UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO STATE UNIVERSITY

The Hierarchical Structure, Dynamics, and Assembly of Spider Silk

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Chemistry

by

David Onofrei

Committee in charge:

University of California San Diego

Professor John Crowell
Professor Stanly Opella

San Diego State University
Professor Gregory Holland, Chair
Professor Christopher Harrison
Professor Marshal Hedin

2020

The Dissertation of David Onofrei is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____
Chair

University of California San Diego

San Diego State University

2020

Dedication

To my parents, Adrian and Manuela Onofrei without whose unending support I could never have reached my goals. To my family and friends who pushed me to pursue my degree. And to Christy Milani who encouraged and supported me all the way.

Epigraph

"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, in full, is a reprint of the material as it appears in Proceedings of the National Academy of Sciences. Lucas R. Parent, David Onofrei, Dian Xu, Dillan Stengel, John D. Roehling, J. Bennett Addison, Christopher Forman, Samrat A. Amin, Brian R. Cherry, Jeffery L. Yarger, Nathan C. Gianneschi, and Gregory P. Holland ca. David Onofrei was a co-primary author of this paper.

Chapter 3 is currently being prepared for submission for publication of the material. The dissertation/thesis author was the primary investigator and author of this material. David Onofrei, Dillan Stengel, Jia Di, S. Trescott, A. Soni, B. Addison, M. Muthukumar, and Gregory P. Holland. David Onofrei was a co-primary author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in Proceedings of the National Academy of Sciences. B. Addison, D. Onofrei, D. Stengel, B. Blass, B. Brenneman, J. Ayon and G. P. Holland. David Onofrei was a co-author of this paper.

Vita

2011	Associate of Arts, Portland Community College
2014	Bachelor of Sciences, Washington State University
2020	PhD Chemistry, UCSD/SDSU

Abstract of the Dissertation

The Hierarchical Structure, Dynamics and Assembly of Spider Silks

by

David Onofrei

Doctor of Philosophy in Chemistry

University of California San Diego, 2020
San Diego State University, 2020

Professor Gregory Holland, Chair

Spider silks are biological protein polymers which are spun into fibers with an incredibly diverse range of mechanical properties and functions. Spider silks have long been recognized to have mechanical properties that rival most, if not all man-made materials. However, it is not yet possible to make synthetic spider silk fibers that mimic the impressive properties of their native counterparts. This is due in part to the hierarchical nature of spider silk, which introduces a complex organization of structures at nearly every length scale from the atomic to the micron level. The research in this dissertation has been

in pursuit of understanding the structure, dynamics and assembly of silk proteins into high performance fibers at the molecular and nanoscale levels. The mechanical properties of a number of spider silks were measured and are discussed and placed intheir biological context. The underlying molecular interactions that give rise to these fibers are investigated by DLS, cryo-TEM, solution and solid-state NMR. The solution NMR work combined with cryo-TEM illustrates that silk proteins are stored as protein pre-assemblies in the silk gland and are the fundamental precursors to silk fibrils. The application of denaturant was used to disrupt these protein superstructures and understand the forces and dynamics that facilitate assembly. Finally, solid-state NMR was used to structurally characterize prey wrap spider silk, the toughest of the spider silks, and illustrate its α-helical coiled-coil hierarchical structure that helps explain this silk's high extensibility and impressive mechanical performance.

Chapter 1 Introduction

Spider Silk Introduction

The natural world contains an astounding number of biological polymers that fulfill elegant roles in their environment. Silk is one example of a protein-based biopolymer which humans have successfully put to use, in the textile industry (typically as garments) and in the medical industry (as sutures and wraps) .[1-3] Despite the diversity of silks found in nature, only silkworm silk varieties have historically seen large-scale industrial and commercial scale utilization. Unfortunately, silkworms have a single silk gland which produces one silk type only once during their entire lifespan. By contrast, an individual spider can have up to seven silk glands producing seven unique silk types most of which they use on a daily basis.[4, 5] These fibers have a diverse range of material properties and are highly specialized biopolymers adapted for a diverse range of applications from web construction to fabricating egg cases. The silk used to construct the radii of orb-webs is one of the strongest silks and one of the most impressive materials known in terms of mechanical properties while, the one used to wrap prey has a toughness rivaling all known man-made fibers including Kevlar®.[6-8] Egg cases, also made from silk, have comparable toughness to orb-web fibers because of considerable extensibility but are not as strong. Nevertheless, these egg case silks serve to modulate temperature fluctuations, and protect against predators and parasites.[9-12] Other mechanical properties of spider silk range from highly flexible and elastic to tough and brittle fibers.[13, 14] The need for green chemistry and the search for new, safe, and versatile materials makes the diverse range of favorable physical and

mechanical properties found in spider silks a rich target for bio-inspired materials research and development.

