSD and CSD. The signal intensity for CSD was increased 10X.

assembly process ensuring that the protein transitions from RC-like structure to β-sheet during fiber formation to create a fiber with maximum toughness.

To further understand the differences between the BSD and CSD, we probed the dynamics of both using $T_1$ and $T_2$ relaxation. Generally, $T_1$ relaxation reports on the overall tumbling of the molecular system whereas $T_2$ reports on local domain motions.[26] As we began the experiment, we noticed that it was difficult to obtain enough points to obtain an accurate measurement of $T_1$ for CSD (**Figure 56**) due to the broad nature of these resonances as discussed above. We therefore integrated the entire region from 7.7-8.7 ppm to get a general fit for the overall $T_1$ time of the backbone for the CSD material. Figure 57 shows the difference in $T_1$ relaxation times obtained for several sequences within $(AQ)_{12}NR3$ assigned from Table 8. Both CSD and BSD have comparable backbone $T_1$ relaxation times of approximately 0.7 s (**Figure 57**). Analyzing the Gln side chain, we see that the values are slightly different between the two materials. Gln is of interest to us because it is found ubiquitously throughout silks and can engage in hydrogen bonding which may facilitate protein-protein interactions. The BSD Gln side chains can be consistently fit to a single component, with a relaxation time between 0.5-0.7 s. In the CSD

material, we were able to fit the Gln side chains to two populations with vastly different relaxation

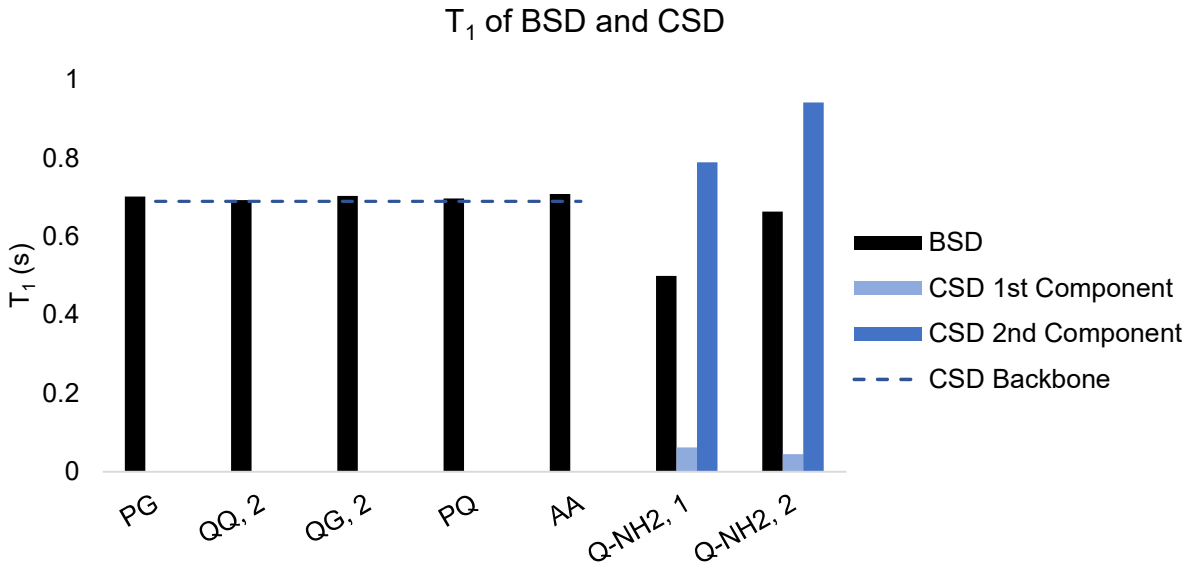



**Figure 57.** $T_1$ fits for BSD (black) and CSD (blue) assigned from Figure 54. Some of the BSD and several of the CSD peaks did not have enough points to be fit. Therefore, the CSD amide region was integrated as a whole, between 7.7 – 8.7 ppm. The dashed line corresponds this $T_1$ time for CSD. Numbers correspond to dipeptide repeats in Table 8. The glutamine side chain consistently fit to two components in the CSD material.

times of 0.05 s and between 0.8-1.0 s. The two different relaxation times can be ascribed to two populations of Gln; one experi encing fast tumbling with a $T_1$ of 0.8-1.0 s, and the other experiencing slow tumbling with a $T_1$ of 0.05 s. The implication here is that one population of Gln may be interacting with other nearby residues through hydrogen-bonding which would slow the rotation of the molecule leading to a shorter relaxation time. Overall, these $T_1$ relaxation times measured for both BSD and CSD are longer than $T_1$ times in native glands which are between 0.3-0.5 s.[19] This suggests that the tumbling of these recombinant proteins is faster than with native systems as expected because of their lower MW. This seems likely considering native protein concentrations are between 25-50 wt%,[27] whereas the BSD and CSD here are between 10-15 wt%.

$T_2$ times were calculated for several sequences within the BSD (**Figure 58**). Only a few of the same sequences were able to be determined in the CSD material, again due to the fast

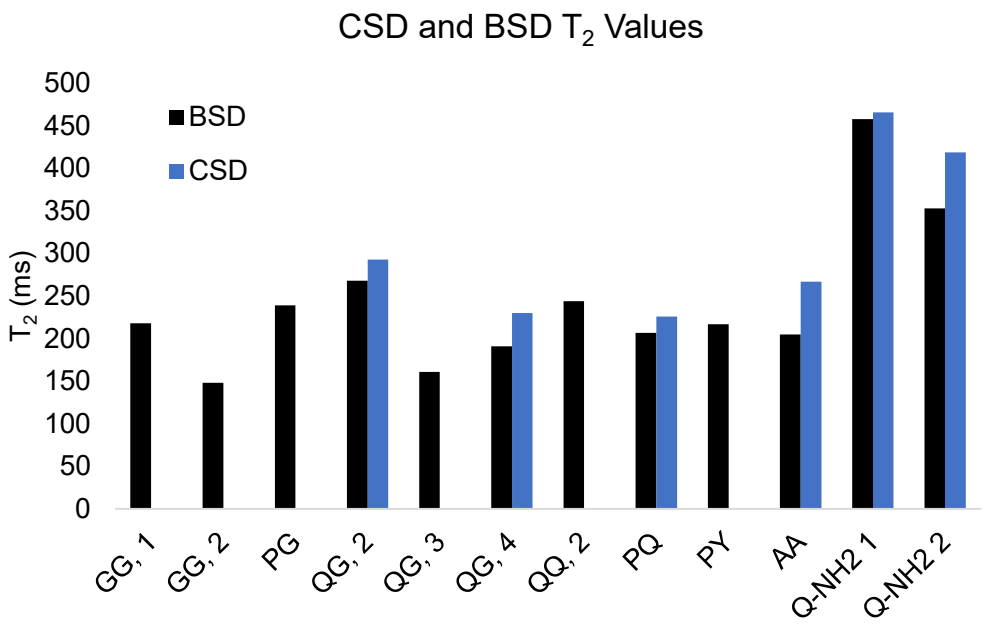



**Figure 58.** Reported $T_2$ values for several residues in $(AQ)_{12}NR3$ prepared as BSD (black) or CSD (blue). Errors were within 1-2 ms for all samples. Because of the low signal, several BSD and most CSD times could not be determined. Numbers next to dipeptide repeats correspond to Table 8.

exchange leading to short relaxation times. Of the CSD times that were determined, all were similar to the BSD material (**Figure 58**) and between 150-500 ms. Overall, these $T_2$ times are similar to those in native glands which are also between 150-500 ms.[19] This reveals that the local motions – side chain rotations, loop motions – are similar between these recombinant BSD and CSD systems and native glands.

Together, these data reveal that there is considerable exchange occurring in the CSD material, but not the BSD material. The downstream effect of this reduced exchange could encourage hydrogen-bonding interactions that form during the spinning process to create solid fibers. The role of pH and buffer choice has been explored previously and is correlated to the quality of fiber.[6, 28] Generally, kosmotropic salts like phosphate encourage the stabilization of water, thus creating protein-protein interactions through a “salting-out” effect.[6]

In addition to characterizing recombinant proteins in the liquid state, we also investigated differences in secondary structure between sparsely-labeled $(AQ)_{12}NR3$ proteins spun into fibers.

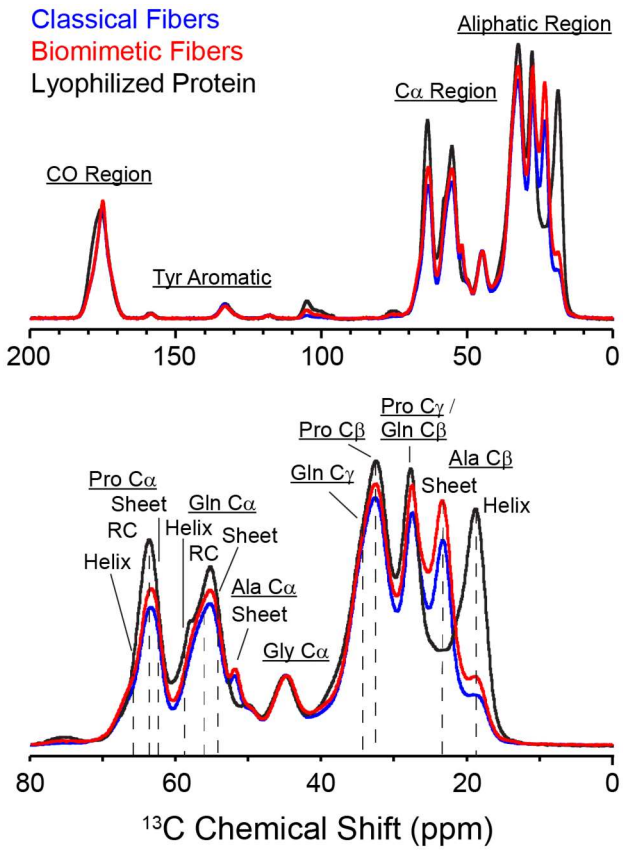



**Figure 59.** $^{1}H$-$^{13}C$ CP-MAS experiments on labeled silks. Each spectra is scaled to the Gly peak at approximately 45 ppm.

Previous studies by our collaborators have determined that fibers spun from BSD show superior mechanical properties over their CSD counterparts.[6] We wanted to follow this by understanding the molecular differences between the fiber types and attribute them to differences in mechanical properties.

Using solid-state NMR (SSNMR) we investigated the secondary structure differences between BSD fibers, CSD fibers, and the lyophilized protein before fiber spinning. Figure 59 shows the $^{1}H$-$^{13}C$ CP-MAS results for the $^{13}C/^{15}N$ enriched lyophilized protein, CSD and BSD fibers. Several differences can be seen between the lyophilized material and fiber types. In both

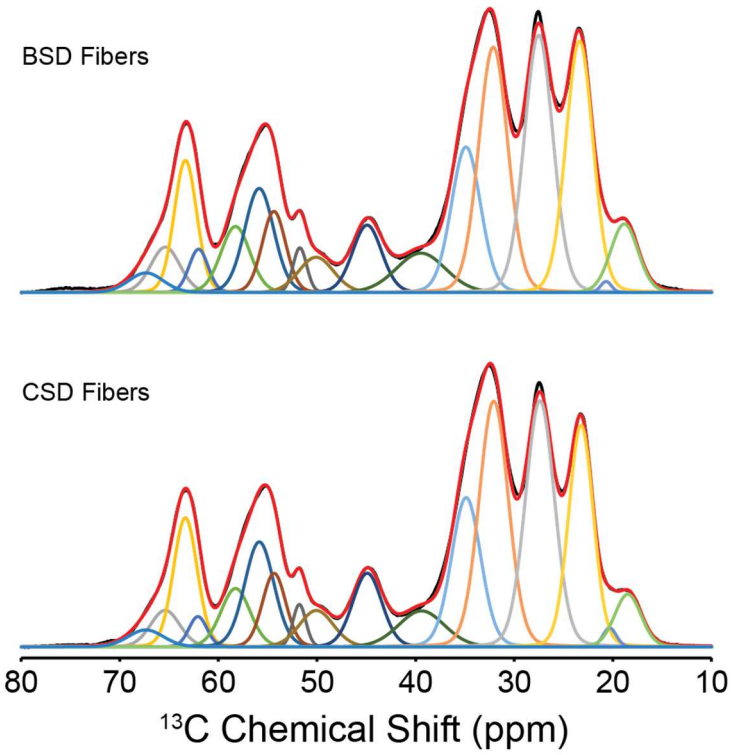



**Figure 60.** Deconvolutions of BSD (top) and CSD (bottom). Secondary structure assignments were done using chemical shifts from the DARR (**Figure 61**) and DARR difference (**Figure 63**). Once the positions were fixed, the linewidths and amplitudes were allowed to adjust freely using a 0.9 ratio for Gaussian/Lorentzian line shape. Linewidths greater than 3.5 ppm were fixed at 3.5 ppm. Assignment based on DARR is shown in Figure 61.

fiber samples, the Ala Cβ resonance near 23 ppm shows a considerably higher amount of β-sheet structure compared to the lyophilized powder, which is primarily RC near 18 ppm. Both materials, BSD and CSD, have the same amount of Ala Cβ in at 77%. This value is approximately 5% lower than what is found in native dragline silk.[29] The Gln Cα near 58 ppm also shows a moderate shift from a helical environment to RC / β-sheet between the fiber samples and powder, respectively. With s parse $^{13}C$ enrichment, we are also able to see several resonances we can attribute to Pro. Pro Cα shows a slight but still identifiable shift from α-helical to RC / β-sheet type structures. Interestingly and perhaps surprisingly, we do not see a substantial difference between the CSD and BSD samples even after comparing secondary structures through deconvolution (**Figure 60**).

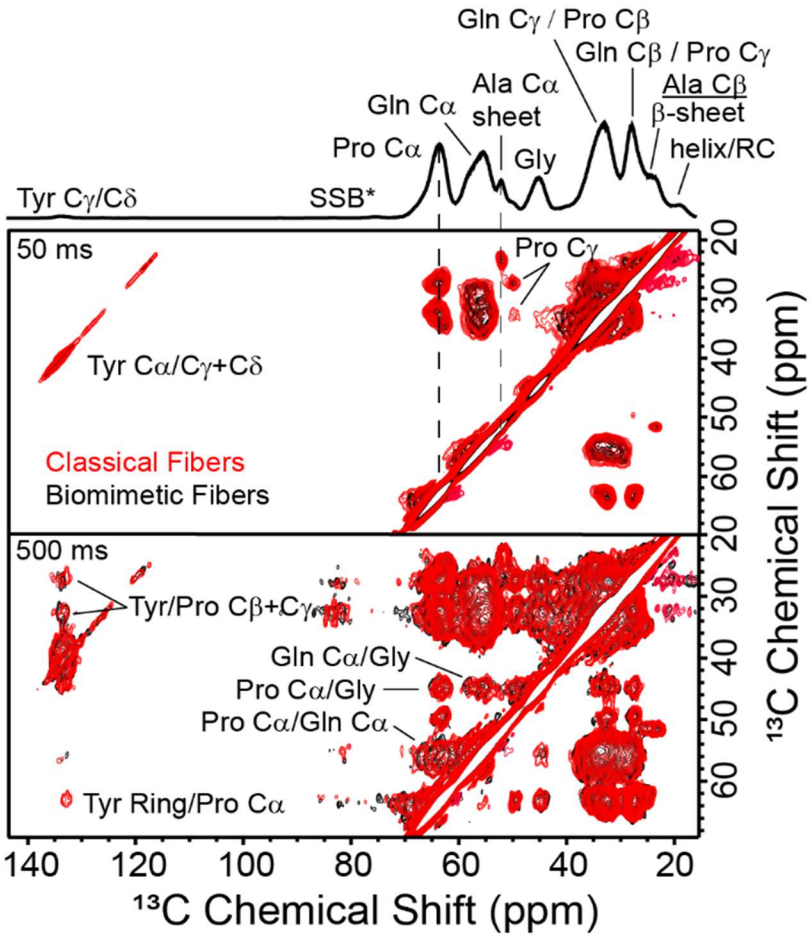



**Figure 61.** Through-space $^{13}C$-$^{13}C$ DARR correlation. At the two mixing times of 50 ms (top) and 500 ms (bottom) it is possible to accurately determine the chemical shifts for several amino acids in the BSD (black) and CSD (red) fibers. Overall, the chemical shifts and extracted linewidths look similar in both materials. The extracted linewidths were used as a starting place for deconvolutions (figure 59) and for integration (Table 7).

Table 7 lists the secondary structure assignments and the normalized areas of the BSD and CSD fibers referenced to DSS.[10, 30]

To learn what interactions may be present between strands, we conducted a series of though-space $^{13}C$-$^{13}C$ correlation spectra at both short (50 ms) and long (500 ms) mixing times which helps us to identify both intra- and inter-residue correlations. In the dipolar assisted rotational resonance experiment (DARR), signal is passed from one $^{13}C$ nuclei to others near in space over a mixing period. Increasing the mixing period allows us to see further distances through-space. Specifically, short mixing times (50 ms) allow us to determine the level of

**Table 7.** Secondary structure assignments of BSD and CSD fibers from DARR (Figure 61) as well as relative secondary structure after deconvoluting CP data (Figure 60). Only amino acids that could be confidently deconvoluted for each secondary structure are provided. Proline was found to be in the same structure as collagen/elastin as measured from the Cβ and Cγ chemical shift distance (Figure 62).

| | | | | Classical Fibers | | | Biomimetic Fibers | | |
|---|---|---|---|---|---|---|---|---|---|
| **Residue Site** | **Chemical Shift (ppm)*** | | | **Normalized Area (%)** | | | **Normalized Area (%)** | | |
| | α-helix | RC | β-sheet | α-helix | RC | β-sheet | α-helix | RC | β-sheet |
| Ala Cα | - | - | 51.46 | | | | | | |
| Ala Cβ | 18.43 | 20.4 | 22.82 | 19 | 4 | 77 | 21 | 2 | 77 |
| Ala CO | | | 175.1 | | | | | | |
| Pro Cα | 65.40 | 63.35 | 62.80 | 23 | 65 | 12 | 25 | 58 | 16 |
| Pro Cβ | 32.05 | 32.10 | 32.59 | | | | | | |
| Pro Cγ** | 27.65 | 27.39 | 27.14 | | | | | | |
| Pro Cδ** | - | 50.06 | - | | | | | | |
| Pro CO | | | 176.00 | | | | | | |
| Gln Cα | 58.26 | 55.85 | 54.35 | 26 | 47 | 27 | 28 | 44 | 28 |
| Gln Cβ | 33.18 | 33.25 | 33.11 | | | | | | |
| Gln Cγ** | 29.67 | 30.68 | 29.71 | | | | | | |
| Gln CO | | | 174.70 | | | | | | |
| Gly Cα | 44.8 | - | - | | | | | | |
| Gly CO | 173.01 | | | | | | | | |
| Tyr Cβ | - | - | 40.29 | | | | | | |
| Tyr Cδ** | - | 133.61 | - | | | | | | |

- Signal too low
* From DSS
** Unknown 2° assignments listed as RC

enrichment and accu rately assign chemical shifts within amino acids, whereas long mixing times (500 ms) answer questions about contacts between amino acids that are close in space.[29, 31]

At the short mixing time of 50 ms, we are able to see a number of sites labeled including Ala Cα/Cβ, Pro Cα/Cβ/Cγ, Gln Cα/Cβ/Cγ and Gly Cα in both BSD and CSD fibers (**Figure 61, 50 ms**). The signal intensity is similar in proportion to the predicted amino acids that would be labeled using 1-$^{13}$C-glucose.[8] The chemical shifts of these sites also appear similar in both CSD and BSD fibers. At longer mixing times, we can see inter-residue correlations between several residues

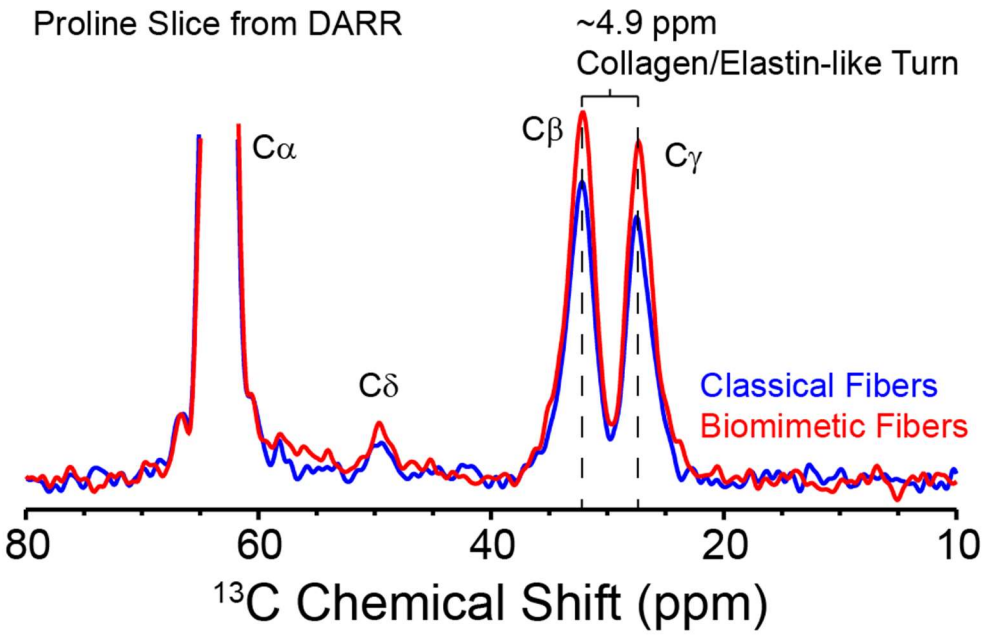



**Figure 62.** Slice extracted from DARR spectra at Pro Cα resonance (~63 ppm) for both CSD (blue) and BSD (red) fibers. By taking the difference in chemical shift between the Cβ and Cγ resonances, it is possible to determine the secondary structure of Pro. Our results indicate that Pro in both BSD and CSD fibers have a collagen/elastin-like secondary structure.

approximately 0.3-0.5 nm away.[31] This has helped us to identify several key amino acids and their correlations such as Gln Cα/Gly, Pro Cα/Gly Cα, Pro Cα/Gln Cα, and Tyr Cγ+δ/Pro Cβ+γ (**Figure 61, 500 ms**). Unfortunately, it was difficult to determine if inter-residue interactions (Tyr/Pro, Pro/Gln) were occurring between strands or within the same strand because these residues are neighbors in the primary sequence (**Figure 49**). With improved resolution from the 2D DARR, we were able to extract slices correlating the Pro Cα / β / γ sites and determined its secondary structure (**Figure 62**). By calculating the difference in chemical shift between the Pro Cβ and Cγ resonances, we were able to determine that Pro in both CSD and BSD fibers adopt a collagen/elastin-like turn structure.[32] This is similar in to previous reports of Pro in spider dragline silk.[33] We also see an additional correlation between the Tyr ring and Pro Cα in the CSD fibers, however this peak is on the threshold of noise and further characterization of this interaction is most likely due to intra-strand correlations (see below).