Before looking at these silks in detail, a note on naming conventions is necessary. Spider silks are called by different names, leading to some confusion in the

Figure 1 Images showing argiope prey-wrapping silk at different length scales. (a). Individual Silk fibers of AC silk are < 1 µm (Scale bar is 1 µm). Together they can be woven into individual threads which can be used as a protective lining of an egg sac (b). Swaths of these threads are used to immobilize their prey (c). The individual fibers at this level also interact with secretions from the spider's mouth (mostly digestive enzymes and water). These bulk changes are translated all the way down to the molecular level.

literature. For example, the term “dragline silk” is often used to describe the orb-web type but has vaguely been applied to a collection of silk types. The convention in this dissertation will be to refer to silks by the gland type from which they are primarily produced. Dragline silks, produced from the major ampullate glands, are called MA silk however, it should be noted that dragline which describes the silk spiders “drag” along while walking is usually comprised of mostly MA silk but, also a small amount of Mi silk. Prey wrapping silks, mainly synthesized from the aciniform gland are AC or aciniform silk. Egg case silks originating in the tubuliform gland will be referred to as TU silks although AC silks are also a minor component of egg cases.

In order to be able to harness the exceptional properties of spider silk, we must first understand the mechanism by which they are produced. This is especially crucial given that individual spiders do not produce industrial quantities of silk and are not easily farmed due to their cannibalistic nature, environmental requirements and the small amount of material produced. Therefore, spider silk materials must be developed using biotechnology and recombinant production for large-scale applications. Because replicating nature in a test tube is never trivial, we must have a firm grasp of the principles that guide silk formation from the molecular level all the way to the macroscopic scale. **Figure 1** shows AC silk at several different length scales to demonstrate how emergent properties are present in each domain. The first step to developing a complete understanding of silk formation is to understand the protein components at the sequence level, which to a large degree, has been accomplished for a number of silk types and will now be described.

As previously stated, silk proteins are produced from several different types of specialized silk glands.[4] Each gland is associated with a different main silk protein or combination of proteins, known as a “spidroins”. The most studied gland, the major ampullate (MA) gland, is the primary site of synthesis of MA silk proteins. MA silk is primarily composed of two spidroins: major ampullate spidroins 1 and 2 (MaSp1 and MaSp2), although recent research has shown there may be several other MaSp and MaSp-like proteins present depending on the spider species.[15-17] **Figure 2** gives an overview of the MA gland with the important regions described in the caption.