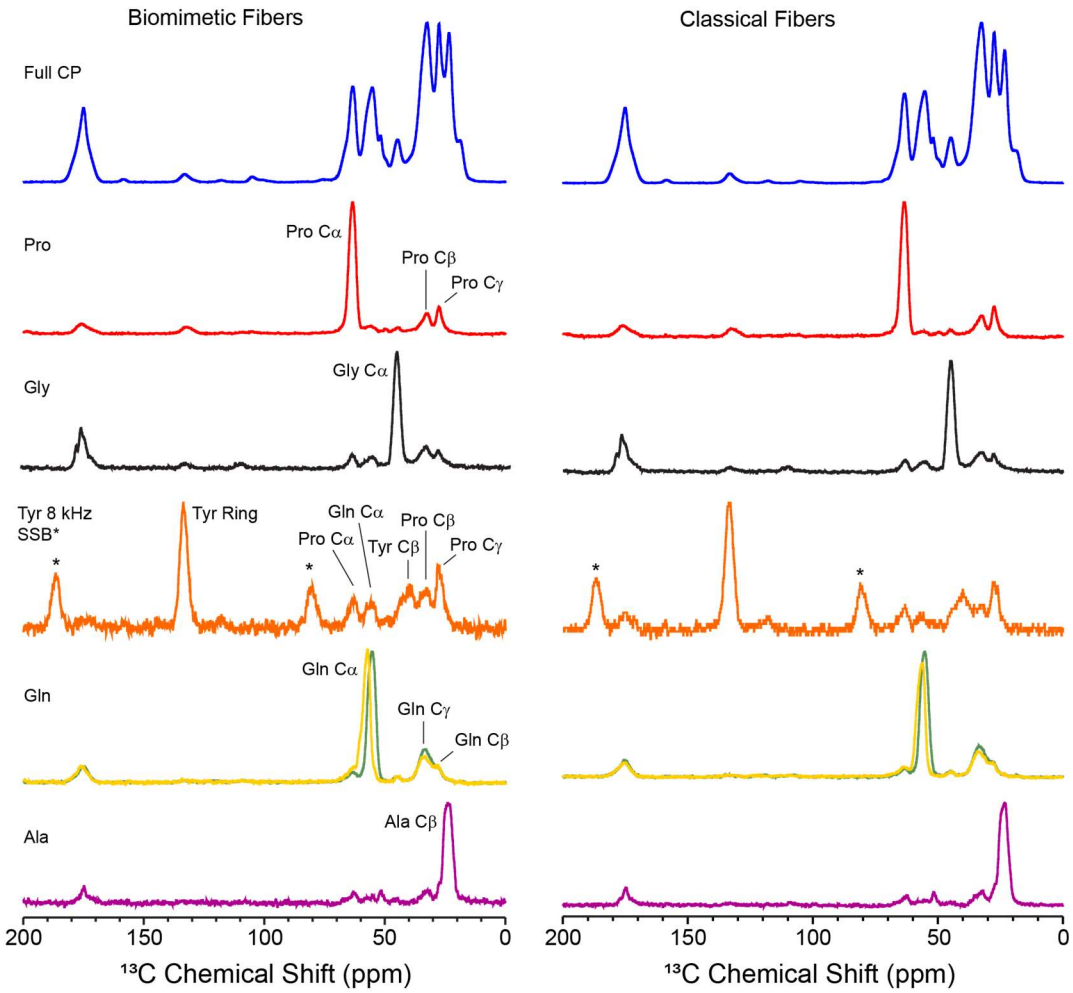



**Figure 63.** DARR difference experimental results for isotopically labeled BSD and CSD fibers. The CP spectra for each sample is at the top (blue), followed by each 1D experiment where the largest peak was selected for spin-diffusion. Two experiments were collected for the Gln Cα resonance since the chemical shift range is broad (~5 ppm). The asterisks note spinning side bands.

To more accurately reveal neighboring amino acids in the silk fibers, we utilize an experiment developed in our group called DARR difference.[13] In this experiment, we first select $^{13}C$ peaks in our 1D spectrum and subsequently allow spin-diffusion to occur to nearby nuclei. The result is a 1D spectra where each signal corresponds to only nuclei close in space allowing for selection of just a single a mino acid site in a typically crowded spectra if one of the amino acid sites is resolvable in the spectrum. By changing the mixing time, it is possible to probe further

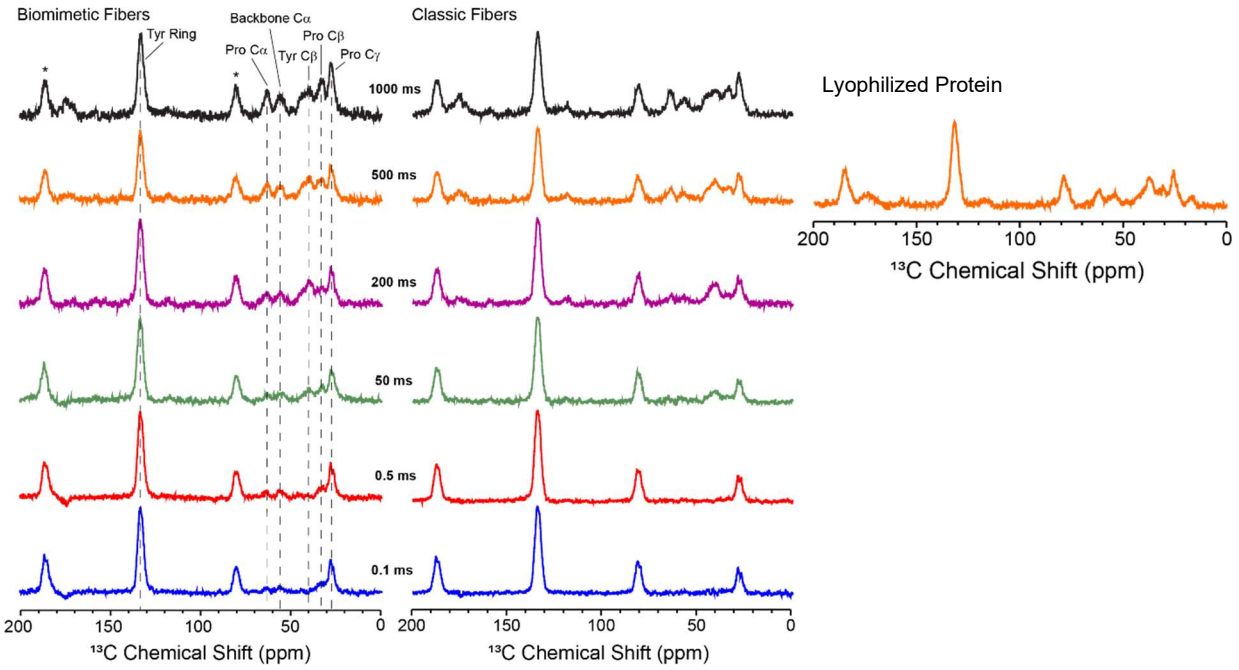



**Figure 64.** Through-space $^{13}C$-$^{13}C$ selective correlation experiment designed at SDSU called DARR difference. In this experiment, the Tyr sidechain at ~131 ppm is selected by the pulse sequence and spin-diffusion occurs within an amino acid and between adjacent amino acids. By plotting the signal versus mixing time (from 0.1-1000 ms) we can measure distances using Tyr as an internal ruler. From this data, we see that signal builds in immediately from Pro Cγ (blue, 0.1 ms), leading us to conclude that Tyr ring – Pro Cγ are nearby in space. However, this is also seen in the lyophilized power (right, orange, 500 ms) suggesting that this interaction is present in the starting material. The buildup rates for both BSD and CSD fibers look similar at all sites.

distances from the peak of interest. This experiment is similar in principle to a 2D DARR, with added selectivity for resonances of interest analogous to NOE in a one-dimensional fashion.

By targeting a $^{13}C$ site that is resolvable for each amino acid found in the primary sequence, we are able to assign all the chemical shifts for several nuclei within amino acids in our structure that would normally not be resolved due to spectral overlap. Figure 63 shows the results of this experiment for several nuclei on most amino acids found in the primary sequence of the sparsely labeled $(AQ)_{12}NR3$ protein. Not only do we see nuclei corresponding to intra-residue contacts, but several inter-residue assignments can be made. For example, when we select the Tyr ring (**Figure 63, orange**) we see that several amino acids are close in space and not only those corresponding to Tyr. However, further investigation has revealed that these

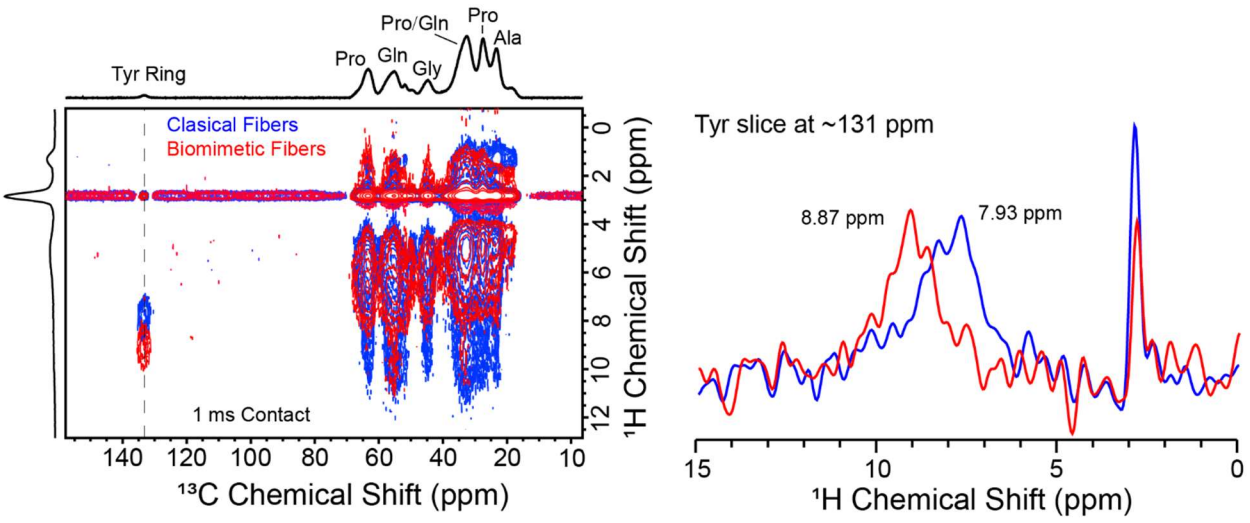



**Figure 65.** $^{1}$H-$^{13}$C heteronuclear correlation (HETCOR) experiment. This experiment correlates $^{1}$H and $^{13}$C signals through-space, where the distance is determined by the contact time (1 ms here). Slices were extracted at the Tyr sidechain resonance (right) to illustrate the differences between the materials. Tyr $^{1}$H signals in the BSD fibers are more downfield and therefore suggest they are in a more de-shielded (polar) environment. This downfield $^{1}$H shift in the BSD fibers is also seen in the C$\alpha$ region, but not in the aliphatic region.

interactions are also p resent in the lyophilized powder, suggesting that these are likely due to intra-strand contacts along the protein backbone where next nearest neighbor amino acids are observed (**Figure 64**). Overall, we were unable to determine if the signals arising near several residues were from neighboring strands. The difficulty is due to the repetitiveness and location of the amino acids in the primary sequence. In the future, labeling strategies that only label amino acids in high fidelity may be useful here. In this way, we could determine inter-strand distances between amino acids between the different materials.

We further investigated the difference in secondary structure using a heteronuclear correlation experiment ($^{1}$H-$^{13}$C HETCOR) to correlate $^{1}$H nuclei to $^{13}$C nuclei close in space. Because of the vast dipole-coupling network of $^{1}$H in proteins, high spinning speeds are required to average out these effects albeit still produce large streaks as well as artifacts due to the LG sequence (**Figure 65**). From our HETCOR experiment, we see two major difference between the BSD and CSD materials. First, the region between approximately 50-65 ppm corresponding to

the Ca backbone of the protein appears shifted downfield in the $^{1}H$ dimension for BSD (red) from the CSD material (blue). The same is true for the Tyr ring region at 131 ppm $^{13}C$. In fact, the shift downfield for the protons on the Tyr ring are more dramatic than those seen in the C$\alpha$ backbone region. This suggests that the Tyr ring as well as the C$\alpha$ resonances in BSD are located in a more polar environment shifting their $^{1}H$ resonances downfield. Tyr is of interest to us because previous molecular modeling has shown that Tyr residues in native silk proteins are organized to the periphery of the collapsed protein, interacting with water. We believe this organization might help to drive liquid-liquid phase separation through pi-pi interactions. It is not immediately clear what is causing this shift and how this is correlated to toughness, but we do know that the fibers made via BSD show greater toughness than with CSD.[6] It might be possible more water trapped within the BSD fibers which could produce this effect. The effect of water on the mechanical properties of silk fibers has been studied previously, where the addition of water causes supercontraction in addition to a marketable differences in extensibility.[34, 35] These differences have also been characterized with NMR, where there is a clear difference in relaxation time between dry and wetted silks.[36] In the recombinant system here, it is possible that the post-stretching and subsequent drying methods could play a dramatic role in the fiber quality and would account for the differences in toughness between BSD and CSD.

**Conclusions**

The results presented here lay the foundation for a holistic picture for silk fiber formation from lyophilized recombinant protein to final fiber. By generating silk fibers spun from native-like conditions we can begin to understand what is necessary to produce synthetic silk fibers with material properties that rival native systems. In this work, we compared recombinant silk dopes prepared in similar conditions to those found in native spider silk glands containing phosphate. Light microscopy and DLS reveal that these recombinant silk proteins transition from a low to high density phase through two different processes depending on buffer conditions. Protein dopes

containing phosphate buffer transition from monomer to pre-assemblies gradually over time and ultimately achieve liquid-liquid phase separation in the final highly-concentrated dope. Dopes that do not contain phosphate transition from monomer to pre-assemblies but stop short of forming large liquid-liquid droplets on the micron scale. Liquid NMR experiments reveal that the backbone amide exchange for phosphate-containing dopes is much slower than dopes that do not contain phosphate which has a direct impact on the ability to form hydrogen bonds necessary for fiber formation. The secondary structure of the protein dopes with and without phosphate generally appear disordered with a slight disposition of the leading Ser-Ala amino acids in the poly(Ala) motif of the phosphate-containing dopes to be $\alpha$-helical. In the final fibers, the secondary protein structure between fibers spun from dope with or without phosphate appeared the same. The only difference we were able to identity between the two different fibers was the downfield shift in $^{1}H$ signal in the HETCOR experiment for several resonances including the Tyr sidechain in the fiber spun from phosphate-containing dope. This suggests that Tyr may play an important role in achieving greater toughness in fibers which we plan to investigate in the future.

Although our results are generally inconclusive to what major differences there are between these two fibers and how that impacts mechanical properties, there are a few outstanding experiments we wish to explore in our future work. For example, the alignment of the crystalline $\beta$-sheet motifs formed by the poly(Ala) motifs has not been explored for this system. Alignment of these motifs help to impart greater stiffness to the material which is introduced with post-stretching the fibers immediately after extrusion. Another example is the measurement of domain sizes within the fiber. That is, measuring the volume of both the crystalline and amorphous domains within the fiber. Although we have reported the amount of Ala in a $\beta$-sheet, the size of these crystalline regions is unknown. The size of the $\beta$-sheet crystallites has a direct impact on the quality of the fiber. If the crystallites are too short, the toughness will be attenuated. If the crystallites are too large, the material will be too brittle.

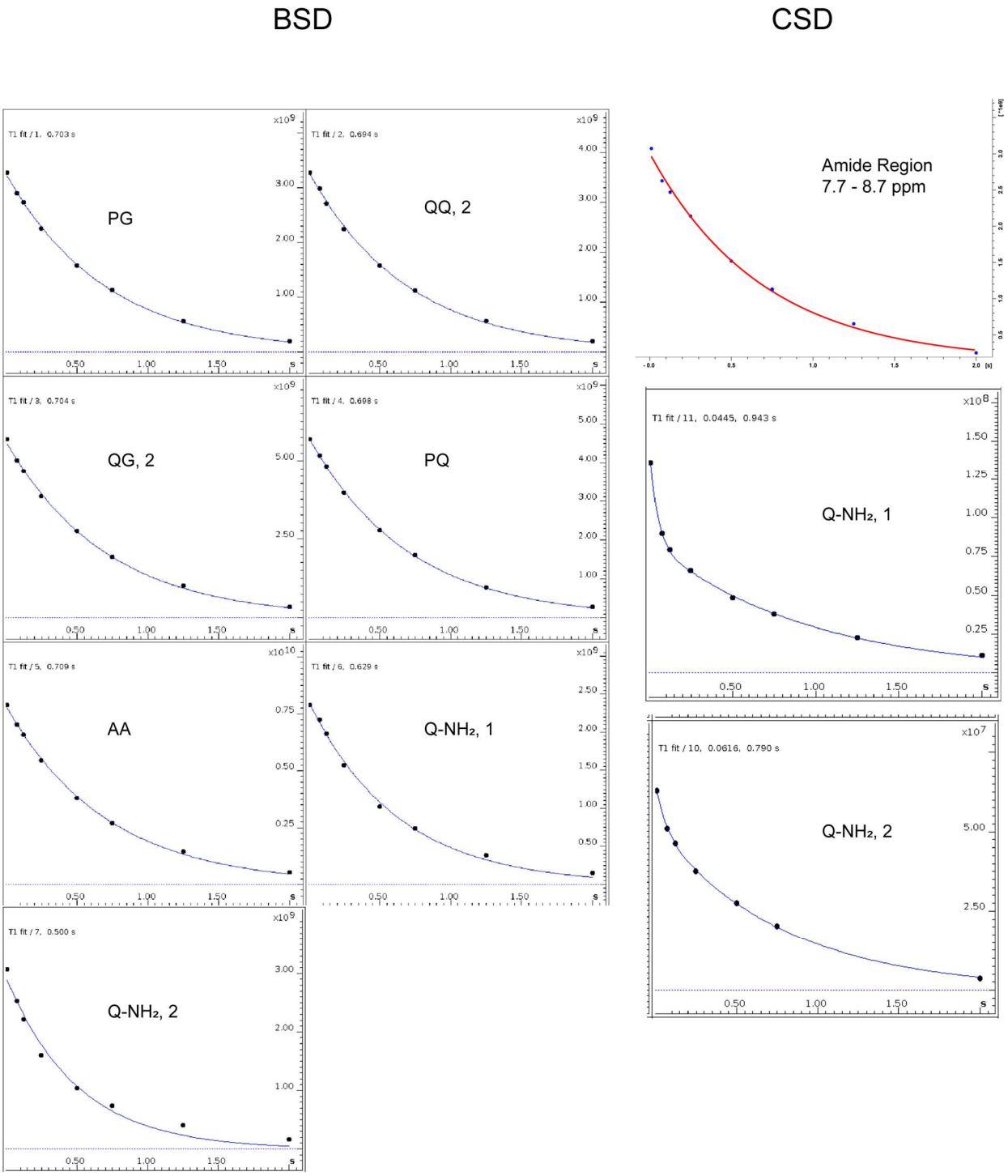



**Figure 66.** $T_1$ fits for BSD and CSD. The CSD material does not have resolved amide peaks, so the entire amide region was fit for $T_1$ measurements between 7.7-8.7 ppm. Numbers correspond to dipeptide repeats in Table 8.

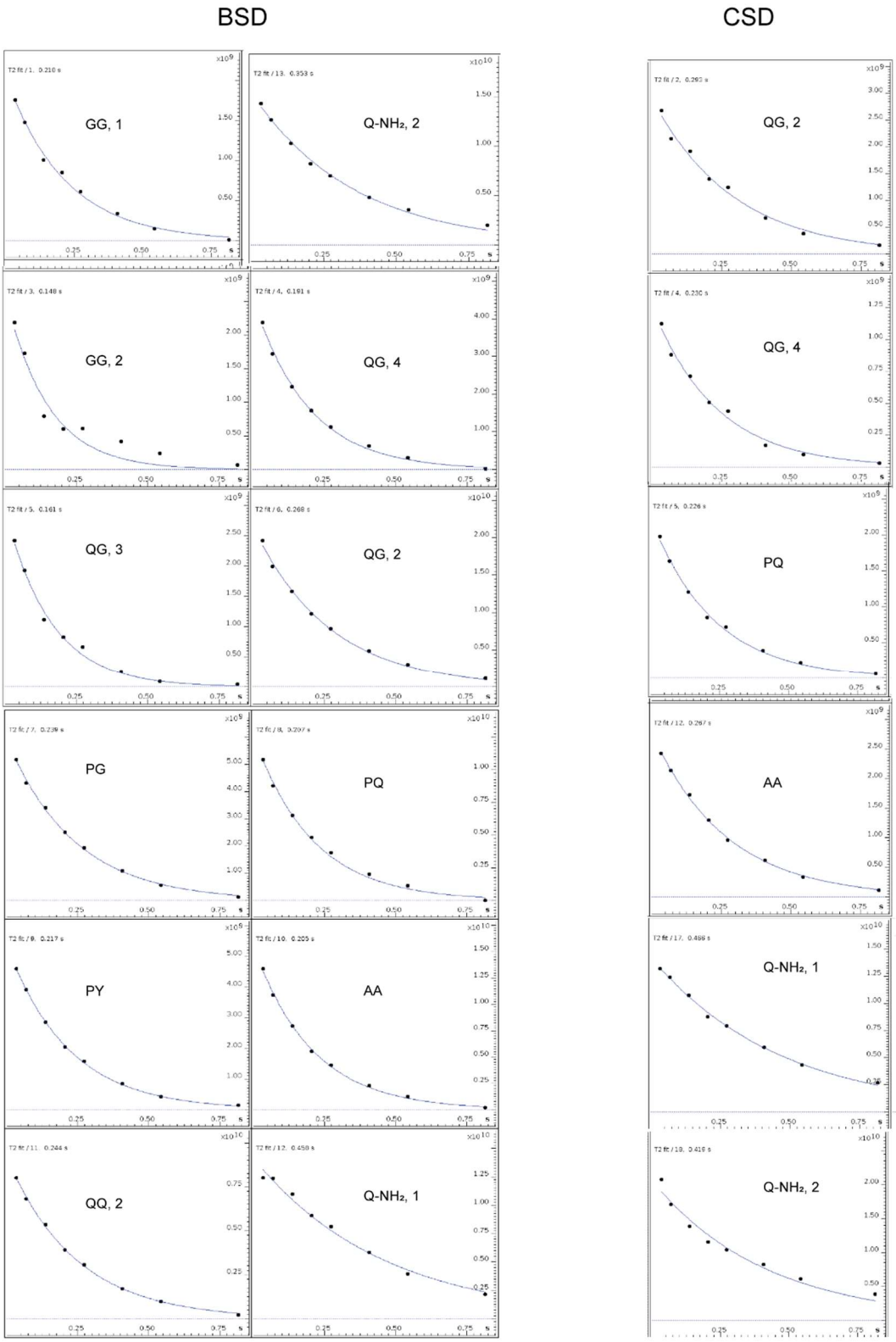



**Figure 67.** $T_2$ fits for BSD and CSD. Some of the BSD and several of the CSD peaks did not have enough points to be fit. Numbers correspond to dipeptide repeats in Table 8.