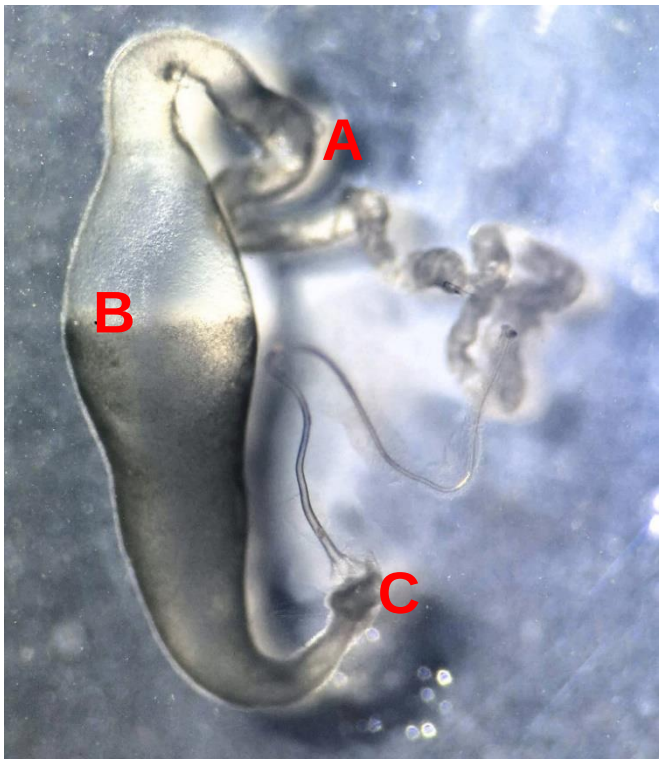


Figure 1 Black widow MA gland. (A) shows the tube from which the solution of fiber spinning dope is created. (B) is the storage sac where the spidroins are stored at very high concentration (~25-50 wt.%). The sac opens to the spinning ducts (C) which leads to the spinnerets. The transition from the soluble form to the insoluble fiber occurs very quickly along the duct with water content dropping, exchange of Na^+ ions for K^+ ions, and continuous pH gradient that gets more acidic as the dope moves towards the spinnerets where the fiber is extruded.

MaSp proteins are very large (>250 kDa) and all share three common characteristics:

1. A short highly conserved, non-repetitive amino terminus containing a bundle of alpha helices which forms dimers in response to pH.[18, 19] The length of this regions is between 120-130 residues.
2. A large repetitive core region, highly variable in length but typically 2000+ AA's
3. A short non-repetitive, carboxy terminus which may be involved in dimerization of the chains between 80-110 residues.[20]

The repetitive cores of different spidroins have different compositions but for MaSp-type proteins they generally consist of very short motifs composed of simple amino acid residues. In fact, glycine and alanine compose over 70% of the sequence of

…YGRGGYGQGGAGQGGAGAAAAATAAGAGQGGQGGYGQGGYGQGGAGQ
GGAAAAAAAAAGGAGQGGYGRGGAGQGGAAAAAAAAAGAGQGGYGGQGAG
QGGAGAAAAAAGGAGQGGQGGYGRGGYGQGGAGQGGAGAAAAAAAAGGA
GQGGQGGYGQGGYGQGGAGQGGAAAAAAAAAGGAGQGGYGRGGAGQGG
AAAAAGAGQGGYGGQGAGQGGAGA…

…GGQSGYG**P**GGAGAAAAAAAGGAG**P**GRQQEYG**P**GGSGAAAAAAAAAGSG**P**
SGYG**P**GAAG**P**IG**P**GGAGAAAAGGSG**P**VGYGQG**P**SGYGASGTGGEQDYG**P**G
GSGAAAAAAAAAGGAG**P**GRQQGYG**P**GGSGAAAAAAAGG**P**GYGGQQGYG**P**G
GAGAAAAAAAGGAG**P**GRQQ**P**YG**P**GGAGAAAAAAGGSG**P**GGYGQG**P**SGYGA
SG**P**GGQQGFG**P**GGSGAAAAAAAGGAG**P**G…

Figure 2 Short excerpts from the full sequences of MaSp1 (top) and MaSp2 (bottom). Gly and Ala are by far the largest components of these spidroins. The primary difference in composition between the two is the presence of proline rich motifs in MaSp2 which are completely absent in MaSp1.

MaSp1 in *Latrodectus Hesperus* (commonly known as the Western Black Widow). The repetitive domains contain small blocks of several simple amino acid repeats: Gly-Gly-X, Gly-Ala-Gly, $(GA)_n$ and poly(Ala). X here represents either Ala, Tyr, or Gln. MaSp2 also contains Gly-Pro-Gly-X motifs which are not found in MaSp1. Examples of the two spidroins are shown in **Figure 3**. While many spiders seem to contain analogues to these, amino acid analysis of glands of different spiders suggest certain species, such as *A. aurania* may have a greater proportion of MaSp2.[21] Variation can also occur between individuals of the same species in response to environmental stimuli, life stage, or diet, but long term variations are thought to be low.[22]

Aciniform fibers have only one known spidroin component, AcSp1. While also having a large repetitive core domain, the repeat unit of AC silk differs greatly from MA silk. The spidroin sequence contains a much greater diversity of amino acids and the repeat unit is much shorter as seen in **Figure 4** containing a couple hundred AAs. In *A.*

LISRVANALSNTSTLRTVLRRGVSQQIASSILRRAAQTLASTLGVDGNNLSRVAL
QAISQVPTGSDTSAYAQAFSSALFNAGVLNASNIDTLGSRVLSAVLNGVSSAAQGLGINV
DTGSVQSDISSSSSFLSTSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSA
SGQSSFGQTSGLTASSGGQAAFGGTSGASAG

Figure 3 The sequence of the repeating unit in AC silk of *Argiope argentata.* This sequence is well conserved among Argiope *spp*. and represents.

argentata, the species used in this work, the repeat unit consists of 199 amino acids and is iterated 14 times between the termini with virtually no variation.[7, 23] It should come as no surprise, looking at this sequence, that the mechanical properties of this fiber type are very different from the MA fibers. The sequence and its molecular

structure will be discussed in chapter 4. Next we will discuss these properties and how they are determined.