**Table 8.** Chemical shift values and reported structure for residues in $(AQ)_{12}$NR3 prepared as CSD or BSD. Residue assignments were done with 3D datasets and are reported on figure 1. Assignments with a similar i residue (bolded) are grouped, and similar sequences are numbered. None of the CSD $^{1}$H-$^{15}$N HSQC peaks could be accurately determined with the 3D datasets. We therefore assumed the overlapping chemical shifts between CSD and BSD were the same residue.

| $^{15}$N | BSD | CSD | α-helix | RC | β-sheet | $^{1}$H | BSD | CSD | α-helix | RC | β-sheet |
|---|---|---|---|---|---|---|---|---|---|---|---|
| G**G**, 1 | 107.08 | 107.20 | 107.34 | 109.94 | 110.19 | | 8.22 | 8.22 | 8.23 | 8.34 | 8.27 |
| G**G**, 2 | 107.76 | 107.89 | | | | | 8.11 | 8.11 | | | |
| P**G** | 108.92 | 108.91 | | | | | 8.54 | 8.54 | | | |
| Q**G**, 1 | 110.25 | | | | | | 8.53 | | | | |
| Q**G**, 2 | 109.84 | 109.82 | | | | | 8.32 | 8.31 | | | |
| Q**G**, 3 | 109.69 | 109.64 | | | | | 8.15 | 8.16 | | | |
| Q**G**, 4 | 109.67 | 109.67 | | | | | 8.05 | 8.05 | | | |
| **S**, 1 | 114.41 | | 114.78 | 115.94 | 117.44 | | 8.29 | | 8.11 | 8.26 | 8.57 |
| **S**, 2 | 115.22 | | | | | | 8.21 | | | | |
| Q**Q**, 1 | 119.58 | | 118.59 | 119.73 | 123.14 | | 8.45 | | 8.11 | 8.25 | 8.51 |
| Q**Q**, 2 | 120.90 | 120.84 | | | | | 8.47 | 8.46 | | | |
| P**Q** | 119.21 | 118.89 | | | | | 8.19 | 8.18 | | | |
| P**Y** | 118.99 | 119.00 | 119.67 | 120.05 | 122.55 | | 8.09 | 8.08 | 8.10 | 7.90 | 8.69 |
| A**A** | 122.12 | 122.08 | 121.65 | 123.52 | 125.57 | | 8.08 | 8.08 | 7.99 | 8.11 | 8.59 |
| **Q-N**$H_2$, 1 | 111.93 | 111.47 | | | | | 7.52 | 7.51 | | | |
| **Q-N**$H_2$, 2 | 111.93 | 111.47 | | | | | 6.86 | 6.86 | | | |

## References


1. Blackledge, T. A.; Swindeman, J. E.; Hayashi, C. Y., Quasistatic and continuous dynamic characterization of the mechanical properties of silk from the cobweb of the black widow spider Latrodectus hesperus. *J Exp Biol* **2005,** *208* (Pt 10), 1937-49.

2. Gosline, J.; Lillie, M.; Carrington, E.; Guerette, P.; Ortlepp, C.; Savage, K., Elastic proteins: biological roles and mechanical properties. *Philos Trans R Soc Lond B Biol Sci* **2002,** *357* (1418), 121-32.

3. Swanson, B. O.; Blackledge, T. A.; Beltrán, J.; Hayashi, C. Y., Variation in the material properties of spider dragline silk across species. *Applied Physics A* **2005,** *82* (2), 213-218.

4. Vollrath, F.; Knight, D. P., Liquid crystalline spinning of spider silk. *Nature* **2001,** *410* (6828), 541-8.

5. Johansson, J.; Rising, A., Doing What Spiders Cannot-A Road Map to Supreme Artificial Silk Fibers. *ACS Nano* **2021,** *15* (2), 1952-1959.

6. Heidebrecht, A.; Eisoldt, L.; Diehl, J.; Schmidt, A.; Geffers, M.; Lang, G.; Scheibel, T., Biomimetic fibers made of recombinant spidroins with the same toughness as natural spider silk. *Adv Mater* **2015,** *27* (13), 2189-94.

7. Huemmerich, D.; Helsen, C. W.; Quedzuweit, S.; Oschmann, J.; Rudolph, R.; Scheibel, T., Primary Structure Elements of Spider Dragline Silks and Their Contribution to Protein Solubility. *Biochemistry* **2004,** *43*, 13604-13612.

8. Lundstrom, P.; Teilum, K.; Carstensen, T.; Bezsonova, I.; Wiesner, S.; Hansen, D. F.; Religa, T. L.; Akke, M.; Kay, L. E., Fractional 13C enrichment of isolated carbons using [1-13C]- or [2- 13C]-glucose facilitates the accurate measurement of dynamics at backbone Calpha and side-chain methyl positions in proteins. *J Biomol NMR* **2007,** *38* (3), 199-212.

9. Meissner, A.; Sorensen, O. W., Sequential HNCACB and CBCANH protein NMR pulse sequences. *J Magn Reson* **2001,** *151* (2), 328-31.

10. Wang, Y.; Jardetzky, O., Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein Sci* **2002,** *11* (4), 852-61.

11. Takegoshi, K.; Nakamura, S.; Terao, T., $^{13}C$-$^{1}H$ dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chemical Physics Letters* **2001,** *344*, 631-637.

12. Takegoshi, K.; Nakamura, S.; Terao, T., 13C–1H dipolar-driven 13C–13C recoupling without 13C rf irradiation in nuclear magnetic resonance of rotating solids. *The Journal of Chemical Physics* **2003,** *118* (5), 2325-2341.

13. Addison, B.; Stengel, D.; Bharadwaj, V. S.; Happs, R. M.; Doeppke, C.; Wang, T.; Bomble, Y. J.; Holland, G. P.; Harman-Ware, A. E., Selective One-Dimensional (13)C-(13)C Spin-Diffusion Solid-State Nuclear Magnetic Resonance Methods to Probe Spatial Arrangements in Biopolymers Including Plant Cell Walls, Peptides, and Spider Silk. *J Phys Chem B* **2020,** *124* (44), 9870-9883.

14. Parent, L. R.; Onofrei, D.; Xu, D.; Stengel, D.; Roehling, J. D.; Addison, J. B.; Forman, C.; Amin, S. A.; Cherry, B. R.; Yarger, J. L.; Gianneschi, N. C.; Holland, G. P., Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks. *Proc Natl Acad Sci U S A* **2018,** *115* (45), 11507-11512.

15. Lin, T. Y.; Masunaga, H.; Sato, R.; Malay, A. D.; Toyooka, K.; Hikima, T.; Numata, K., Liquid Crystalline Granules Align in a Hierarchical Structure To Produce Spider Dragline Microfibrils. *Biomacromolecules* **2017,** *18* (4), 1350-1355.

16. Malay, A. D.; Suzuki, T.; Katashima, T.; Kono, N.; Arakawa, K.; Numata, K., Spider silk self-assembly via modular liquid-liquid phase separation and nanofibrillation. *Science Advances* **2020,** *6*, 1-12.

17. Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J., Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry* **1999,** (38), 16424-16431.

18. Alexandra H. Simmons, C. A. M., Lynn W. Jelinski, Molecular Orientation and Two-Component Nature of the Crystalline Fraction of Spider Dragline Silk. *Science* **1996,** *271* (5245), 84-87.

19. Xu, D.; Yarger, J. L.; Holland, G. P., Exploring the backbone dynamics of native spider silk proteins in Black Widow silk glands with solution-state NMR spectroscopy. *Polymer* **2014,** *55* (16), 3879-85.

20. Onofrei, D.; Stengel, D.; Jia, D.; Johnson, H. R.; Trescott, S.; Soni, A.; Addison, B.; Muthukumar, M.; Holland, G. P., Investigating the Atomic and Mesoscale Interactions that Facilitate Spider Silk Protein Pre-Assembly. *Biomacromolecules* **2021**.

21. Yang, Z.; Grubb, D. T.; Jelinski, L. W., Small-Angle X-ray Scattering of Spider Dragline Silk. *Macromolecules* **1997,** *30*, 8254-61.

22. Addison, B.; Onofrei, D.; Stengel, D.; Blass, B.; Brenneman, B.; Ayon, J.; Holland, G. P., Spider prey-wrapping silk is an alpha-helical coiled-coil/beta-sheet hybrid nanofiber. *Chem Commun (Camb)* **2018,** *54* (76), 10746-10749.

23. Addison, J. B.; Ashton, N. N.; Weber, W. S.; Stewart, R. J.; Holland, G. P.; Yarger, J. L., b-Sheet Nanocrystalline Domains Formed From Phosphorylated Serine-rich Motifs in Caddisfly Larval Silk: A Solid State NMR and XRD Study. *Biomacromolecules* **2013,** *14* (4), 1140-8.

24. Suzuki, Y.; Yamazaki, T.; Aoki, A.; Shindo, H.; Asakura, T., NMR study of the structures of repeated sequences, GAGXGA (X = S, Y, V), in Bombyx mori liquid silk. *Biomacromolecules* **2014,** *15* (1), 104-12.

25. Stengel, D.; Addison, J. B.; Onofrei, D.; Huynh, N. U.; Youssef, G.; Holland, G. P., Hydration-Induced β-Sheet Crosslinking of α-Helical-Rich Spider Prey-Wrapping Silk. *Advanced Functional Materials* **2021**.

26. Kay, L. E.; Torchia, D. A.; Bax, A., Backbone Dynamics of Proteins As Studied by $^{15}N$ Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **1989,** (28), 8972-8979.

27. Lewis, R., Spider Silk: Ancient Ideas for New Biomaterials. *Chem. Rev.* **2006,** *106* (9), 3762−3774.

28. Eisoldt, L.; Hardy, J. G.; Heim, M.; Scheibel, T. R., The role of salt and shear on the storage and assembly of spider silk proteins. *J Struct Biol* **2010,** *170* (2), 413-9.

29. Holland, G. P.; Jenkins, J. E.; Creager, M. S.; Lewis, R. V.; Yarger, J. L., Quantifying the fraction of glycine and alanine in beta-sheet and helical conformations in spider dragline silk using solid-state NMR. *Chem Commun (Camb)* **2008,** (43), 5568-70.

30. Morcombe, C. R.; Zilm, K. W., Chemical shift referencing in MAS solid state NMR. *Journal of Magnetic Resonance* **2003,** *162* (2), 479-486.

31. Hong, M.; Schmidt-Rohr, K., Magic-Angle-Spinning NMR Techniques for Measuring Long-Range Distances in Biological Macromolecules. *Acc Chem Res* **2013,** *46*, 2154-2163.

32. Kricheldorf, H. R.; Müller, D., Secondary Structure of Peptides. 3. $^{13}C$ NMR Cross Polarization / Magic Angle Spinning Spectroscopic Characterization of Solid Polypeptides. *Macromolecules* **1983,** *16* (4).

33. Jenkins, J. E.; Creager, M. S.; Butler, E. B.; Lewis, R. V.; Yarger, J. L.; Holland, G. P., Solid-state NMR evidence for elastin-like beta-turn structure in spider dragline silk. *Chem Commun (Camb)* **2010,** *46* (36), 6714-6.

34. Liu, Y.; Shao, Z.; Vollrath, F., Relationships between supercontraction and mechanical properties of spider silk. *Nat Mater* **2005,** *4* (12), 901-5.

35. Blackledge, T. A.; Boutry, C.; Wong, S. C.; Baji, A.; Dhinojwala, A.; Sahni, V.; Agnarsson, I., How super is supercontraction? Persistent versus cyclic responses to humidity in spider dragline silk. *J Exp Biol* **2009,** *212* (Pt 13), 1981-9.

36. Zhitong Yang; Oskar Liivak; Andreas Seidel; George LaVerde; David B. Zax; Jelinski, L. W., Supercontraction and Backbone Dynamics in Spider Silk: $^{13}C$ and $^{2}H$ NMR Studies. *Journal of the American Chemical Society* **2000,** *122*.

**Chapter 6.** Spider Prey-Wrapping Silk is an α-Helical Coiled-Coil/β-Sheet Hybrid Nanofiber

## Abstract

Solid-State NMR results on $^{13}$C-Ala/Ser and $^{13}$C-Val enriched Argiope argentata prey-wrapping silk show that native, freshly spun aciniform silk nanofibers are dominated by α-helical (~50% total) and random-coil (~35% total) secondary structures, with minor β-sheet nanocrystalline domains (~15% total). This is the most in-depth study to date characterizing the protein structural conformation of the toughest natural biopolymer: aciniform prey-wrapping silks.

## Introduction

Spiders have evolved to produce up to 7 different types of silk with a wide range of impressive mechanical properties.[1,2] To date, the vast majority of structural data on as-spun silk fibers has been on dragline, or Major Ampulate (MA), silks due to their unmatched strength and ease of study. However, through a unique combination of high strength (~700 MPa) and high

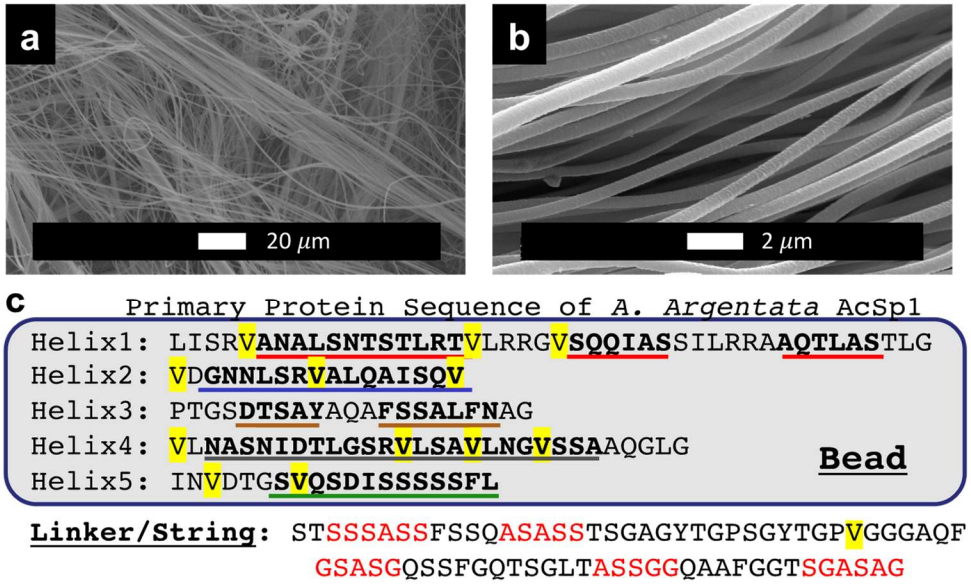



**Fig 68.** Scanning Electron Microscopy (SEM) images (a, b) of native *A. argentata* prey-wrapping silk bundles. Fibers are roughly 500 nm in diameter. (c) Primary-protein sequence for *A. argentata* AcSp1 repeat unit. Regions determined to be helical in solution are bolded and color-underlined based on structure homology between *A. trifaciata* (PDB Code 2MU3) and *A. argentata*. Val residues are highlighted in yellow and proposed β-sheet-forming motifs from the linker region are shown with red lettering.

extensibility before breaking (~60-80%),3 lesser-studied aciniform silk, utilized for prey-wrapping and egg-case lining, is actually the toughest of the spider silks and boasts mechanical properties that surpass the toughest man-made materials.[1,4] A complete molecular-level understanding of native aciniform silk fibroin is lacking, although significant insight into its protein structure can be gained from the current consensus model for spider dragline fibers and from solution-NMR work conducted on recombinant aciniform spidrion 1 (AcSp1) protein. The high strength and moderate extensibility of dragline fibers is largely attributed to common structural motifs arising from short repetitive protein units; high fiber strength is thought to arise from aligned nanocrystalline β-sheet structures comprised of poly(Ala) and poly(Gly-Ala), while Gly-Gly-X and Gly-Pro-Gly-X-X repeats contribute to fiber elasticity in the form of randomly-oriented domains and elastin-like β-turns.[1,2,5] On the other hand, the consensus sequence of AcSp1, identified by Hayeshi et al., is composed of a string of ~14 much longer repeats of about 200 amino acids, flanked by non-repetitive C- and N-terminal regions.[3,6] The AcSp1 sequence contains both Ser-rich and Ala-Ser rich motifs (**Figure 68C**, red lettering), but aciniform silks are entirely deficient in the traditional poly(Ala), poly(Gly-Ala), Gly-Gly-X and Gly-Pro-Gly-X-X repeats.§ It is assumed through crude sequence-based structure predictions that AcSp1 repeats are likely to be rich in α-helices, and in 2011 this assumption was proven correct when Xu et. al solved the liquids-NMR structure of a single 200-AA repeat unit for recombinant aciniform spidroin 1 from *Argiope trifaciata* spiders.[7] The authors found that in solution the AcSp1 recombinant protein construct indeed is helix-rich, existing as a multi-domain “beads on a string” structure composed of a well-defined 5-helix globular domain (bead, helices 1-5) and a disordered linker domain (string). A very similar structure was solved by the same group in 2012 for the *Nephila antipodiana* AcSp1 repeat unit, highlighting a high degree of similarity across spider species.[8] These structures were improved in subsequent studies where researchers showed through solution-NMR structural and dynamical data that the serine-rich terminal helix of the globular bundle, denoted helix-5 (**Figure 68a**, **green**), is more dynamic, and thus possibly primed to form β-sheet nanostructures upon fiber aggregation.[9,10] While extremely

enlightening, these studies were performed on recombinant silk protein in solution, thus probing the secondary protein structures of native as-spun aciniform silk fibers is largely lacking. Polarized Raman studies do suggest that the silk protein dope in the gland is α-helix rich, while both helical and β-sheet structures are observed after fibrillization,[11,12] however these Raman data are not capable of directly obtaining amino-acid specific structural data nor correlating data to the primary protein sequence.

**Methods**

To gain a more detailed view into the molecular structure of native prey-wrap silks we utilized solid-state NMR techniques. *Argiope argentata* (this text) and *Argiope aurantia* spiders were fed ~50 microliters of a saturated solution containing A) $^{13}$C-labelled alanine (serine is also labelled through metabolism of alanine), and/or B) $^{13}$Clabelled valine every few days. Prey-wrap silk was collected by simulating prey using vibrating tweezers or a vibrating electric toothbrush. These labelling schemes were chosen to highlight the two distinct domains found in the aciniform repeat sequence (**Figure 68**); namely that alanine and serine are found dispersed throughout both the globular helical domain and the disordered linker region, while valine (and other hydrophobic residues) are found almost exclusively in the helical globular region (bead); notably in helices 1, 2, 4 and 5 but not in helix 3 (**Figure 72**).

**Results**

Alanine: $^{1}$H-$^{13}$C CP-MAS NMR data on $^{13}$C Ala-Ser enriched aciniform silk reveals a dominant Ala Cβ resonance at 16.3 ppm with a minor α-sheet shoulder at 21.0 ppm. The observed Ala Cβ chemical shift is suggestive of helical structure, but differentiating between RC and α-helical structures is more easily visualized by Cα chemical shifts; one expects a 2-3 ppm downfield shift for helical Cα resonances relative to RC.[13,14] Spectral overlap in the 1D data inhibits accurate

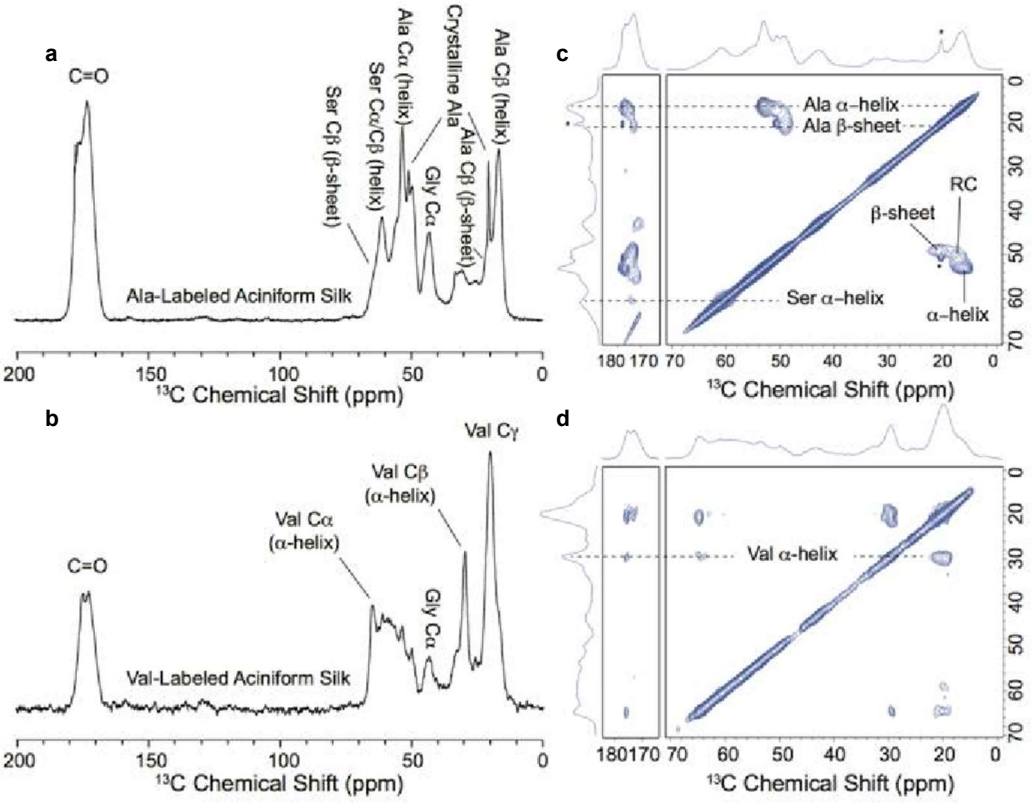



**Figure 69.** $^{1}$H-$^{13}$C CP-MAS spectra (a, b, 30 kHz MAS) and 2D $^{13}$C-$^{13}$C DARR (c, d, 14 kHz MAS, 100 ms DARR mixing time) spectra of freshly-spun *Argiope argentata* aciniform silks that were isotopically enriched with either $^{13}$C-Ala (a, c) or $^{13}$C-Val (b, d). Signal from crystalline $^{13}$C-Ala contaminant is indicated with an asterisks.

extraction of chemical shifts, thus we collected two-dimensional (2D) $^{13}$C-$^{13}$C through-space correlation experiments using dipolar-assisted rotational resonance (DARR) recoupling (**Figure 69**).[15,16] From the 2D DARR data we clearly see distinct Ala Cα / Cβ / CO dipolar-coupling cross-peaks. The dominant Cα / Cβ / CO cross-peaks (16.3 / 53.3 / 176.8 ppm) align with alanine adopting α-helical secondary structure. There are additional minor cross peaks at (20.9 / 49.3 / 172.8 ppm) and (17.5 / 50.5 ppm) representing alanine adopting β-sheet and random-coil structures, respectively. While one can visually interpret that the dominant Ala secondary structure is α-helical, we were interested in a more quantitative interpretation. Using precise chemical shift information extracted from 2D DARR data we could deconvolute the 1D CP-MAS data to estimate the percent representations of Alanine adopting helical, β-sheet and RC secondary structures.

For alanine, the Ala Cα resonance is the most well resolved and therefore is most likely to give confident fitting results. We find through deconvolution of the Ala Cα resonance that approximately 39% of all Ala residues adopt α-helical structures, with roughly 29% residing in β-sheet domains and the remaining 32% is unstructured (**Figure 70**).

Serine: Similar to alanine, serine amino acids are found distributed throughout both the globular (bead) domain and the disordered linker in the primary protein sequence of the AcSp1 repeat unit. Unsurprisingly our NMR data tells a similar story to alanine, namely that $^{13}C$ chemical shifts of the dominant serine resonance in native aciniform silks align with a true α-helical secondary structure, with minor β-sheet and RC subdomains present. We note that in the 2D DARR data, at the contour level displayed in **Figure 42c** we do not see any Ser Cα / Cβ cross-peaks for β-sheet environments. However, if we increase the contour level to just below the noise threshold those correlations are observed. Hence, it is clear from a simple visual interpretation of the 2D DARR spectra that serine β-sheet content is minimal compared to α-helical. To gain a more quantitative picture we peak-fit the seryl resonances. Serine is more difficult to peak fit than alanine due to substantial spectral overlap; serine Cα and Cβ resonances in an α-helix, β-sheet, and RC conformations all reside between 55 - 65 ppm. However, since the chemical shifts for Serine β-sheet structures could be extracted from $^{13}C$-$^{13}C$ DARR data, we could at minimum estimate β-sheet versus nonβ-sheet structures by deconvoluting the Ser Cβ resonance, which is the most downfield and best resolved and therefore most likely to yield accurate results. It is clear that for native prey-wrapping silk serine is mostly α-helical (~55%) with minor β-sheet (~30%) and RC (~15%) content. More confidently, we can state that roughly 30% of Ser residues exist in β-sheet structures, while the remaining 70% is either helical (dominant) or unstructured (minor). We note that the most dominant seryl side-chain / carbonyl DARR cross-peak is at 174.0 ppm, aligning with α-helical secondary structures, further supporting that the dominant Ser environment is α-helical.