Mechanical Properties of Spider Silks

Understanding the mechanical properties of spider silks and, crucially, their correlations to the underlying atomic, molecular and nanoscale structures of the silk

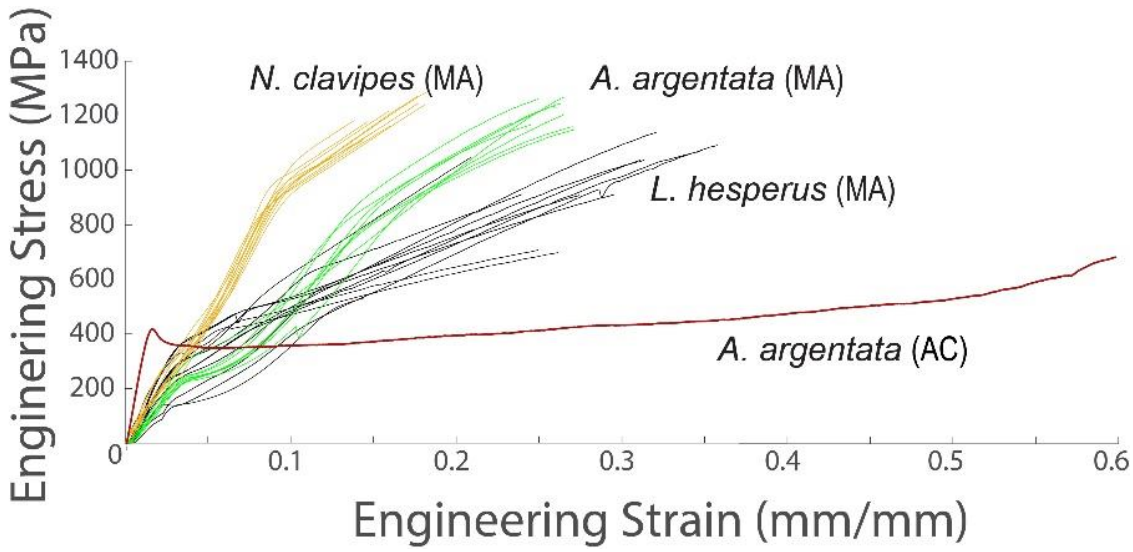


Figure 5 Mechanical stress-strain curves for dragline (major ampullate, MA) silk collected from three different spider species and *A. argentata* prey wrap (aciniform, AC) silk. Differences in the ratio of different MA spidroins contained in the different dragline silks are likely to lead to the different mechanical properties observed. See text for sample prep details.

proteins is a critical key step to generating synthetic silk fibers with a range of desired physical properties. Much of the interest in the material properties of spider silk is founded on testing of the mechanical properties of MA silk which is one of the toughest materials known. This silk type can be forcibly extracted from spiders and thus, easily collected in reasonable amounts for experiments. MA silk has incredible strength for a biopolymer fiber compared to man-made materials[24, 25], but AC silks have both high strength and extensibility making them significantly tougher.[13, 26] These silks are used to

wrap prey and also line the inner compartment of the spider egg sacs (**Figure1A, C**). [27] These two silk types are arguably the best industrial silk production candidates where the best mechanical properties are the end goal. A summary of the testing of various spider silks is shown in **Figure 5**. A brief description of how these fibers were tested will now be given.

A KeySight UTM T150 mechanical testing system was used to measure the tensile properties of MA silk from three different species of spiders, *N. clavipes, L. hesperus,* and *A. argentata*. Mechanical testing was performed as close as possible to the method described by Blackledge et al.[24] Briefly, spiders were anesthetized with CO_2, restrained on their backs and forcibly silked at a rate of 1.5 cm/s once the spider came to. To ensure our tests were single fiber measurements we used a needle to gently break all but a single fiber. While silking, pieces of cardstock with rectangular cutouts were placed underneath the extended fibers. Acrylic glue was then used to secure the strand of silk in place on either side of the cutouts. These cardboard cutouts were then placed vertically in the tensile tester and an increasing load was placed on the fibers such that they were strained at 1% per second until failure.