Valine: There are 13 total valine residues in the *Argiope argentata* AcSp1 repeat unit, 12 of which are found in the globular bead region and only one found in the linker. With alanine and serine amino acids well represented in both the bead and string regions, we collected data on $^{13}$C-Val enriched prey-wrapping silk to better understand the divide between the bead and linker after fibrillization (**Figure 69**). Again, we extracted exact $^{13}$C chemical shifts for the dominant Val Cα, Cβ, and CO resonances, which as anticipated, align very clearly with α-helical secondary

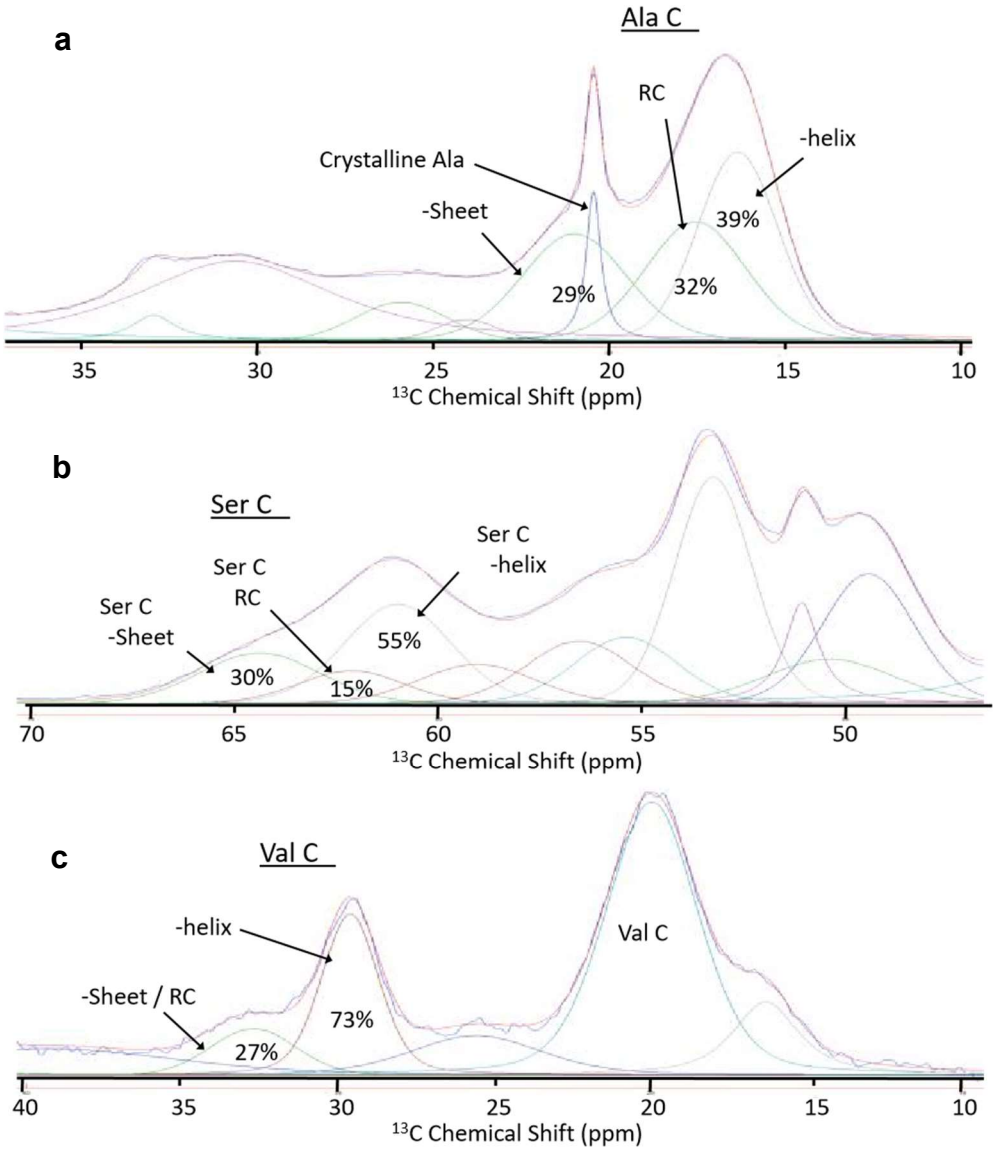



**Figure 70.** Spectral deconvolutions of Ala, Ser and Val Cβ resonances from $^{1}$H-$^{13}$C CPMAS data collected on isotopically-enriched freshly-spun *Argiope argentata* aciniform silks.

structure. We utilized the Val Cβ resonance to quantify helical versus non-helical structures because it is the most well resolved (**Figure 70c**). Peak fitting results suggest that at minimum, 73% (likely 80-90%) of all Val residues adopt an α-helical environment, with a minor RC and/or β-sheet component. We could not confidently assign the likely protein structure of the Val Cβ shoulder because no clear DARR correlations with Cα or CO are observed, thus the minor Val environment is assumed to be a combination of random-coil and/or β-sheet structures in addition to background signal from natural-abundant and partially-labelled signals (**Fig. S6**).

To make sense of these results we directly compared our quantifications to the primary protein sequence of the *A. argentata* AcSp1 repeat unit. Solid-state NMR data on silk biopolymers has repeatedly been used to quantify secondary structure content at the amino-acid level, and also to quantitatively correlate the primary amino acid sequence to said secondary structures.[17-20] Our peak-fitting results, when taken together, point to a protein structural model of native aciniform silk in which A) the majority of the silk fibroin exists as fiber-aligned α-helices (~45-50%) or loosely structured regions (~35%), B) the helical domains likely form some higher-order coiled-coil suprahelical structure, C) β-sheet aggregation occurs upon fibrillization, accounting for about 15% of the total fiber, likely from possible sheet-forming sequences in the linker domain, and D) the Poly(Ser) region of helix-5 remains helical or partly unstructured in the native silk. The arguments for this proposed model go as follows. To generate a complete model of AcSp1 in its final fibrous form we first consider that ~73% of all Val residues adopt true α-helical structures. Interestingly, the solution-NMR structure of the AcSp1 W unit identifies 6 Val residues in the "bead" region adopting a true helical conformation, 6 additional Val residues in the bead region that are on the ends of individual helices and are loosely helical or unstructured, and finally one clearly unstructured Val in the linker domain (**Figure 68**). Considering our 73% observation, it is

clear that the loosely structured Val residues on the edges of helices are enticed to adopt well defined helices upon fiber formation. We also note that hydrophobic residues Val, Leu and Ile exist predominantly in helices 1, 2, 4 and 5, often in patterns expected for coiled-coil suprahelical structures,[21] while such hydrophobic-rich residues are nearly non-existent in helix 3 and in the linker domain (**Figure 72**). With this data we can hypothesize which specific regions in the AcSp1 primary protein sequence adopt αhelical structures in the final fiber; we propose that helices 1, 2, 4 and 5 of the 5-helix bundle (bead) remain in a helical conformation as the fiber is pulled from the spider (**Figure 44**), most likely as a coiled-coil suprahelical structure where hydrophobic residues are buried in the coiled-coil core (**Figure 72, 73**). We fully expect these helices to be fiber-aligned, as previously suggested.[11,12] In addition to valine information we can map onto the protein sequence our fitting results for Ala (~40% helical, 30% β-sheet) and Ser (~55% helical, 30% β-sheet) residues. From the primary protein sequence of *A. argentata* prey-wrapping silk repeat unit we identified 5 possible short stretches in the linker region and one from helix-5 that might be prone to form β-sheets similar to well-known sheet-forming motifs found in other common

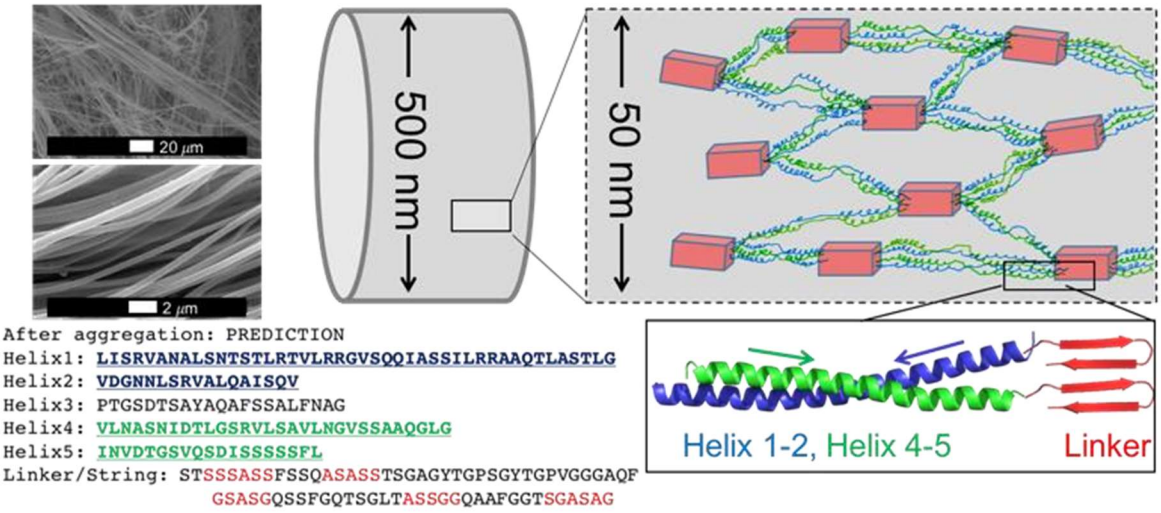



**Figure 71.** Proposed hierarchical molecular protein structure of aciniform prey-wrapping silks as a hybrid coiled-coil and nanocrystalline β-sheet fibroin. Helix 1 and 2 form a helical coil (blue), Helix 3 is unstructured, acting as a possible turn, helices 4 and 5 form another single coil (green), and the Ser / Ala-rich regions of the linker form pleated β-sheet (red) subunits.

silks: SSSASS, ASASS, GSASG, ASSGG, SGASAG, and SSSSS. Indeed, these motifs (highlighted with red lettering in **Figure 68, red lettering**) show a high propensity for β-sheet aggregation in our preliminary Monte-Carlo Molecular Dynamics simulations (**Figure 75, 76**). From these regions we can estimate 7 Ala residues out of a total of 28 (25%) might form β-sheet nanostructures upon fibrillization, which agrees reasonably well with our experimental data (~29%). However, if we expect that all of the possible β-sheet prone serine-rich motifs including the Poly(Ser) run in helix-5 indeed form β-sheets, then roughly 40% of all seryl residues should be β-sheet, 30% helical and 30% RC (**Fig 74**). This prediction does not agree with our experimental results. We therefore revisited the repeat sequence and proposed multiple iterations of how the AcSp1 repeat might exist in its final fiber (**Figure 74**). A sequence-based structure prediction that best agrees with our experimental NMR data is one in which the β-sheet aggregation occurs in the linker domain while the poly(Ser) motif in helix-5 remains helical or loosely-structured (**Figure 71**). If we also consider the density of hydrophobic residues Val, Leu and Ile in the globular region (found in helices 1, 2, 4, 5 but not in helix 3), and also the location of charged residues Arg (higher density in helices 1 and 2) and Asp (higher density in helices 4, 5), we can propose an antiparallel coiled-coil motif in which helices 1 and 2 form a single α-helical coil, helix 3 is more randomly-oriented, helices 4 and 5 form a second antiparallel coil, and finally, the Ser / Ala rich motifs at the beginning and end of the linker domains align to form antiparallel pleated sheets.

**Conclusion**

In conclusion, through our Solid-State NMR data we have shown that prey-wrapping silks, which are actually the toughest of the spider silks yet are surprisingly understudied, are dominated by α-helical secondary structures (~40-50% of total fiber) with only minor β-sheet content (~15% total). This is in loose agreement with Raman spectroscopy studies on aciniform fibers (~25% α-helices, ~30% β-sheets), 9,11,12 although in contrast to the authors conclusions, here we find

that α-helix-to-β-sheet conversion is minimal, and prey-wrapping silk β-sheet content is low compared to other spider silks. Spectral deconvolutions of NMR data show that β-sheet nucleation indeed occurs during fibrillization, but this is likely from serine- and alanine-rich domains found in the linker region while the helical domains remain largely in-tact. True α-helical silks are observed in other insect silks, [21] but never before have they been observed in spiders. Aciniform silks are capable of extending about twice as far before breaking compared to dragline fibers, a property that is most certainly a direct product their high helical content.

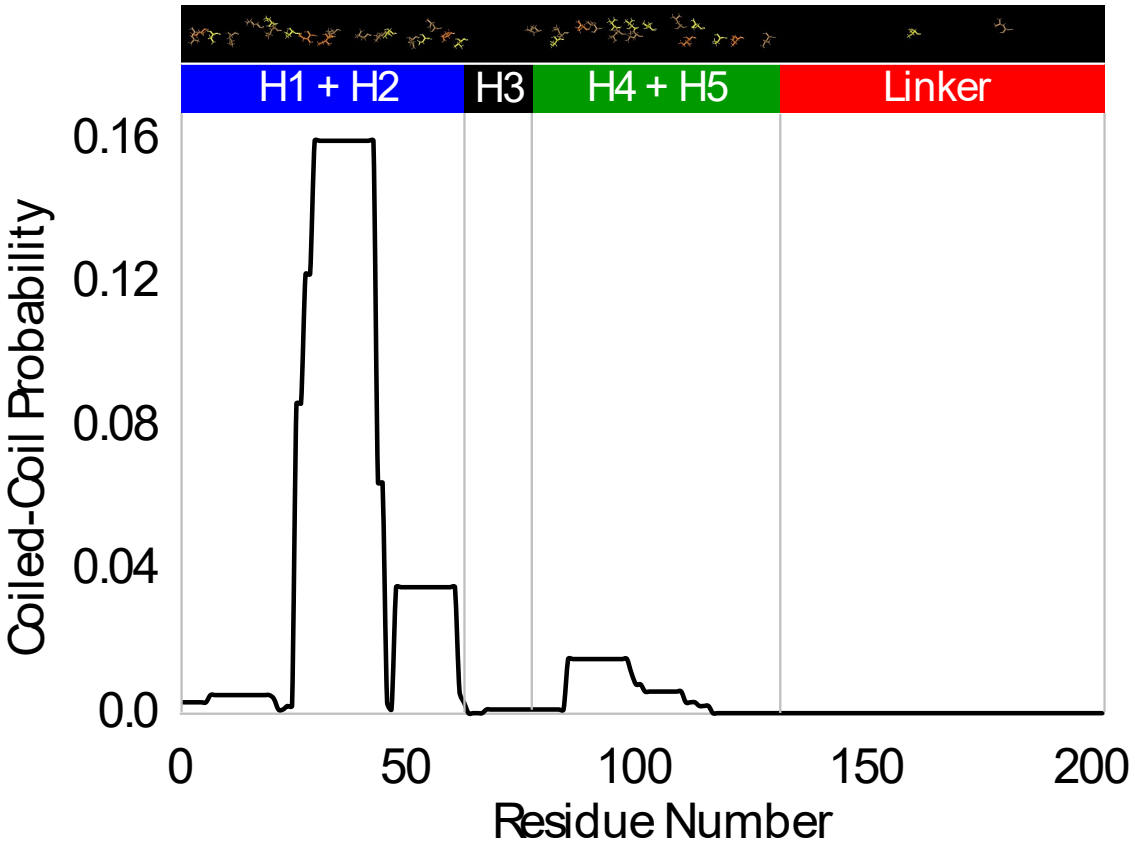



**Figure 72.** Primary protein sequence alignment comparing A. argentata and A. trifasciata repeat units.

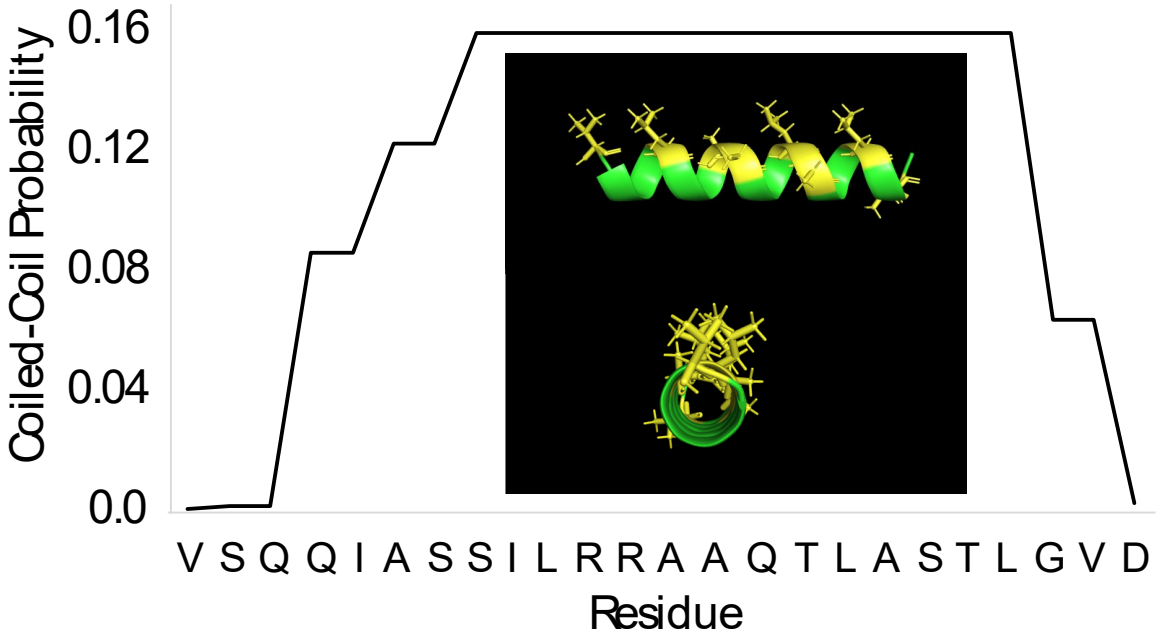



**Figure 73.** Prediction of coiled-coil probabilities of residues 23 – 46 from the AcSp1 repeat unit from A. argentata, along with a cartoon structure of the sequence modeled as an ideal α-helix. Hydrophobic residues Val, Ile, Leu and Ala are shown in yellow, while the rest of the sequence is green. Here we can see a clear alternating hydrophobic - polar $(HHPP)_x$ pattern, resulting in opposing hydrophobic and polar faces.

After aggregation: PREDICTION 1
Helix1: **LISRVANALSNTSTLRTVLRRGVSQQIASSILRRAAQTLASTLG**
Helix2: **VDGNNLSRVALQAISQ**V
Helix3: PTGSDTSAYAQAFSSALFNAG
Helix4: **VLNASNIDTLGSRVLSAVLNGVSSAAQGLG**
Helix5: **INVDTGSVQSDISSSSSFL**
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

| Residue | Total | % RC | % Helical | % Sheet |
|---|---|---|---|---|
| Ala | 28 | 32 | 43 | 25 |
| Ser | 47 | 28 | 43 | 30 |
| Val | 13 | 15 | 85 | 0 |
| All | 204 | 33 | 53 | 13 |

After aggregation: PREDICTION 2
Helix1: **LISRVANALSNTSTLRTVLRRGVSQQIASSILRRAAQTLASTLG**
Helix2: **VDGNNLSRVALQAISQ**V
Helix3: PTGSDTSAYAQAFSSALFNAG
Helix4: **VLNASNIDTLGSRVLSAVLNGVSSAAQGLG**
Helix5: **INVDTGSV**QSDISSSSSFL
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

| Residue | Total | % RC | % Helical | % Sheet |
|---|---|---|---|---|
| Ala | 28 | 32 | 43 | 25 |
| Ser | 47 | 28 | 32 | 40 |
| Val | 13 | 15 | 85 | 0 |
| All | 204 | 36 | 48 | 16 |

After aggregation: PREDICTION 3
Helix1: **LISRVANALSNTSTLRTVLRRGVSQQIASSILRRAAQTLASTLG**
Helix2: **VDGNNLSRVALQAISQ**V
Helix3: PTGSDTSAYAQAFSSALFNAG
Helix4: **VLNASNIDTLGSRVLSAVLNGVSSAAQGLG**
Helix5: **INVDTGSVQSDISSSSSFL**
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

| Residue | Total | % RC | % Helical | % Sheet |
|---|---|---|---|---|
| Ala | 28 | 36 | 43 | 21 |
| Ser | 47 | 32 | 43 | 26 |
| Val | 13 | 15 | 85 | 0 |
| All | 204 | 36 | 53 | 11 |

After aggregation: PREDICTION 4
Helix1: **LISRVANALSNTSTLRTVLRRGVSQQIASSILRRAAQTLASTLG**
Helix2: **VDGNNLSRVALQAIS**QV
Helix3: PTGSDTSAYAQAFSSALFNAG
Helix4: VLN**ASNIDTLGSRVLSAVLNGVSSAAQGLG**
Helix5: **INVDTGS**VQSDISSSSSFL
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

| Residue | Total | % RC | % Helical | % Sheet |
|---|---|---|---|---|
| Ala | 28 | 36 | 43 | 21 |
| Ser | 47 | 32 | 32 | 36 |
| Val | 13 | 23 | 77 | 0 |
| All | 204 | 41 | 46 | 13 |

After aggregation: PREDICTION 5
Helix1: LISR**VANALSNTSTLRTVL**RRG**VSQQIASS**ILRRA**AQTLAS**TLG
Helix2: **VDGNNLSRVALQAISQV**
Helix3: PTGS**DTSAY**AQA**FSSALFN**AG
Helix4: **VLNASNIDTLGSRVLSAVLNGVSSA**AQGLG
Helix5: INVDTGSVQ**SDISSSSSFL**
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

| Residue | Total | % RC | % Helical | % Sheet |
|---|---|---|---|---|
| Ala | 28 | 32 | 43 | 25 |
| Ser | 47 | 23 | 47 | 30 |
| Val | 13 | 23 | 77 | 0 |
| All | 204 | 41 | 46 | 13 |

---

Before Fibrilization, from Solution-NMR Structure
Helix1: LISRV**ANALSNTSTLRT**VLRRGV**SQQIAS**SILRRA**AQTLAS**TLG
Helix2: VD**GNNLSRVALQAISQV**
Helix3: PTGS**DTSAY**AQA**FSSALFN**AG
Helix4: VL**NASNIDTLGSRVLSAVLNGVSSA**AQGLG
Helix5: INVDTG**SVQSDISSSSSFL**
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

| Residue | Total | % RC | % Helical | % Sheet |
|---|---|---|---|---|
| Ala | 28 | 57 | 43 | 0 |
| Ser | 47 | 53 | 47 | 0 |
| Val | 13 | 54 | 46 | 0 |
| All | 204 | 56 | 44 | 0 |

**Figure 74.** Five sequence-based structure predictions of the A. argentata wrapping silk repeat unit after fiber formation (above horizontal line), with solution-NMR conformation added for context (below horizontal line, see **Figure 68** in main text). Residues that are bolded and underlined are helical, while residues in red lettering are proposed to form β-sheets upon fibrilization. For each variant, tables are shown summarizing the total number of Ala, Ser, and Val residues adopting random-coil (RC), α-helical and β-sheet structures, as well as a sum total of all residues. A green checkbox or a red cross indicates which predictions agree or disagree with experimental NMR data, and the size of those icons loosely indicates our confidence in agreement or disagreement.

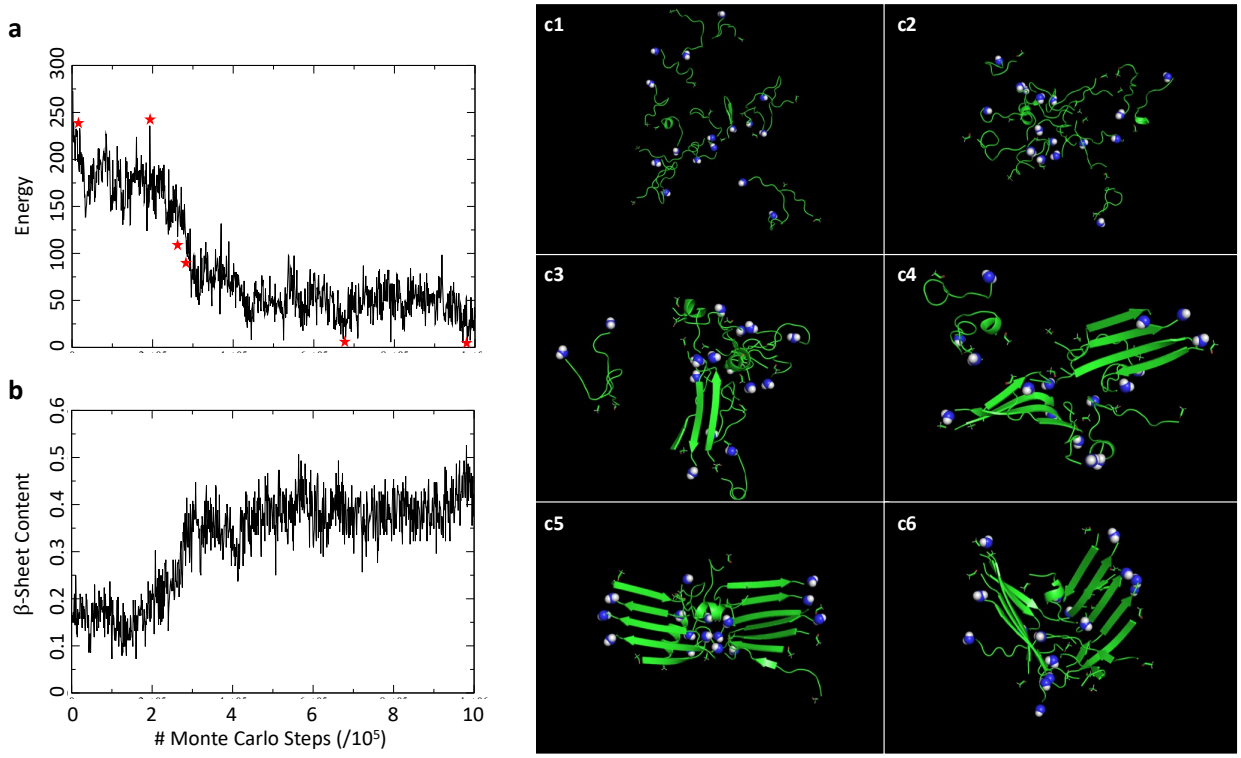



**Figure 75.** Preliminary Monte-Carlo Simulation results on possible β-sheet forming domains from the AcSp1 linker, in this case 8x Acetyl-STSSSASSFSS-NH2 (chain 1) and 8x Acetyl-AAFGGTSGASAG-NH2 (chain 2), representing the beginning and ending β-sheet prone motifs in the linker domain. Acetyl and NH2 capping groups were included to eliminate unwanted charge-induced stabilization. Monte-Carlo evolution of the energies (a) and β-sheet content (b) are shown. Red stars indicate time points for snapshots in c1 – c6. In this particular simulation, β-sheet nucleation occurs between two strands chain 1 and chain 2 (c3), which then seeds larger scale aggregation. Monte-Carlo simulations were conducted using the software package PROFASI version 1.5, using 16 total chains (see above) in a 60 Angstrom box at 298 Kelvin. 1,000,000 Monte-Carlo cycles were executed, corresponding to 1,000,000,000 elementary Monte-Carlo updates. The simulation was performed utilizing the San Diego State Computing Cluster. All images were rendered in Pymol version 2.1.1. By no means should these computational results be considered conclusive or exhaustive, but they do support our hypothesis that the Ala- and Ser-rich motifs in the linker domain are responsible for β-sheet aggregation upon AcSp1 fibrilization.

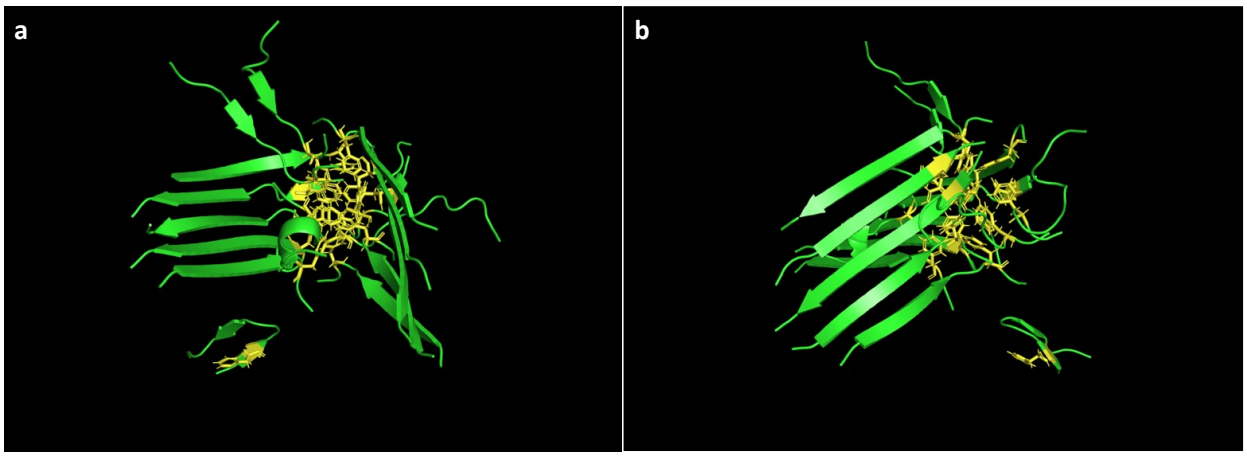



**Figure 76.** Representative minimum energy structure demonstrating AcSp1 linker aggregation from the MD simulation displayed above (frame c6), with Phe residues highlighted in yellow.

**Acknowledgements**

The above was published in Chemical Communications which is publication of the Royal Society of Chemistry. The following authors contributed to this work. B. Addison, performed the ssNMR experiments, compiled the fitting results and wrote the text. D. Onofrei, labeled silk samples and performed fitting for different amino acids D. Stengel provided another set of fit analysis. B. Blass, B. Brenneman, and J. Ayon, provided valuable data that helped shape the manuscript. G. P. Holland conceived and led the project.

Chapter 6 is a full reprint of “Spider Prey-Wrapping Silk is an α-Helical Coiled-Coil/β-Sheet Hybrid Nanofiber” Published in Chem. Commun. 2018. Addison, B.; Onofrei, D.; Stengel, D.; Blass, B.; Brenneman, B.; Ayon, J.; Holland, G. P. Bennett Addison was the primary author of this paper.

**References**


1. R. V. Lewis, Chemical reviews, 2006, 3762–3774.

2. L. Eisoldt and A. Smith, Materials Today, 2011.

3. C. Y. Hayashi, T. A. Blackledge and R. V. Lewis, Mol. Biol. Evol., 2004, 21, 1950–1959.

4. J. Gosline, M. Lillie, E. Carrington, P. Guerette, C. Ortlepp and K. Savage, Philosophical Transactions of the Royal Society B: Biological Sciences, 2002, 357, 121– 132.

5. J. L. Yarger, B. R. Cherry and A. van der Vaart, Nature Reviews Materials 2018 3:3, 2018, 3, 18008.

6. R. C. Chaw, Y. Zhao, J. Wei, N. A. Ayoub, R. Allen, K. Atrushi and C. Y. Hayashi, BMC Evolutionary Biology 2010 10:1, 2014, 14, 31.

7. L. Xu, M.-L. Tremblay, Q. Meng, X.-Q. Liu and J. K. Rainey, Biomol NMR Assign, 2011, 6, 147–151.

8. M.-L. Tremblay, L. Xu, T. Lefèvre, M. Sarker, K. E. Orrell, J. Leclerc, Q. Meng, M. Pézolet, M. Auger, X.-Q. Liu and J. K. Rainey, Sci. Rep., 2015, 5, srep11502.

9. M. Sarker, K. E. Orrell, L. Xu, M.-L. Tremblay, J. J. Bak, X.-Q. Liu and J. K. Rainey, ACS Publications, 2016.

10. M.-E. Rousseau, T. Lefèvre and M. Pézolet, Biomacromolecules, 2009, 10, 2945–2953.

11. T. Lefèvre, S. Boudreault, C. Cloutier and M. Pézolet, Journal of Molecular Biology, 2011, 405, 238–253.

12. D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes, Journal of biomolecular NMR, 1995, 5, 67–81.

13. Y. Wang, Protein Science, 2002, 11, 852–861.

14. K. Takegoshi, S. Nakamura and T. Terao, Chemical Physics Letters, 2001, 344, 631–637.

15. K. Takegoshi, S. Nakamura and T. Terao, The Journal of Chemical Physics, 2003, 118, 2325–2341.

16. G. P. Holland, J. E. Jenkins, M. S. Creager, R. V. Lewis and J. L. Yarger, Chem. Commun., 2008, 5568.

17. J. E. Jenkins, M. S. Creager, R. V. Lewis, G. P. Holland and J. L. Yarger, Biomacromolecules, 2009, 11, 192–200.

18. J. B. Addison, T. M. O. Popp, W. S. Weber, J. S. Edgerly, G. P. Holland and J. L. Yarger, RSC Advances, 2014, 4, 41301–41313.

19. J. E. Jenkins, S. Sampath, E. Butler, J. Kim, R. W. Henning, G. P. Holland and J. L. Yarger, Biomacromolecules, 2013, 14, 3472–3483.

20. D. N. Woolfson, Fibrous Proteins: Structures and Mechanisms, Springer, Cham, 2017, vol. 82.

21. T. D. Sutherland, T. D. Sutherland, S. Weisman, S. Weisman, A. A. Walker, A. A. Walker, S. T. Mudie and S. T. Mudie, Biopolymers, 2012, 97, 446–454.

**Chapter 7:** Hydration-Induced β-Sheet Crosslinking of α-Helical-Rich Spider Prey-Wrapping Silk

**Abstract**

Due to its moderate strength (~700 MPa) and impressive extensibility before breaking (~60-80%), orb-weaving spider aciniform (AC) prey-wrapping silks are actually the toughest of the spider silks but are remarkably understudied. Our previous results indicate that native AC silk fibers are an α-helix rich coiled-coil / β-sheet hybrid nanofiber, and that conversion of disordered or helical domains to β-sheet aggregates is surprisingly minimal and overall β-sheet content is low (~15%). In this work, we demonstrate through Scanning Electron Microscopy (SEM) that native AC silk fibers undergo matted cross-linking upon exposure to moisture that increases silk stiffness. The unique molecular mechanism of water-induced cross-linking is revealed with solid-state NMR (SSNMR) methods; water-induced morphological changes are correlated with an increase in AC silk protein β-sheet content, and additionally we observe a minor unfolding of coiled-coil regions. Continued and increased β -sheet cross-linking is observed upon application of mechanical shear. We determine the size of these β-sheet domains to be 4-6 nm using WISE SSNMR. The observation that merely water treatment can be used to convert a protein-based material from a flexible/extensible α-helix-rich fiber to a crossed-linked β-sheet mat is a novel observation that should provide new avenues in bioinspired materials design.

**Introduction**

The mysteries of spider silks have intrigued researchers and inspired polymer engineers for decades due to their impressive and wide ranging mechanical properties, their biodegradability and biocompatibility, and the fact that nature's super fibers are spun on-demand at ambient temperature and pressure from an aqueous silk dope.[1, 2] It has been proposed that spider silks could be a suitable material for protective clothing, parachutes and cables, wound treatment, tissue scaffolding, and medical implants.[3-5] Spiders are unfortunately territorial creatures and therefore domestication and silk harvesting is not a practical strategy to bring the material to market. Artificial production and utilization of spidroin-based recombinant proteins are therefore

the only alternative. Imperative to attaining this goal is a more complete understanding of the molecular structure, function, storage mechanism and spinning process of the native material. Of the seven types of silks and glues produced by orb-weaving spiders, Aciniform (AC) prey-wrapping spider silks, utilized for prey-wrapping and egg-case linings, are the toughest of the spider silks, but surprisingly little information exists on their molecular structure and detailed structure-function relationship.[6-8] Remarkably, AC silks boast mechanical properties that weight for weight surpass the toughest natural and man-made materials, including Kevlar®, high-tensile steel and even the most studied spider silk, dragline silk.[6, 9] Due to their moderate (~700 MPa) mean ultimate strength and high (~60-80%) extensibility before breaking, AC silks are reportedly 50% tougher than the highly-coveted spider dragline fibers.[9, 10]

Generally, silk protein biopolymers are composed of highly repetitive, high molecular weight proteins (250 – 500 kDa) rich in small-sidechain amino acids Ala, Gly (~60-75%) and to a

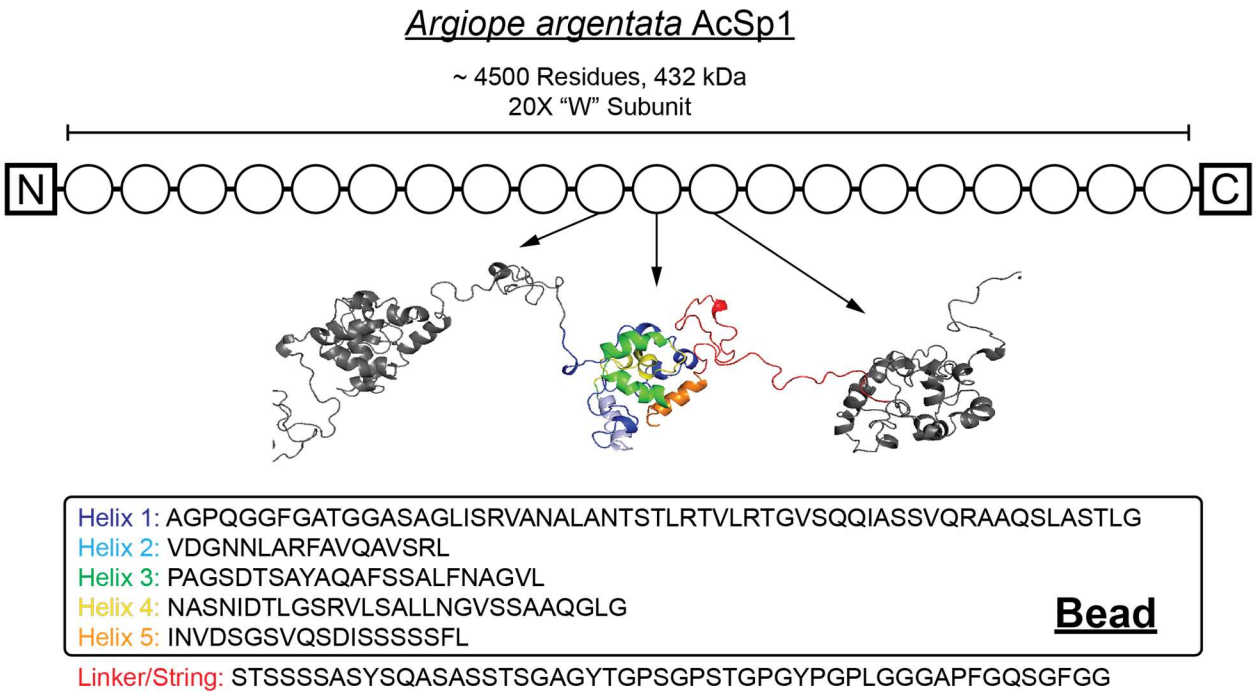



**Figure 77.** Schematic of *Argiope* AcSp1 prey wrapping silk. In the *A. argentata* species, there are 20 iterated repeats of the wrapping ("W") subunit. The solution NMR structure from a single *A. trifaciata* W subunit (PDB 2MU3) was used to create the theorized "beads-on-a-string" model hypothesized to occur in solution and colored to identify each area. The colored sequences are for *A. argentata* AcSp1.

variable extent Gln, Ser, Pro, Tyr and Arg.[11, 12] The major ampullate (MA) silk proteins that comprise dragline silk are highly dynamic, intrinsically disordered proteins (IDPs) as stored in the MA silk gland prior to fiber spinning.[13] In the final spun fibers, the poly(Ala)$_n$ (n = 4-9) and flanking (Gly-Ala) repeats assemble into β-sheet nanocrystallites (~2-7 nm)[14-17] dispersed throughout a Gly-Gly-X (X = Gln, Tyr, Pro, Ser and Arg)-rich matrix that is non-β-sheet and best described as a disordered helix that is not α-helical.[18, 19] Aciniform spidroin 1 (AcSp1), the primary protein making up spider wrapping silk, is markedly different than dragline silk and is composed of a string of 14-20 identical subunits (depending on species) of about 200 amino acids (called a W unit for wrapping), flanked by non-repetitive N- and C-terminal regions (**Figure 77**).[20] Solution-state NMR studies on a single recombinantly produced AcSp1 wrapping unit suggest that each 200 amino acid W subunit folds into a multi-domain structure composed of a well-defined helix-rich globular domain and a glycine/serine-rich disordered linker region.[21-24] It was proposed that each W subunit forms a "beads on a string" model when the protein is stored in the gland.[22] Synthetic versions of AC silk fibers have previously been generated from an aqueous dope in order to study the mechanical properties imparted by different buffers as well as differences in the number of W subunits.[25-27] Although recombinant AcSp1 W subunits have been well characterized as recombinant protein constructs in a solubilized state, a molecular-level understanding of the native AC silk protein after conversion to insoluble fiber is quite minimal. Recently, we applied Solid-State Nuclear Magnetic Resonance (SSNMR) techniques to isotopically-enriched native prey-wrapping silks and quantitatively correlated the secondary protein structures of Ala, Ser and Val residues to the primary protein sequence of the W subunit. We proposed a hybrid α-helical coiled-coil / β-sheet hierarchical model for *A. argentata* and *A. aurantia* spider prey-wrapping silk fibers, in which 1) the globular helix-rich domains (bead) mostly retain their helical structures in the final fiber, 2) where the Ala / Ser-rich motifs from the string / linker region aggregate into β-sheet nanostructures, and 3) a significant fraction of the protein remains disordered.[28] We

intuitively argued that the high extensibility of AC silk fibers is a direct result of high α-helical content, while its moderate mean ultimate strength arises from β-sheet nanostructures and possibly coiled-coil interactions between fiber-aligned helices.

It is likely that when stored in the AC gland, the silk protein is highly helical with no evidence of β-sheet secondary structure, but there is clear evidence of a minor population of β-sheet aggregates in the final fiber.[28-30] Hence, a structural conversion from either RC and/or α-helical regions into β-sheet nanostructures takes place upon fibrilization. There is some debate over the precise mechanism. For example, Raman studies suggest that wrapping silks from *N. clavipes* are roughly 50% helical in the gland and 24% helical and 30% β-sheet in the fiber,[29, 30] which are in poor agreement with our recent results where we determine the native fiber is ~50% helical and ~15% β-sheet, albeit from a different spider species.[28] Additionally, solution-state NMR data on *A. trifaciata* recombinant W subunits show higher dynamics of the Ser-rich stretch in the globular domain, denoted helix 5 (**Figure 77**), and the authors suggest that the Ser-rich helix 5 motif is a likely candidate for α-helix to β-sheet aggregation as a driving event for fibrilization.[22, 24] However, this again conflicts with our SSNMR results, namely that in order for Ala, Val and Ser secondary structures to all agree with SSNMR observations, the Ser-rich region in the helix 5 domain must remain helical or loosely-helical but not β-sheet in the fiber.[28] Perhaps the true fibril protein structure is somewhere in between, existing as heterogeneous structures with incomplete conversion of poly(Ser) and poly(Ala-Ser) motifs from RC and/or α-helices in the gland to β-sheet structures in the final fiber. As such, there are clearly regions of the primary protein sequence that may be primed to β-sheet aggregate if given the appropriate conditions.

In this work we report further molecular and nanoscale structural insights into native AC silk fibers and demonstrate that AC silks exhibit a remarkable β-sheet fiber cross-linking property when the silk is exposed to water that has not been observed in any other silk to the best of our knowledge. Remarkably, single ~500 nm wrapping fibers are visibly cross-linked into fibrous

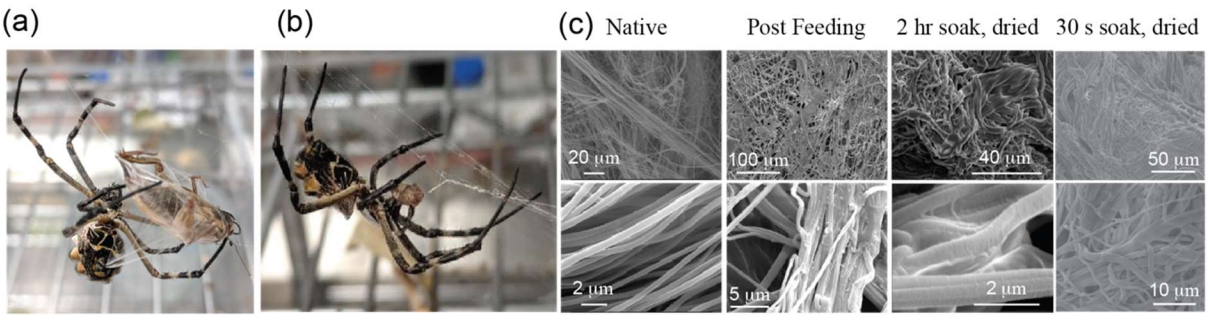



**Figure 78.** a), *A. argentata* spinning a cricket with AC silk. b), the same spider returning to eat its prey through EOD. c), SEM images comparing native prey-wrapping AC silk from *A. argentata* following feeding or water treatment where the silk was soaked for 2 hr or 30 s and then dried for one day.

sheets and mats when exposed to water. SSNMR results on native and water-induced cross-linked fibers illustrate that disordered linker regions and α-helical structures decrease with a corresponding increase in β-sheet content for Ala, Ser, Gly, Gln and Val residues. This suggests that the linker/string environments and loosely-structured helices are most sensitive to water-induced β-sheet crosslinking. When mechanical stress is applied to hydrated silks, the data shows a further increase in β-sheet structures and a corresponding reduction in α-helical and random coil content. We hypothesize that exposure to moisture provides disordered and semi-stable helical regions of the wrapping silk with the necessary degrees of freedom to access lower-energy β-sheet aggregates. We think this remarkable moisture-induced cross-linking behavior is an evolutionarily adapted trait that benefits the spider during Extra-Oral Digestion (EOD) and feeding, and should be of great interest to the synthetic biomimetic materials community; converting a malleable and extensible fibers into rigid crossed-linked biomaterial simply by hydrating with water and drying.

**Results and Discussion**

Freshly-collected AC silk is shown as a dense mesh of fine (~500 nm in diameter) fibers (**Figure 78C**). When we attempted to image silk on a wrapped and consumed prey (cricket), we were surprised to observe clear morphological changes to regions of the silk. Dense bundles of individual AC fibers remain, but areas of wrapping silk fibers appear to have fused together into

**Table 9.** Nano-indentation mechanical properties of native and water treated AC spider silks.

| ***Native Aciniform (as-spun)*** | | ***Aciniform wetted (1-day dry)*** | | ***Aciniform wetted (2-day dry)*** | |
|---|---|---|---|---|---|
| ***Material Stiffness (GPa)*** | ***Std Dev (GPa)*** | ***Material Stiffness (GPa)*** | ***Std Dev (GPa)*** | ***Material Stiffness (GPa)*** | ***Std Dev (GPa)*** |
| 4.40 | 0.28 | 4.76 | 0.15 | 4.11 | 0.05 |

fibrous sheets or mats, especially in regions where the silk makes direct contact with the cricket. Individual fibers clearly have fused together to form larger, uniaxial fibers and fibrous mats. Additionally, inter-fiber space appears to have collapsed. To determine the effect of water on AC fiber morphology, native prey-wrapping silk was briefly soaked in deionized water for two hours without mechanical agitation and dried for 1 day. SEM images of our water-wetted silk bundles confirm that this observed cross-linking property is water-induced. The effect of hydration on silk structure has been investigated for several silk-based materials including native silk fibers, recombinant silk protein films, and peptides.[31-34] The effect of water on protein secondary structure has been conclusive, and results in an increase in β-sheet content of the materials, including silkworm silk fibers.[31, 33] However, these changes occur within a single fiber, not between adjacent strands. After allowing our AC silk fibers to air-dry, the material became noticeably stiffer and was no longer malleable like the native silk fiber. It is possible that other factors, such as dissolved salts and pH used in the native (wrapped cricket) fibers, may fine tune the local silk fusion and aggregation to minimize fluid loss from the prey. The appearance of nearly identical morphological structures in both native and water-treated samples, however, leads us to the conclusion that AC silk protein-water interactions drive the mechanism for fiber fusion and cross-linking. Lastly, the fiber fusion and cross-linking process occurs rapidly with similar observations made for AC silks water treated for 2 hr and 30 s (**Figure 78c**).

The mechanical properties of native and hydration-induced cross-linked AC silks were determined with Force-displacement AFM (see Figure 83 for AFM topography images). As shown

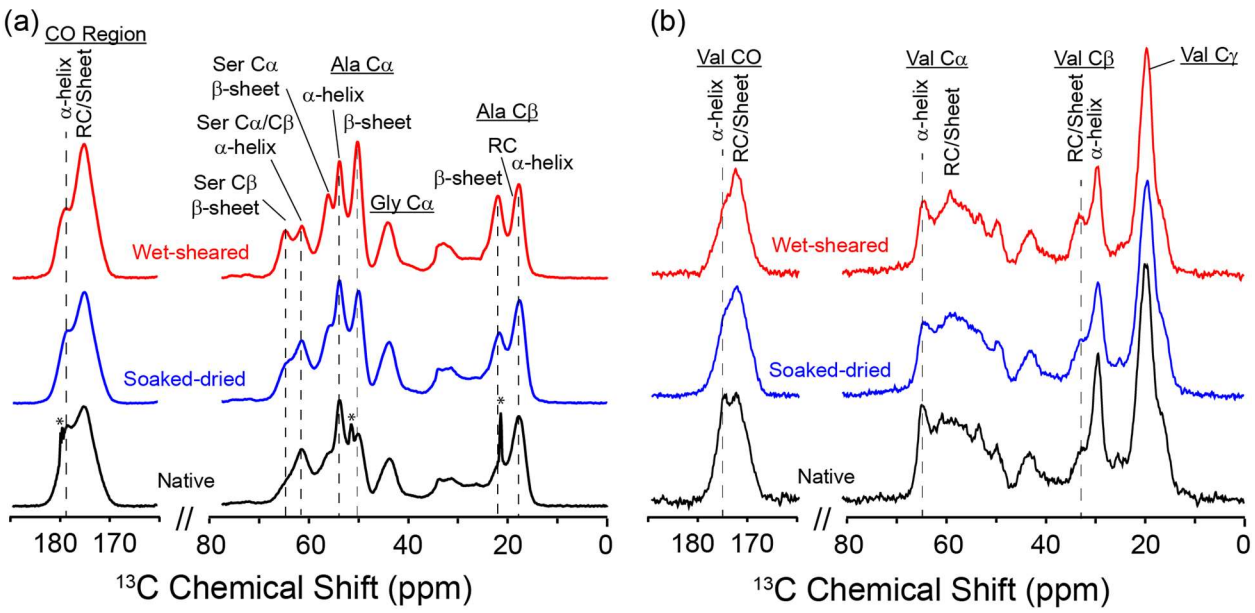



**Figure 79.** $^{13}$C CP-MAS SSNMR spectra of *A. argentata* AC silk. a) AC silk labeled with $^{13}$C-Ala. b) AC silk labeled with $^{13}$C-Val. Peaks in a) with asterisks are from crystalline Ala contamination.

in the model (see Methods), the resulting material stiffness is dependent on the value of the Poisson's ratio. The Poisson's ratio is not known explicitly for AC spider silks however, it has been determined for a number of dragline (MA) spider silks from a range of species.[35] Considering only the linear elastic response and assuming the material is isotropic and homogeneous, the assumption of Poisson's ratio of 0.3 is reasonable and consistent with the underpinnings of the model, resulting in a material stiffness of 4.40±0.28 GPa (n=29), 4.76±0.15 GPa (n=31), and 4.11±0.05 GPa (n=17) for native AC, 1-day dehydration, and 2-day dehydration conditions, respectively **(Table 9).**

These results illustrate that the stiffness of AC silk following water treatment increases and this increase is consistent with the water content of the cross-linked silk. Thermal analysis was used to determine the water content (**Figure 86**, **Table 13**) and it was found that the native AC silk, AC silk water-wetted for 2 hr and dried for 1 or 2 days have water contents of 3.2, 5.8

**Table 10.** Summary of conformation dependent $^{13}C$ isotropic chemical shifts and secondary structure populations determined from peak deconvolutions of $^{13}C$ CP-MAS spectra for AC silks (Figure 80).

| Residue | Chemical Shift[a] (ppm) | | | Peak Width (ppm) | | | Normalized Peak Area (%) | | |
|---|---|---|---|---|---|---|---|---|---|
| **Ala Cβ** | Native | SD[b] | WSD[c] | Native | SD[b] | WSD[c] | Native | SD[b] | WSD[c] |
| **α-helix** | 18.2 | 18.3 | 18.3 | 2.7 | 2.7 | 2.6 | 38.0 | 32.7 | 26.5 |
| **Random Coil** | 19.6 | 19.6 | 19.5 | 3.3 | 3.7 | 2.9 | 32.9 | 26.1 | 19.9 |
| **β-sheet A** | 22.9 | 22.9 | 22.8 | 3.4 | 3.2 | 3.2 | 24.0 | 35.6 | 48.1 |
| **β-sheet B** | 25.8 | 25.8 | 25.8 | 2.8 | 2.3 | 2.1 | 5.1 | 5.6 | 5.5 |
| **Val Cβ** | | | | | | | | | |
| **α -helix** | 31.6 | 31.6 | 31.7 | 1.9 | 1.9 | 1.9 | 62.0 | 54.4 | 50.9 |
| **RC / β-sheet / Background** | 34.8 | 34.8 | 35.0 | 3.1 | 3.1 | 3.1 | 38.0 | 45.6 | 49.1 |

a) from DSS
b) SD, Soaked-dried
c) WSD: Wet-sheared, dried

and 4.7%, respectively. This correlates with the material stiffness and implicates the role of water in impacting fiber cross-linking and the resulting mechanical properties where the 1 day dried (5.8 wt%) displays a higher stiffness than the 2 day dried (4.7 wt%). Further mechanical characterization is ongoing in our labs and will be the subject of future publications.

The likely biological reasoning behind the observed phenomenon is that *Argiope* spiders, including the *A. argentata* species in this study, will first immobilize their prey using wrapping silk in an act called "attack wrapping" (Figure 78a), then bite to inject their venom.[36, 37] After paralyzing their prey, they then engage in EOD, whereby they digest their food outside of their bodies by regurgitating digestive fluids into or onto the prey to consume them, during which prey-package manipulation is common adding a degree of mechanical stimulation through physical agitation and rotation (**Figure 78a**).[38] Therefore, to better understand the structural changes induced by both moisture and mechanical agitation, SSNMR data was collected on AC silks after two types of sequential treatment. After initial SSNMR analysis on the native fibers (Native, **Figure 79**, black), the AC silk was removed from the NMR rotor and soaked with DI water at room temperature for two hours without any mechanical agitation and allowed to air-dry ("Soaked-

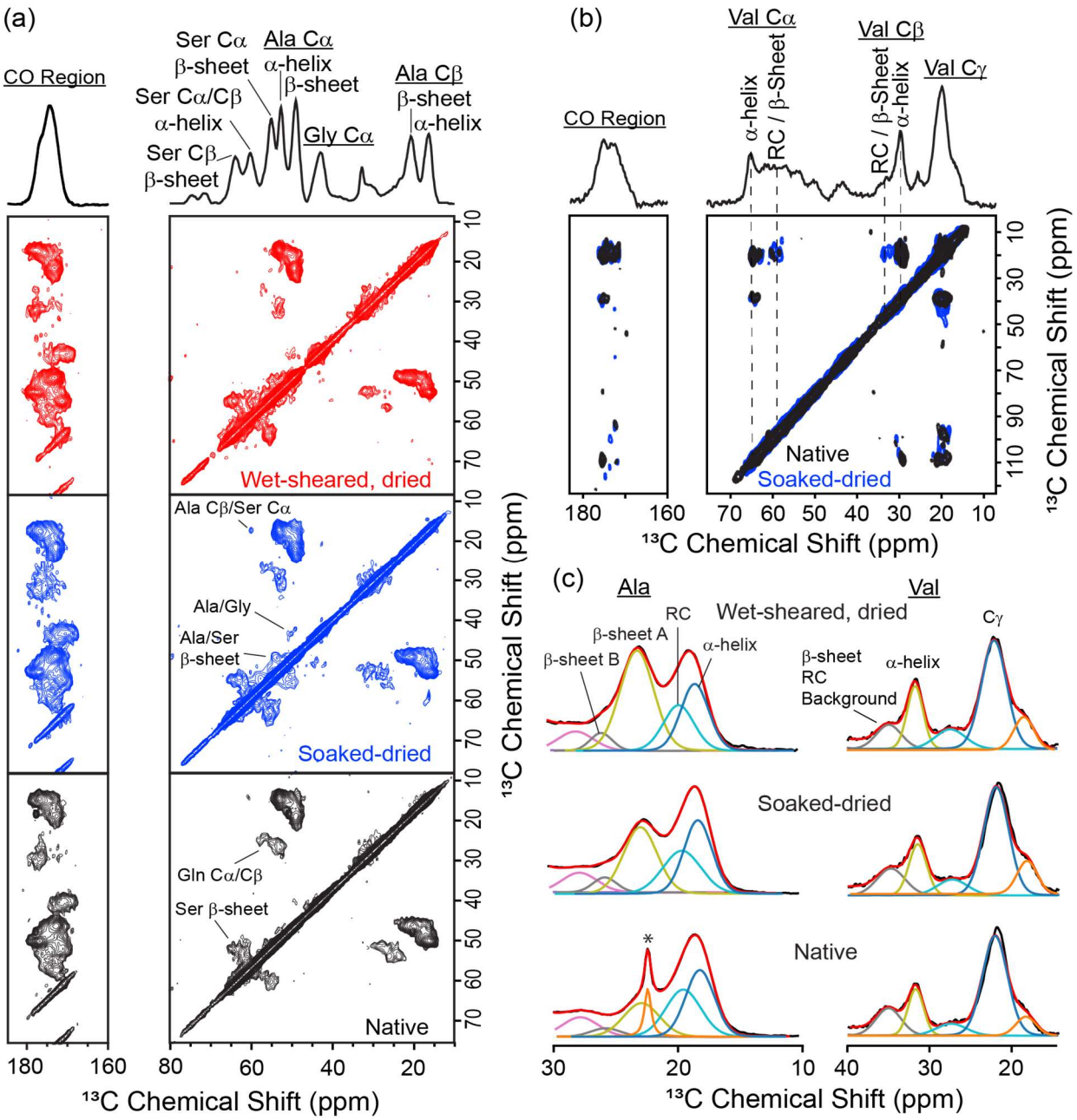



**Figure 80.** a), 2D $^{13}C$-$^{13}C$ DARR SSNMR spectra collected with a 100 ms mixing time for $^{13}C$-Ala-labeled *A. argentata* AC silk after different treatments. b), the same as in a) but only $^{13}C$-Val labeled. c), spectral deconvolutions of 1D $^{13}C$ CP-MAS data showing qualitative changes to the Ala Cα (left) and Val Cα (right) resonance after different treatments. Peak in c) with an asterisk are from crystalline Ala contamination. The fit parameters are summarized in Table 10.

dried", Figure 79, blue). After data collection, the AC silk was re-hydrated and loaded into the rotor while still wet which allows the spinning rotor to induce physical shear, removed and dried for two days, then reloaded. ("Wet-sheared, dried", Figure 79 red). The changes to the protein secondary

structure for several amino acids found throughout the W subunit were monitored throughout this process and a summary of these conformational changes (**Table 10**).

Several conclusions can be drawn from comparing 1D $^{1}H/^{13}C$ cross polarization magic angle spinning (CP-MAS) spectra between the native and water-treated silk (**Figure 79a**). When scaling the 1D $^{13}C$-Ala-labeled CP-MAS spectra based on the Gly C$\alpha$ resonance, it becomes clear that Ala $\beta$-sheet content increases moderately upon water wetting (**Figure 79a**). The reader is reminded that this sample was only submerged in water for two hours with no agitation and allowed to air-dry prior to data collection, thus these structural changes are truly a result of exposure to water and not from shearing or mechanical stress caused by MAS. It is not immediately clear if this observed structural transition is due to unfolding of $\alpha$-helical environments in favor of $\beta$-sheets, or if previously disordered regions of the silk protein are enticed to aggregate. Insight is gained through spectral deconvolutions of the data. To understand water-induced structural changes we first extracted precise chemical shifts from 2D $^{13}C$-$^{13}C$ DARR spectra (**Figure 80**a), and then peak-fit the 1D $^{13}C$ CP-MAS spectra using chemical shifts and estimated linewidths extracted from the 2D spectrum (**Figure 80c**). Two peaks are required to fit the Ala C$\beta$ $\beta$-sheet resonance because of different side-chain packing arrangements that have been proposed from studies on silk-like model peptides.[39] This procedure, outlined in the supplemental section of our prior work, was applied.[28] These results show that the $\beta$-sheet content for Ala increases from ~30% for the native silk to ~41% after wetting, and reaches a maximum of ~53% after wet-shear and drying (**Figure 79a**, **Table 10**). As a comparison, Ala within poly(Ala)-rich spider dragline silks show much higher $\beta$-sheet content at 82%.[40] This increase in $\beta$-sheet content for AC silks is mirrored by a decrease of Ala in RC structures, followed by $\alpha$-helical structures after shear (**Table 10**). Ala occupies $\alpha$-helical, RC and $\beta$-sheet structures in approximately equal proportions after wetting. After wet-shearing and drying, Ala adopting $\alpha$-helical and RC structures further decreases while $\beta$-sheet content increases to nearly half. The solution structure of the W

subunit shows that helix 1 contains 11 of the 26 Ala residues, where the linker contains five.[20] Therefore, this data implies Ala conversion likely comes from the partially-coiled helix 1 and linker regions.

Similar to Ala, there is a clear but minor increase in Ser adopting β-sheet structures after exposure to water. The relative intensity of the Ser Cα and Cβ β-sheet resonances at 55 and 65 ppm increase slightly upon exposure to water, and then increase dramatically after wet-shearing (Figure 79a). Unfortunately, due to the high spectral overlap in that region, we cannot confidently determine if this is from a partial disruption of α-helical structures or a collapse of disordered regions into β-sheet nanostructures. In our previous model of native AC silks, we demonstrated that the poly(Ser) motif at the end of helix 5 remains helical in the final fiber.[28] This is intriguing, since poly(Ser) motifs found in other insect silks have been shown to form β-sheet structures.[41, 42] While a partial α to β conversion of this poly(Ser) helix 5 may occur during water-wetting, the surprisingly low conversion towards β-sheets with only soaking suggests that a full conversion does not occur. After wet-shearing and drying, we see a more substantial decrease in Ser α-helix content in favor of β-sheet structures, implying collapse of the poly(Ser) helical motif.

1D $^{13}C$ CP-MAS data for $^{13}C$-Val enriched AC silk before and after water-washing suggests that Val helical structures partially collapse from 62% to ~51% (**Figure 79b**, **Table 10**), where broad and weak Val Cα/Cγ and Val Cβ/Cγ cross-peaks in the 2D DARR become slightly stronger (**Figure 80b**). This qualitative observation is supported by spectral deconvolutions of the Val Cβ signal, although we could not confidently separate β-sheet and RC structures (**Figure 80c**, **Table 10**). Since Val residues are almost exclusively found in the globular “bead” region and adopt coiled-coil like superstructures in the final fiber, it is likely that a minor conversion of Val from helical to RC and/or β-sheet motifs occurs in this region upon exposure to water. One explanation for this conversion is that several of the Val residues in the “bead” region are found as Gly-Val/Val-Gly pairs. Perhaps the juxtaposition of Gly and Val create a more favorable RC or β-sheet

environment during water-wetting, since Gly is known to be unstructured in many proteins. Indeed, closer inspection of the Gly peak in the Ala-labeled native AC silk (**Figure 79a**) shows a shoulder at ~47 ppm, corresponding to Gly in an α-helical environment, which decreases after treatment with water and suggests Val in Gly/Val pairs are no longer α-helical. This is an interesting observation in light of recent work into the silks of aquatic spiders, where Gly-Val motifs are common and are suggested to be involved in efficiency of underwater webs.[43] Such a response to water by Gly-Val motifs could be a common trigger among these silks. After wet-shearing and drying, a slight collapse in helices occurs from 54% to ~51% favoring RC and/or β-sheets, however the predominant structure for Val still remains α-helical. This observation strengthens our prior conclusion that Val residues predominantly exist in well-folded highly-stable α-helical coiled-coil hierarchical structures within spider prey-wrapping silks and conversion to β-sheet due to wetting or wet-shearing is minimal. The high degree of order for Val in α-helical coiled-coil assemblies is further supported by the linewidth of the Val Cβ resonances that are considerably narrower (285 Hz) compared to all other Cβ amino acid linewidths (315-510 Hz) in other structures (see **Table 10**).

Gln residues are distributed fairly evenly throughout the ”W” subunit, with seven in the bead region and two in the linker. In the freshly-spun fibers, 2D DARR data identifies multiple Gln environments including β-sheets, α-helices and disordered structures (**Figure 80a**). Low signal to noise makes it difficult to obtain quantitative measurements, but visual inspection of the 2D $^{13}C$-$^{13}C$ DARR SSNMR spectra seems to suggest a fairly even distribution of structures (Figure 80a, black). After water-wetting, Gln cross peaks look fairly similar with perhaps a slight but non-definitive increase in β-sheet (**Figure 80a**, **blue**). After wet-shearing however, Gln Cα/Cβ contacts become much more resolved and there is a clear bias for Gln Cα/Cβ β-sheet environments over helices (**Figure 80a**, **red**).

Several inter-residue cross peaks emerge or become clearer in the 2D $^{13}C$-$^{13}C$ SSNMR correlation spectra after the native silk is cross-linked with water. First, after water treatment a correlation between Ser C$\alpha$ (57.4 ppm) and Ala C$\alpha$ (51.4 ppm) in a $\beta$-sheet environment is considerably stronger (**Figure 80a**, **blue**). These cross correlations must arise from regions in the silk where Ala and Ser residues are near in space. The DARR 100 ms mixing period applied in these experiments is sufficiently long for $^{13}C$-$^{13}C$ contacts to occur within a single amino acid residue, and in some cases between neighboring amino acids, but not long enough to diffuse between differing domains.[44] Thus, we safely assume that these observed inter-residue correlations are from amino acid couplets in the primary sequence for AcSp1. Many Ala-Ser/Ser-Ala pairs exist in the W wrapping repeat and are concentrated in the aforementioned Ala-Ser/Ser-Ala-rich motifs in the linker domain, but some also are found in the globular helical bead. The simplest and most likely explanation for this emerging inter-residue correlation is that a higher fraction of Ser/Ala-rich motifs adopt $\beta$-sheets after the silk is exposed to water. An additional correlation between Gly C$\alpha$ and Ala C$\alpha$ becomes more evident after wetting (**Figure 80a**, **blue**). Based on the $^{13}C$ Ala C$\alpha$ chemical shift (~52.1 ppm), this most likely represents Ala/Gly pairs in disordered structures. Ala/Gly pairs are concentrated in the linker region along with a small grouping located in helix 1. Since helix 1 and the linker are continuous in the primary protein sequence, we suggest that Ala/Gly intermolecular interactions arise from these regions, possibly from different fiber strands. Typically, Gly is found in disordered regions within proteins and does not measurably contribute to $\alpha$-helical structures. It may be possible therefore, that upon water wetting Gly in the Ala/Gly pairs act as a trigger to unravel from the $\alpha$-helical structures to form inter-strand contacts similar to the Gly/Val motifs previously discussed above. Regardless, the increase in intensity of this inter-residue correlation after water wetting strongly indicates that overall order and rigidity has increased around these Gly/Ala amino acid pairs and spin-diffusion data for Gly indicates that it is located on the periphery of the $\beta$-sheet domains with a more rapid

buildup compared to Ala. This also appears to be the case in Black Widow MA silk where narrow $^{1}$H components spin-diffuse more rapidly into the broad Gly components that represent Gly in β-sheet domains (**Figure 85 and Table 12**).

Since β-sheet formation is the dominant mechanism behind water-induced cross-linking of prey-wrapping silk, we next implemented the 2D Wide-Line Separation (WISE) NMR technique with $^{1}$H spin-diffusion to better understand the domain size and overall water-accessibility of these β-sheet nanostructures. The concept behind the technique is outlined elsewhere,[45-47] but briefly, by observing the $^{13}$C chemical shift in the direct dimension and measuring the associated $^{1}$H profile in the indirect dimension, the 2D WISE NMR technique is capable of correlating local

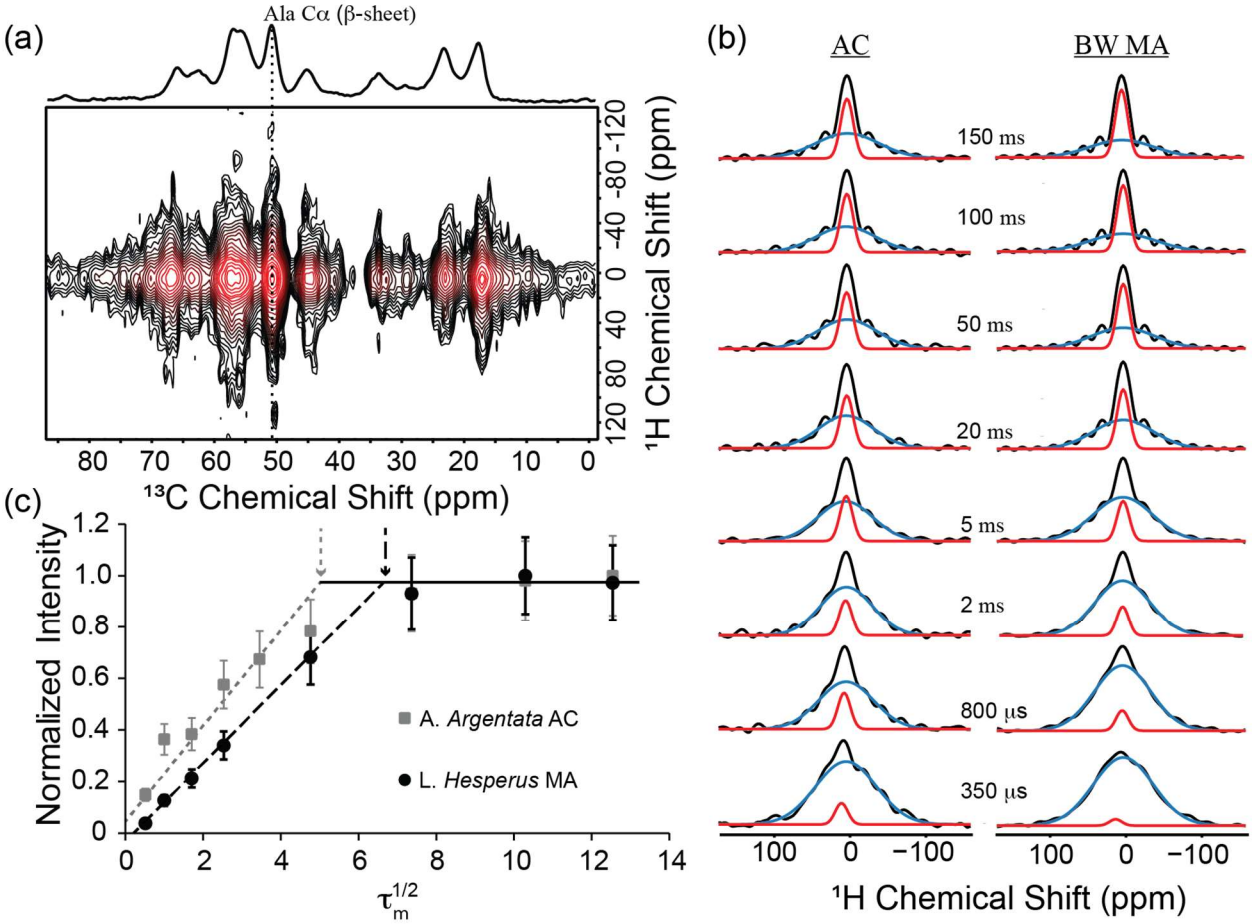



**Figure 81.** (**a**) A 2D $^{1}$H/$^{13}$C WISE spectrum with a 50 ms spin-diffusion time for hydrated *A. argentata* AC silk. (**b**), deconvoluted fits of $^{1}$H slices extracted at the Ala Cα β-sheet resonance in the WISE spectrum for AC silk and BW dragline. (**c**), build-up curves of narrow component (red) from deconvolutions for the two silks. The arrows indicate $\tau_m^*$ for each sample.

biopolymer conformational structure with mobility and dynamics. A 1D $^{1}H$ slice is typically extracted from the 2D $^{1}H/^{13}C$ WISE SSNMR spectrum to reveal the $^{1}H$ line shape at a specific $^{13}C$ chemical shift where broad $^{1}H$ line shapes are associated with “rigid” $^{1}H$ environments, while narrow $^{1}H$ profiles would arise from more "mobile" regions.

Importantly, if a $^{1}H$ spin-diffusion period is included in the WISE experiment it is possible to observe magnetization from the "mobile" water-plasticized regions diffuse into these water-inaccessible domains which ultimately allows for domain-size estimates of the rigid structures.[45] Here we apply the WISE technique to both Black Widow (BW) dragline silk fibers and cross-linked AC silks and compared the respective β-sheet domain sizes and overall fiber rigidity. Work by Holland (2004) was followed[48]; we collected 2D WISE data on $D_2O$-hydrated isotopically-enriched silks at a series of increasing mixing times and observed the associated $^{1}H$ profile of the Ala Cα β-sheet environment.[15] $D_2O$ was used as the plasticizing solvent over $H_2O$ so that the narrow component can be properly attributed to silk polymer protons and not water. The extracted $^{1}H$ profiles were fit to both broad and narrow components for each mixing time, and the area of the narrow component was plotted against the square root of the spin-diffusion mixing time.[45] The point ($\tau_m^*$) at which the narrow component reaches a plateau implies that diffusion of magnetization from the "mobile" domain into the "rigid" domain has reached an equilibrium. Larger domains would require more time to reach equilibrium. This technique has successfully been applied to determine the domain sizes in a number of heterogeneous materials.[15, 45, 49]

WISE experiments were collected on AC silk and BW dragline fibers using spin-diffusion times between 0.05-150 ms (**Figure 81**a). Slices were extracted from WISE experiments at the Ala Cα β-sheet resonance and fit to a broad and narrow component (**Figure 81b**). We found the following major conclusions from the WISE buildup data on cross-linked AC silk fibers; 1) a large fraction of Ala residues adopting β-sheet secondary structures within AC silks are inherently water accessible, 2) Gly residues are more dynamic within AC silks compared to BW dragline silks, and

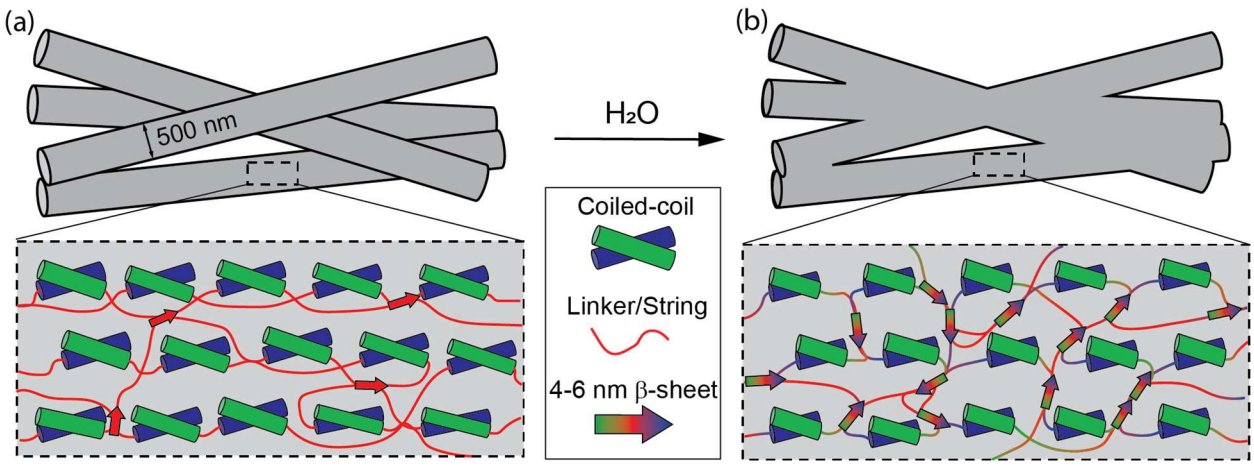



**Figure 82.** Proposed model of AC silks after contact with water. Helices 1-5 are blue and green cylinders that form α-helical coiled-coils. (**a**), before wetting there is little β-sheet content (~15%). (**b**), after water wetting some α-helical and RC structures convert to β-sheet, illustrated as shortened coiled-coils and the multi-colored arrows colored according to their origin from RC and/or helix structures.

3) water-inaccessible β-sheet nanodomains are significantly smaller in cross-linked AC silks compared to well-characterized BW spider dragline β-sheet nanocrystallites. The first conclusion is clear from the higher initial intensity of the narrow component for AC relative to dragline silk. Even at short (**Figure 81b**, 50 and 500 μs) $^{1}$H spin-diffusion times, the $^{1}$H profile of Ala Cα β-sheet environment from PW silk shows a significantly larger mobile component compared to the corresponding mobile component of the BW dragline silk. This observation implies that even when Ala adopts β-sheet secondary structures in water-hydrated AC silks, a larger fraction of these structures are inherently water-accessible relative to those within BW dragline fibers. Similarly, based on the high initial offset and quick rise to equilibrium, it is clear that Gly residues within AC silks are highly mobile even after water-induced crosslinking (**Figure 83**). Third, analysis of the Ala Cα spin-diffusion buildup curves for water-hydrated AC and BW dragline silks suggest that the Ala-rich β-sheet domains in cross-linked AC silks are smaller than in BW dragline fibers. We determined the $\tau_m^*$ for AC and BW dragline fibers to be 4.5 and 6.6 ms, respectively (**Figure 81c**). Using the previously-derived equation to calculate domain size, we determine the β-sheet domains in cross-linked AC silks to be between 4-6 nm (detail in supporting information).

Additionally, we calculated the size of the nanodomains in BW dragline silk to be approximately 11 nm, somewhat larger than reported by XRD results which show a crystallite size of 3.02 x 4.15 x 6.71 nm.[16] The difference in size could be due to the well-diffracting, truly crystalline areas of these nanodomains which may be smaller than the water-inaccessible domains we probe with NMR. Obviously, it was not possible to measure the initial domain sizes of native AC silks because the technique inherently requires water-hydrating and shearing under MAS. Since overall β-sheet content increases upon water contact and shearing, the original β-sheet nanodomains in freshly-spun wrapping silks are either smaller than 3 nm and increase in size upon wetting (unlikely), or more likely there is an increased abundance of small-sized β-sheet domains after water-induced crosslinking.

**Conclusion**

In summary, we have shown that prey-wrapping silks from the garden spider *A. argentata* undergo a remarkable cross-linking behavior when exposed to moisture that has not been observed in other silks to the best of our knowledge. Upon contact with water, SEM images reveal that individual silk fibers fuse together to form cross-linked fibrous sheets and mats. Comprehensive solid-state NMR data on $^{13}$C-enriched wrapping silks has uncovered the molecular mechanism of inter-fiber cross-linking. Orthogonal isotopic enrichment schemes (Ala-labeling and Val-labeling) were chosen to understand these structural changes for Ala, Ser, Gly, and Val residues, shedding light on which regions of the primary protein sequence are most susceptible to water-induced β-sheet crosslinking. Overall, the data reveals that water-induced fiber cross-linking is driven by an increase in β-sheet protein secondary structure at the expense of disordered and loosely-structured α-helical motifs, whereas well-ordered coiled-coil structures remain in-tact for the most part. However, helical coiled-coil motifs also undergo partial α-to-β conversion with the addition of mechanical shear. This behavior is also seen in other structural protein-based biopolymers. For example, keratin has been shown to convert from α-to-β when

stretched in the presence of moisture.[50, 51] Finally, WISE NMR data was used to estimate the β-sheet domain sizes and overall protein dynamics within water-hydrated wrapping silks. Even after water-induced crosslinking, β-sheet regions within AC silks are smaller (~4-6 nm) than those found in BW dragline fibers (~11 nm), suggesting that newly-formed β-sheet structures are small in size and likely between two or more protein chains and potentially involving multiple fibers. Figure 82 shows a proposed model to illustrate the structural changes of AC silks after water treatment. This data supports our prior conclusions that at a minimum there are three distinct domains in wrapping silks: (1) disordered regions that are easily plasticized by water, (2) rigid α-helices from highly-stable coiled-coil motifs, and (3) Ala and Ser-rich β-sheets are present but surprisingly minimal. Future work will further divulge the role of water in this process, turning a flexible silk fiber into a hardened matted sheet.

It is anticipated that the concept of utilizing water treatment to convert protein-based biomaterials from flexible and extensible, dominated by α-helical coiled-coil hierarchy, to rigid crossed-linked β-sheet assemblies, should provide new avenues in bioinspired material design. The potential for this hydration-induced conformational switching behavior in a biomaterial are hypothesized to impact a broad range of fields including defense, biomedical and space exploration that all require ever evolving functional materials. In the biomedical space new biomaterials for micro, nano and even molecular suturing could be possible for tissue repair while, in the defense and space exploration sectors new novel materials based on the aciniform silk system could be utilized in advanced textiles for protective clothing, tents, parachutes and the repairs of such materials.

**Methods**

*Sample Preparation: A. Argentata* spiders were fed aqueous solutions of isotopically enriched amino acid containing either U-[$^{13}C/^{15}N$]-Ala or U-[$^{13}C/^{15}N$-Val]. Ala labeling also isotopically enriches Ser, Gly, Gln through the metabolism of Ala and has the added advantage

of labeling these other amino acids for NMR structural characterization.[52] These orthogonal labeling schemes were chosen because Val is almost exclusively found in the helix-rich globular "bead" domain, Ala and Ser residues are dispersed throughout both bead and "linker / string", Gly is heavily represented in the disordered linker / string region, and Gln is fairly evenly distributed.

*Scanning Electron Microscopy:* Spider silk samples were imaged with a FEI Quana 450 FEG Scanning Electron Microscope (SEM) in the Electron Microscope Facility at San Diego State University, using a 10 kV accelerating voltage. Samples were coated with 8 nm of platinum prior to imaging using an EMS 150 Sputter Coater.

*Force-displacement AFM (Nano-indentation):* The silk fibers were mounted on glass substrates using sodium silicate and interrogated using AFM Workshop TT-2 AFM operated in force-displacement mode at various locations along the length of the fiber with a Bruker RTESPA-525 silicon indenter (k=203 N/m, f0 = 535 kHz). A maximum of 14-20 mN of force was applied at each location, while the indentation depth was measured. The collected data was analyzed using the Oliver and Pharr model[53] according to **Equation 1**:

$$E = \frac{E_i \frac{0.18S^2}{h_{max}S - \varepsilon P_{max}}(1-\nu^2)}{E_i - \frac{0.18S^2}{h_{max}S - \varepsilon P_{max}}(1-\nu_i{}^2)} \quad (1)$$

where, $\mathrm{E_i}$ is the modulus and $\nu_\mathrm{i}$ is the Poisson's ratio of the indenter, which were taken to be 170 GPa and 0.27, respectively, $\mathrm{E}$ is the modulus of the material and $\nu$ is the Poisson's ratio of the material, $\mathrm{S}$ is the slope of the load-elastic displacement curve, $\varepsilon$ is a constant determined based on the geometry of the indenter, and $\mathrm{P_{max}}$ and $\mathrm{h_{max}}$ is the maximum load applied and maximum indentation depth, respectively. As shown in the model, the resulting material stiffness is dependent on the value of the Poisson's ratio.

*Solid-state NMR (SSNMR):* SSNMR experiments were collected with a 600 MHz Bruker AVANCE III HD spectrometer equipped with a 1.9 mm HCN MAS probe. 1D $^{13}$C CP-MAS NMR experiments were collected at 30 kHz MAS and used a 2.45 $\mu$s 90° pulse, 24 ms acquisition time and 3-5 s recycle delay. In order to correctly assign chemical shifts and choose linewidths for

peak fitting of 1D data, we extracted these parameters from 2D dipolar-assisted rotational resonance (DARR) experiments. DARR is a 2D homonuclear spin-diffusion technique that correlates nearby $^{13}C$-$^{13}C$ nuclei.[54] The data is presented in 2D, where the off-diagonal peaks correspond to through-space $^{13}C$-$^{13}C$ interactions. By adjusting the mixing time, one can allow spin-diffusion to occur for longer periods and thus probe longer distances. 2D DARR experiments were collected at 14 kHz MAS with 256 points in the f1 dimension with 256 scans, a mixing time of 100 ms and a recycle delay of 2.5 s. During DARR mixing periods, continuous wave $^{1}H$ irradiation was applied at a rotary-resonance ($\omega_r = \omega_{rf}$) of 14 kHz.

WISE experiments were collected at 5 kHz MAS with a 300 μs CP contact time and 5 s recycle delay. All spectra were referenced to DSS at 0.0 ppm by setting the downfield adamantane resonance to 40.49 ppm.[55] Samples were soaked in $D_2O$ before experiments for 2 hrs and center packed with Teflon inserts. Equation 2 below is used to calculate the β-sheet domain size, $d_{dis}$, for the silks:

$$d_{dis} = \frac{\rho_A^H \phi_A + \rho_B^H \phi_B}{\phi_A \phi_B} \frac{\sqrt{D_A D_B}}{\rho_A^H \sqrt{D_A} + \rho_B^H \sqrt{D_B}} \frac{4\varepsilon \phi_{dis}}{\sqrt{\pi}} \sqrt{\tau_m^*} \quad (2)$$

where $\rho^H$, $\phi_{dis}$, $D$ and $\varepsilon$ are the proton density of the two domains, volume fraction of the two domains, $^{1}H$ spin-diffusion coefficients for both domains, and dimensionality, respectively.[45] A summary of the coefficient values used for the calculations (Table 11) and further explanation is in the Supporting Information.

**Supplemental Figures**

WISE Calculations for Domain Size. Equation 1 below is used to calculate the β-sheet domain size, $d_{dis}$, for the silks:

$$d_{dis} = \frac{\rho_A^H \phi_A + \rho_B^H \phi_B}{\phi_A \phi_B} \frac{\sqrt{D_A D_B}}{\rho_A^H \sqrt{D_A} + \rho_B^H \sqrt{D_B}} \frac{4\varepsilon \phi_{dis}}{\sqrt{\pi}} \sqrt{\tau_m^*} \quad (1)$$

where $\rho^H$, $\phi_{dis}$, $D$ and $\varepsilon$ are the proton density of the two domains, volume fraction of the two domains, $^1$H spin-diffusion coefficients for both domains, and dimensionality, respectively.[46] The value of $t_m^*$ for all silks was extracted from the build-up curve of the narrow component of the data by fitting the Ala Cα resonance with a broad and narrow component (**Figure 84**). The linewidths in for BW were 47 kHz for the broad, and 5.2 kHz for the narrow. In AC fibers, the linewidths were 45 kHz for the broad, and 5 kHz for the narrow. We assume the proton densities for AC β-sheets are similar to those in BW dragline. The proton densities for β-sheets domains and disordered regions in dragline are 0.107 and 0.078 g/cm$^3$, respectively. The spin-diffusion coefficient can be determined by the $^1$H linewidth ($\Delta\nu_{1/2}$) and average minimum proton distance ($r_{HH}$) with the following proportionality: $D \sim \Delta\nu_{1/2}(r_{HH})^2$.[46] The spin-diffusion of PS-PMMA is known to be ~0.8 nm$^2$/ms with a $^1$H linewidth of 38 kHz.[46] Using this proportionality, we estimate the spin-diffusion coefficients for the dynamic and rigid regions of water-treated AC silks to be 0.11 and 0.95 nm$^2$/ms, respectively. The crystalline domain volume of BW dragline fibers were recently determined by X-ray scattering and reported to be 40%.[16] Because there is no XRD data reported for AC silk the morphology of the β-sheet regions is unknown. However, we assume spin-diffusion can occur into the domains predominantly on four faces, giving it a dimensionality of 2 ($\phi_{dis} = 2$)

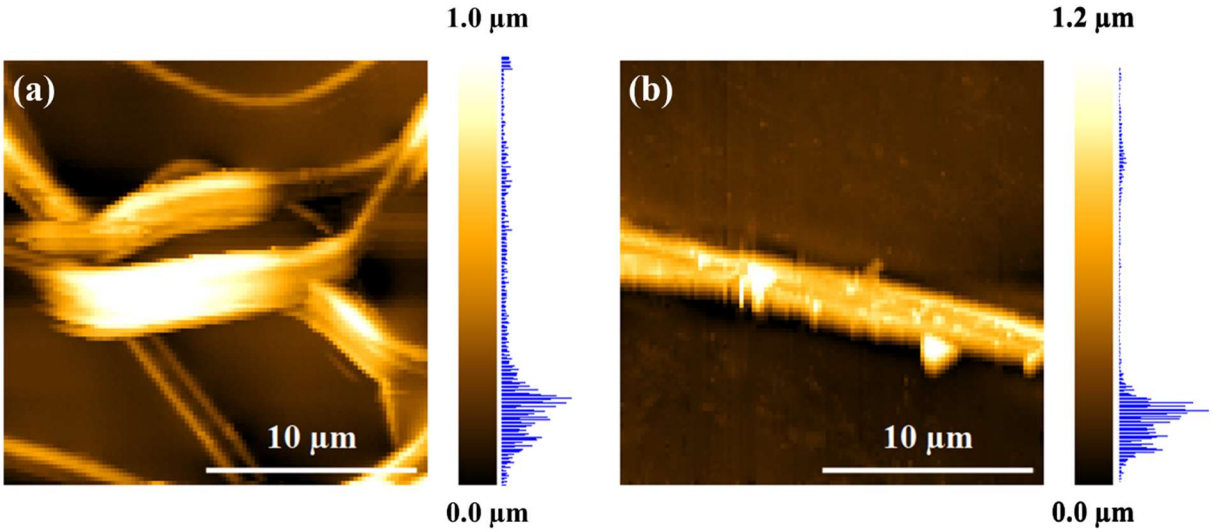



**Figure 83.** AFM topography scans of (a) native prey-wrapping silk and (b) silk water-wetted for 2 hr and dried for 2 days.

based on the location of the β-sheet regions with respect to the rest of the matrix.(REF) The β-sheet fraction for native, as-spun AC fibers was determined from SSNMR in our recent work and is reported to be ~15%.[28] However, we acknowledge that during the WISE experiment the silk may become sheared in the presence of $D_2O$ which could have an impact on the β-sheet content. In the wet-sheared and dried sample, we see an increase in β-sheet from Ala, Val and Ser although deconvolution of Ser is difficult because of spectral overlap as discussed in the main text. Using an approach similar to our previous work, we use the primary sequence and secondary structure for Ala, Val and Ser to roughly assign total β-sheet content,[28] and estimate an upper-limit of the β-sheet fraction to be 25% in the wet-sheared and dried material. With this in mind, the calculated size of these crystallites using a β-sheet fraction between 15-25% produces a range of nanocrystallite sizes between 4.4-6.4 nm. Table 11 summarizes the parameters used for calculating the domain sizes in both Black Widow (*L. hesperus*) dragline MA and *A. argentata* AC silks.

**Table 11.** Summary of parameters used to calculate domain size in BW dragline and AC silks.

| Parameter | Major Ampullate | Aciniform (lower limit) | Aciniform (upper limit) |
|---|---|---|---|
| $\rho_A^H$ (g/cm$^2$) | 0.107 | 0.107 | 0.107 |
| $\rho_B^H$(g/cm$^2$) | 0.078 | 0.078 | 0.078 |
| $\phi_A$ | 0.40 | **0.15** | **0.25** |
| $\phi_B$ | 0.60 | **0.85** | **0.75** |
| $D_A$ (nm$^2$/ms) | 0.99 | 0.95 | 0.95 |
| $D_B$ (nm$^2$/ms) | 0.11 | 0.11 | 0.11 |
| ε | 2 | 2 | 2 |
| $\sqrt{\tau_m^*}$ | 6.6 | 4.5 | 4.5 |
| $\sqrt{\tau_m^*}$ Error | ± 0.3 | ± 0.5 | ± 0.5 |
| $d_{dis}$ **(nm)** | **11.0 ± 0.6** | **4.9 ± 0.5** | **5.8 ± 0.6** |

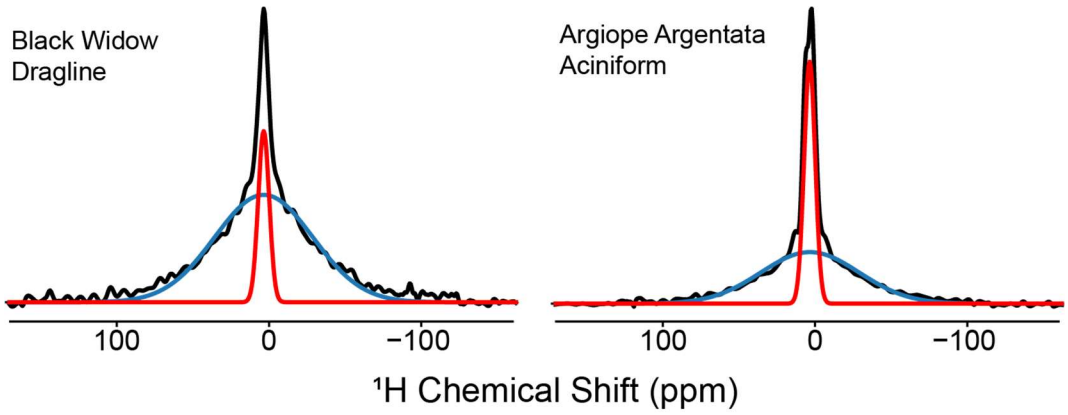



**Figure 84.** $^1$H slice extracted at Ala Cα β-sheet resonance in WISE spectra collected with 50 ms spin-diffusion mixing times for the two silks. Clearly the AC silks exhibit larger narrow component indicating mobile magnetization is able to spin diffuse into Ala β–sheet domains more rapidly because of smaller nanocrystallite dimensions compared to Black Widow (*L. hesperus*) dragline MA fibers.

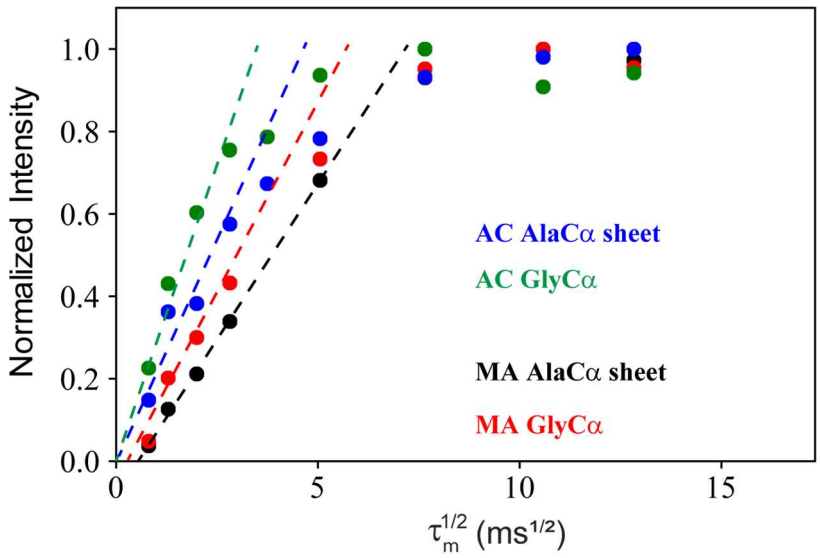



**Figure 85.** WISE build-up curves *for A. argentata* AC and Black Widow (*L. hesperus*) MA silk fibers. The first four points of the initial Gly build-up are fit with a dashed line of the corresponding color and the reported buildup time ($\sqrt{\tau_m^*}$) is listed next to the label. Spin-diffusion occurs rapidly into the Ala Cα and Gly Cα areas of AC silks suggesting they are smaller than in the BW silk fibers. Table 12 has these values.

**Table 12.** Summary of parameters used to calculate domain size in *L. hesperus* MA silks and *A. argentata* AC silks for both Gly Cα and Ala Cα.

| **Parameter** | **MA Ala Cα** | **AC Ala Cα (LL)** | **AC Ala Cα (UL)** | **MA Gly Cα** | **AC Gly Cα (LL)** | **AC Gly Cα (UL)** |
|---|---|---|---|---|---|---|
| $\rho_A^H$ (g/cm$^2$) | 0.107 | 0.107 | 0.107 | 0.107 | 0.107 | 0.107 |
| $\rho_B^H$(g/cm$^2$) | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 | 0.078 |
| $\phi_A$ | 0.40 | **0.15** | **0.25** | 0.40 | **0.15** | **0.25** |
| $\phi_B$ | 0.60 | **0.85** | **0.75** | 0.60 | **0.85** | **0.75** |
| $D_A$ (nm$^2$/ms) | 0.99 | 0.95 | 0.95 | 0.99 | 0.95 | 0.95 |
| $D_B$ (nm$^2$/ms) | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 | 0.11 |
| ε | 2 | 2 | 2 | 2 | 2 | 2 |
| $\sqrt{\tau_m^*}$ | 6.6 | 4.5 | 4.5 | 5.1 | 2.9 | 2.9 |
| $\sqrt{\tau_m^*}$ Error | ± 0.3 | ± 0.5 | ± 0.5 | ± 0.5 | ± 0.3 | ± 0.5 |
| $d_{dis}$ **(nm)** | **11.0 ± 0.6** | **4.9 ± 0.5** | **5.8 ± 0.6** | **8.6 ± 0.8** | **3.1 ± 0.4** | **3.7 ± 0.6** |

LL: Lower limit, UL: Upper-limit

The results presented above illustrate that narrow $^1$H components (disordered, non-β-sheet) spin diffuse more rapidly into broad components for Gly compared to Ala. This is the case for both AC

and MA silks albeit the length scale are shorter for AC silk. In both cases, this implicates that Gly is spatially located at the periphery of the β-sheet (interphase). For, MA silks the interpretation would be that the poly(Ala) is the core of the β-sheet and the flanking Gly-Ala units are β-sheet but located at the β-sheet interface thus, mobile regions in the disordered (amorphous) phase are closer to Gly than Ala. It appears to be something similar in AC silk where for the most part Ala is located in the β-sheet core with Gly-containing units closer to the β-sheet interface with the more disordered or α-helical regions. It should be noted that this is speculative for AC silk however, the result is consistent with our previous studies for MA silk.[15]

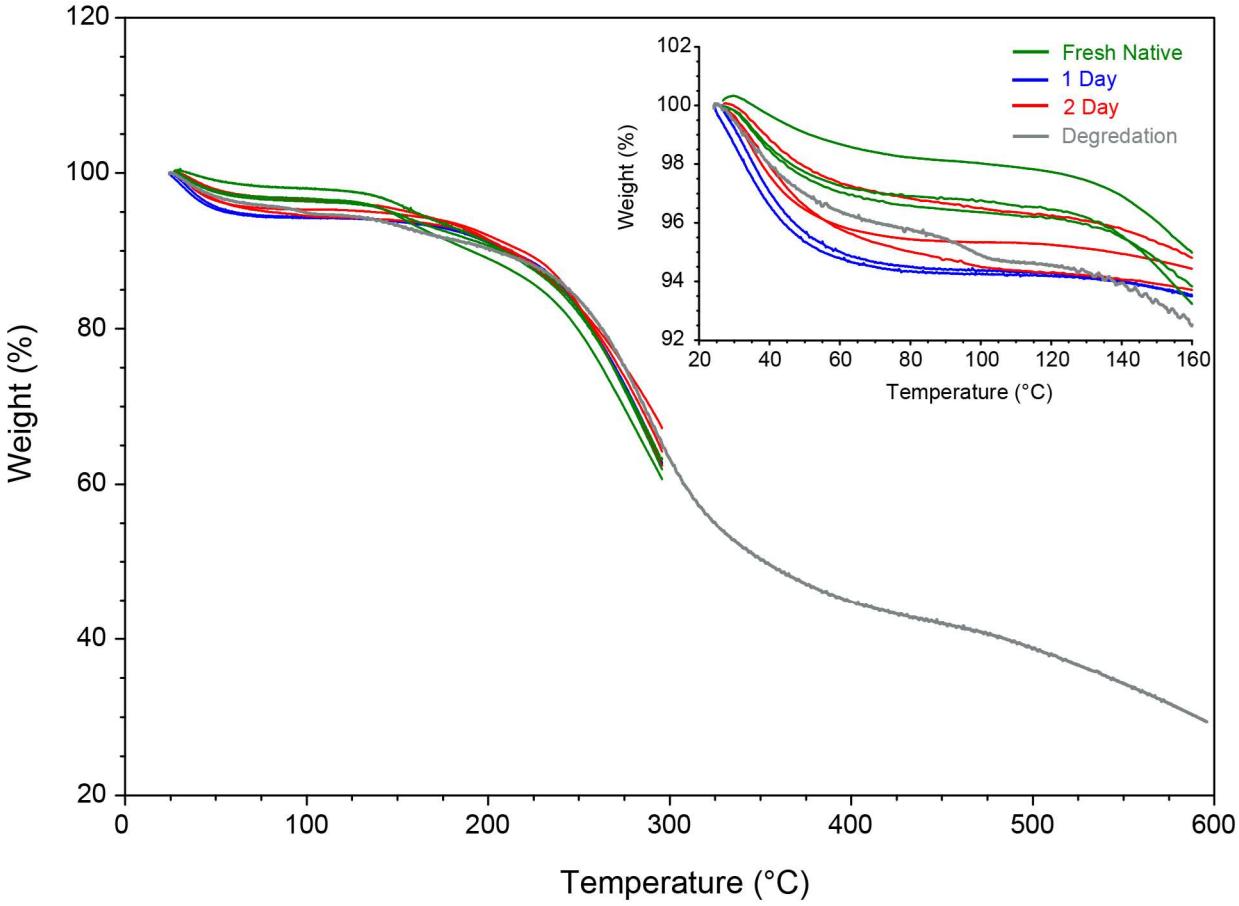



**Figure 86.** Thermogravimetric analysis (TGA) of *A. argentata* AC silks. Native silks were collected and run within 5 minutes of collection. Water loss is reported in Table 13 from the weight loss curve up to 120°C. The data was collected with a TA2910 (TA Instruments Inc.) instrument under a steady nitrogen flow (60 mL/min for furnace and 40 mL/min for balance). Prior to analysis, the sample was kept under $N_2$ flow for 30 minutes to remove weakly-bound, physiosorbed water and obtain a stable baseline.

**Table 13.** Calculated water loss for prey-wrapping silk samples from Figure 86, their average and standard deviation. Native samples were run within 5 minutes of collection. AC silk samples were soaked in DI water for 2 hours and dried at 25°C at a humidity of 30-35% for 1 or 2 days.

| Sample | Weight Loss (%) | Water Content Average ± SD |
|---|---|---|
| Native 1 | 2.359 | 3.2 ± 0.6 |
| Native 2 | 3.778 | |
| Native 3 | 3.413 | |
| 1 Day, 1 | 5.704 | 5.76 ± 0.05 |
| 1 Day, 2 | 5.813 | |
| 2 Day, 1 | 4.729 | 4.7 ± 0.7 |
| 2 Day, 2 | 5.645 | |
| 2 Day, 3 | 3.825 | |

**Acknowledgements**

This research was supported by the US Department of Defense, Air Force Office of Research (DOD-AFOSR) under award numbers FA9550-17-1-0282 and FA9550-20-1-103 and Army Research Office (DOD-ARO), contract number W911NF2010143 (G.P.H.).

Chapter 7 is a full reprint of “Hydration-Induced β-Sheet Crosslinking of α-Helical-Rich Spider Prey-Wrapping Silk” published in Advanced Function Materials in 2021. Dillan Stengel, J. Bennett Addison, David Onofrei, Nha Uyen Huynh, George Youssef, and Gregory P. Holland. Dillan Stengel was the primary author of this paper.

**References**

1. Blackledge, T. A.; Perez-Rigueiro, J.; Plaza, G. R.; Perea, B.; Navarro, A.; Guinea, G. V.; Elices, M., Sequential origin in the high performance properties of orb spider dragline silk. *Sci Rep* **2012,** *2*, 782.

2. Vollrath, F.; Knight, D. P., Liquid crystalline spinning of spider silk. *Nature* **2001,** *410* (6828), 541-8.

3. Vepari, C.; Kaplan, D. L., Silk as a Biomaterial. *Prog Polym Sci* **2007,** *32* (8-9), 991-1007.

4. Aigner, T. B.; DeSimone, E.; Scheibel, T., Biomedical Applications of Recombinant Silk-Based Materials. *Adv Mater* **2018,** *30* (19), e1704636.

5. Andersson, M.; Jia, Q.; Abella, A.; Lee, X. Y.; Landreh, M.; Purhonen, P.; Hebert, H.; Tenje, M.; Robinson, C. V.; Meng, Q.; Plaza, G. R.; Johansson, J.; Rising, A., Biomimetic spinning of artificial spider silk from a chimeric minispidroin. *Nat Chem Biol* **2017,** *13* (3), 262-264.

6. Vasanthavada, K.; Hu, X.; Falick, A. M.; La Mattina, C.; Moore, A. M.; Jones, P. R.; Yee, R.; Reza, R.; Tuton, T.; Vierra, C., Aciniform spidroin, a constituent of egg case sacs and wrapping silk fibers from the black widow spider Latrodectus hesperus. *J Biol Chem* **2007,** *282* (48), 35088-97.

7. Romer, L.; Scheibel, T., The elaborate structure of spider silk: structure and function of a natural high performance fiber. *Prion* **2008,** *2* (4), 154-61.

8. Yarger, J. L.; Cherry, B. R.; van der Vaart, A., Uncovering the structure–function relationship in spider silk. *Nature Reviews Materials* **2018,** *3*.

9. Hayashi, C. Y.; Blackledge, T. A.; Lewis, R. V., Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. *Mol Biol Evol* **2004,** *21* (10), 1950-9.

10. Blackledge, T. A.; Hayashi, C. Y., Silken toolkits: biomechanics of silk fibers spun by the orb web spider Argiope argentata (Fabricius 1775). *J Exp Biol* **2006,** *209* (Pt 13), 2452-61.

11. Xu, M.; Lewis, R. V., Structure of a protein superfiber: Spider dragline silk. *Proc Natl Acad Sci U S A* **1990,** *87*, 7120-7124.

12. Ayoub, N. A.; Garb, J. E.; Tinghitella, R. M.; Collin, M. A.; Hayashi, C. Y., Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One* **2007,** *2* (6), e514.

13. Xu, D.; Yarger, J. L.; Holland, G. P., Exploring the backbone dynamics of native spider silk proteins in Black Widow silk glands with solution-state NMR spectroscopy. *Polymer* **2014,** *55* (16), 3879-85.

14. Xu, D.; Shi, X.; Thompson, F.; Weber, W. S.; Mou, Q.; Yarger, J. L., Protein secondary structure of Green Lynx spider dragline silk investigated by solid-state NMR and X-ray diffraction. *Int J Biol Macromol* **2015,** *81*, 171-9.

15. Holland, G. P.; Lewis, R. V.; Yarger, J. L., WISE NMR Characterization of Nanoscale Heterogeneity and Mobility in Supercontracted Nephila clavipes Spider Dragline Silk. *JACS* **2004,** *126*.

16. Jenkins, J. E.; Sampath, S.; Butler, E.; Kim, J.; Henning, R. W.; Holland, G. P.; Yarger, J. L., Characterizing the Secondary Protein Structure of Black Widow Dragline Silk Using Solid-State NMR and X-ray Diffraction. *Biomacromolecules* **2013,** *14*, 3472-83.

17. Yang, Z.; Grubb, D. T.; Jelinski, L. W., Small-Angle X-ray Scattering of Spider Dragline Silk. *Macromolecules* **1997,** *30*, 8254-61.

18. Riekel, C.; Branden, C.; Craig, C.; Ferrero, C.; Heidelbach, F.; Muller, M., Aspects of X-ray diffraction on single spider fibers. *Biological Macromolecules* **1999,** *24*, 179-186.

19. Gray, G. M.; van der Vaart, A.; Guo, C.; Jones, J.; Onofrei, D.; Cherry, B. R.; Lewis, R. V.; Yarger, J. L.; Holland, G. P., Secondary Structure Adopted by the Gly-Gly-X Repetitive Regions of Dragline Spider Silk. *Int J Mol Sci* **2016,** *17* (12).

20. Chaw, R. C.; Zhao, Y.; W, J.; Ayoub, N. A.; Allen, R.; Atrushi, K.; Hayashi, C. Y., Intragenic homogenization and multiple copies of prey-wrapping silk genes in Argiope garden spiders. *BMC Envolutionary Biology* **2014,** *14* (31).

21. Sarker, M.; Orrell, K. E.; Xu, L.; Tremblay, M. L.; Bak, J. J.; Liu, X. Q.; Rainey, J. K., Tracking Transitions in Spider Wrapping Silk Conformation and Dynamics by $^{19}F$ Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **2016,** *55* (21), 3048-59.

22. Tremblay, M. L.; Xu, L.; Lefevre, T.; Sarker, M.; Orrell, K. E.; Leclerc, J.; Meng, Q.; Pezolet, M.; Auger, M.; Liu, X. Q.; Rainey, J. K., Spider wrapping silk fibre architecture arising from its modular soluble protein precursor. *Sci Rep* **2015,** *5*.

23. Xu, L.; Tremblay, M. L.; Meng, Q.; Liu, X. Q.; Rainey, J. K., $^{1}H$, $^{13}C$ and $^{15}N$ NMR assignments of the aciniform spidroin (AcSp1) repetitive domain of Argiope trifasciata wrapping silk. *Biomol NMR Assign* **2012,** *6*, 147-51.

24. Tremblay, M. L.; Xu, L.; Sarker, M.; Liu, X. Q.; Rainey, J. K., Characterizing Aciniform Silk Repetitive Domain Backbone Dynamics and Hydrodynamic Modularity. *Int J Mol Sci* **2016,** *17* (1305).

25. Xu, L.; Weatherbee-Martin, N.; Liu, X. Q.; Rainey, J. K., Recombinant Silk Fiber Properties Correlate to Prefibrillar Self-Assembly. *Small* **2019,** *15* (12), e1805294.

26. Xu, L.; Lefevre, T.; Orrell, K. E.; Meng, Q.; Auger, M.; Liu, X. Q.; Rainey, J. K., Structural and Mechanical Roles for the C-Terminal Nonrepetitive Domain Become Apparent in Recombinant Spider Aciniform Silk. *Biomacromolecules* **2017,** *18* (11), 3678-86.

27. Xu, L.; Rainey, J. K.; Meng, Q.; Liu, X. Q., Recombinant minimalist spider wrapping silk proteins capable of native-like fiber formation. *PLoS One* **2012,** *7* (11), e50227.

28. Addison, B.; Onofrei, D.; Stengel, D.; Blass, B.; Brenneman, B.; Ayon, J.; Holland, G. P., Spider prey-wrapping silk is an alpha-helical coiled-coil/beta-sheet hybrid nanofiber. *Chem Commun (Camb)* **2018,** *54* (76), 10746-10749.

29. Rousseau, M. E.; Lefevre, T.; Pezolet, M., Conformation and Orientation of Proteins in Various Types of Silk Fibers Produced by Nephila clavipes Spiders. *Biomacromolecules* **2009,** *10*, 2945–2953.

30. Lefevre, T.; Boudreault, S.; Cloutier, C.; Pezolet, M., Diversity of molecular transformations involved in the formation of spider silks. *J Mol Biol* **2011,** *405* (1), 238-53.

31. Asakura, T.; Endo, M.; Tasei, Y.; Ohkubo, T.; Hiraoki, T., Hydration of Bombyx mori silk cocoon, silk sericin and silk fibroin and their interactions with water as studied by (13)C NMR and (2)H NMR relaxation. *J Mater Chem B* **2017,** *5* (8), 1624-1632.

32. Lawrence, B. D.; Wharram, S.; Kluge, J. A.; Leisk, G. G.; Omenetto, F. G.; Rosenblatt, M. I.; Kaplan, D. L., Effect of hydration on silk film material properties. *Macromol Biosci* **2010,** *10* (4), 393-403.

33. Yazawa, K.; Ishida, K.; Masunaga, H.; Hikima, T.; Numata, K., Influence of Water Content on the beta-Sheet Formation, Thermal Stability, Water Removal, and Mechanical Properties of Silk Materials. *Biomacromolecules* **2016,** *17* (3), 1057-66.

34. Asakura, T.; Ohgo, K.; Ishida, T.; Taddei, P.; Monti, P.; Kishore, R., Possible Implications of Serine and Tyrosine Residues and Intermolecular Interactions on the Appearance of Silk I Structure of Bombyx mori Silk Fibroin-Derived Synthetic Peptides: High-Resolution 13C Cross-Polarization/Magic-Angle Spinning NMR Study. *Biomacromolecules* **2005,** *6*, 468-474.

35. Koski, K. J.; Akhenblit, P.; McKiernan, K.; Yarger, J. L., Non-invasive determination of the complete elastic moduli of spider silks. *Nat Mater* **2013,** *12* (3), 262-7.

36. Robinson, M. H., Predatory Behavior of Argiope argentata (Fabricius). *Am. Zoologist* **1969,** *9*, 161-173.

37. Eisner, T.; Dean, J., Ploy and counterploy in predator-prey interactions: Orb-weaving spiders versus bombardier beetles. *Proc Natl Acad Sci U S A* **1976,** *73* (4), 1365-1367.

38. Eberhard, W. G.; Barrantes, G.; Weng, J. L., The mystery of how spiders extract food without masticating prey. *Bull. Br. arachnol. Soc.* **2006,** *13* (9), 372–376.

39. Asakura, T.; Matsuda, H.; Aoki, A.; Kataoka, N.; Imai, A., Conformational change of $^{13}C$-labeled 47-mer model peptides of Nephila clavipes dragline silk in poly(vinyl alcohol) film by stretching studied by $^{13}C$ solid-state NMR and molecular dynamics simulation. *Int J Biol Macromol* **2019,** *131*, 654-665.

40. Holland, G. P.; Jenkins, J. E.; Creager, M. S.; Lewis, R. V.; Yarger, J. L., Quantifying the fraction of glycine and alanine in beta-sheet and helical conformations in spider dragline silk using solid-state NMR. *Chem Commun (Camb)* **2008,** (43), 5568-70.

41. Addison, J. B.; Popp, T. M.; Weber, W. S.; Edgerly, J. S.; Holland, G. P.; Yarger, J. L., Structural characterization of nanofiber silk produced by embiopterans (webspinners). *RSC Adv* **2014,** *4* (78), 41301-41313.

42. Addison, J. B.; Ashton, N. N.; Weber, W. S.; Stewart, R. J.; Holland, G. P.; Yarger, J. L., b-Sheet Nanocrystalline Domains Formed From Phosphorylated Serine-rich Motifs in Caddisfly Larval Silk: A Solid State NMR and XRD Study. *Biomacromolecules* **2013,** *14* (4), 1140-8.

43. Correa-Garhwal, S. M.; Clarke, T. H., 3rd; Janssen, M.; Crevecoeur, L.; McQuillan, B. N.; Simpson, A. H.; Vink, C. J.; Hayashi, C. Y., Spidroins and Silk Fibers of Aquatic Spiders. *Sci Rep* **2019,** *9* (1), 13656.

44. Hong, M.; Schmidt-Rohr, K., Magic-Angle-Spinning NMR Techniques for Measuring Long-Range Distances in Biological Macromolecules. *Acc Chem Res* **2013,** *46*, 2154-2163.

45. Clauss, J.; Schmidt-Rohr, K.; Spiess, H. W., Determination of domain sizes in heterogeneous polymers by solid-state NMR. *Acta Polymer* **1993,** *44*, 1-17.

46. Schmidt-Rohr, K.; Clause, J.; Spiess, H. W., Correlation of Structure, Mobility, and Morphological Information in Heterogeneous Polymer Materials by Two-Dimensional Wideline-Separation NMR Spectroscopy. *Macromolecules* **1992,** *25* (12), 3273-3277.

47. K. Schmidt-Rohr, H. W. S., *Multidimensional Solid-State NMR and Polymers*. 1 ed.; Academic Press: London, 1994.

48. Holland, G. P.; Lewis, R. V.; Yarger, J. L., WISE NMR characterization of nanoscale heterogeneity and mobility in supercontracted Nephila clavipes spider dragline silk. *J Am Chem Soc* **2004,** *126* (18), 5867-72.

49. Yan, B.; Stark, R. E., A WISE NMR Approach to Heterogeneous Biopolymer Mixtures: Dynamics and Domains in Wounded Potato Tissues. *Macromolecules* **1998,** *31*, 2600-2605.

50. Kreplak, L.; Franbourg, A.; Briki, F.; Leroy, F.; Dalle, D.; Doucet, J., A New Deformation Model of Hard α-Keratin Fibers at the Nanometer Scale: Implications for Hard α-Keratin Intermediate Filament Mechanical Properties. *Biophysical Journal* **2002,** *82*, 2265-2274.

51. Kreplak, L.; Doucet, J.; Dumas, P.; Briki, F., New Aspects of the a-Helix to b-Sheet Transition in Stretched Hard a-Keratin Fibers. *Biophys J* **2004,** *87* (1), 640-7.

52. Shi, X. Y.; Yarger, J. L.; Holland, G. P., Probing site-specific C-13/N-15-isotope enrichment of spider silk with liquid-state NMR spectroscopy. *Anal Bioanal Chem* **2013,** *405* (12), 3997-4008.

53. Oliver, W. C.; Pharr, G. M., An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *J. Mater. Res.* **1992,** *7*, 1564-1583.

54. Takegoshi, K.; Nakamura, S.; Terao, T., $^{13}C$-$^{1}H$ dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chemical Physics Letters* **2001,** *344*, 631-637.

55. Morcombe, C. R.; Zilm, K. W., Chemical shift referencing in MAS solid state NMR. *Journal of Magnetic Resonance* **2003,** *162* (2), 479-486.