In order to ensure that CO_2 did not affect the fiber's mechanical performance, spiders were not silked until clearly awake and recovered from anesthesia, following established protocols[28]. Relative humidity were measured to be between 39 to 59% although a few measurements were measured as high as 65%. Temperatures for both silking and testing environments remained constant between 21 and 22^{o} C. During sample collection, visual inspection ensured that only MA fibers were collected. Care was taken to maintain a small amount of slack in the fiber prior to loading to prevent

damage or early stretching. The resulting stress-strain relationship was plotted using the Keysight NanoSuite software. Engineering values for stress and strain for spider silks are generally converted to true stress and strain using the constant volume assumption which is made as follows:

True mechanical stress (σ) and strain (ε) are defined as

$$\sigma = \frac{F}{A}$$

$$\epsilon = \int_{L_0}^{L} \frac{dL}{L} = \ln \frac{L}{L_0}$$

Both quantities are very difficult to measure as the area changes as the test proceeds. For this reason, engineering values are typically reported which are use the original fiber length and area, making the assumption of a circular cross-section,

$$\sigma_{eng} = \frac{F}{A_0}$$

$$\epsilon_{eng} = \frac{L - L_0}{L_0}$$

In order to convert these values into the true values, we assume that the volume of the material does not change during the test such that,

$$L_0 A_0 = LA$$

Which rearranges to

$$A = \frac{A_0 L_0}{L}$$

And substituting and rearranging into equation for engineering strain gives,

$$A = \frac{A_0}{1 + \epsilon}$$

Finally, this can be substituted into the original equations to give the conversion formulas,

$$\sigma_{true} = \frac{F}{A} = \frac{F}{A_0}\left(1 + \epsilon_{eng}\right) = \sigma_{eng}\left(1 + \epsilon_{eng}\right)$$

$$\epsilon_{true} = \ln\frac{L}{L_0} = \ln(1 + \epsilon_{eng})$$

Both engineering and true stress strain values of silk testing have been reported in the literature, but the constant volume assumption is generally made for spider silk and has some empirical confirmation.[29]

In order to measure fiber diameter we used optical microscopy on larger diameter individual fibers (MA and TU). Measured MA silks are well within the limits of published values. Humidity, age, anesthesia, reeling speed, and spider condition (gravid, time since last feeding) and many other variables have been shown to result in fibers of varying mechanical properties, even for spiders of the same species and sometimes even among individual spiders.[30-33] Nevertheless, our results hold with the general trends that have been published in the literature. *N. clavipes* and *A. argentata* MA silks are slightly stronger than *L*. *hesperus* while the latter is the more extensible.[6] A summary of the relevant mechanical properties can be found in **Table 1.**

Having validated our methods for measuring the mechanical properties of MA silks with respect to the literature, we next attempted to develop a method for studying the mechanical properties of AC silk. As previously stated, this silk type is much tougher than MA silk and consists of fibers with diameter on the order of 500 nm (**Figure 1a**). Despite this, an exhaustive search of the literature finds only two research studies on the mechanical properties of AC silk.[13, 26] One of the major challenges in studying this

Table 1 Summary of mechanical properties for various spider silk fibers. Aciniform fibers (*) have only a small number of tests performed.

Species	Strength(MPa) Engineering/ True	Extensibility(mm/mm) True/Engineering	Toughness (MJ/m^2)	Fiber Diameter (um)
L. hesperus(MA) N = 24	1008 ±267 1280 ±284	34 ±4% 29 ±4%	164 ±82	2.54 ±0.15
L. hesperus(MA^{sc}) N = 3	602 ±92 919 ±99	53 ±8 42 ±7%	153 ±36	3.43 ±0.3
N. clavipes(MA) N = 28	1146 ±188 1340 ±194	17 ±3% 16 ±3%	113 ±33	3.59 ±0.2
A. argentata(MA) N = 15	1204 ±104 1524 ±106	27 ±2% 24 ±2%	177 ±15	1.84 ±0.84
Acin N=2	650	60%	230*	0.7
A. aurantia(TB) N = 4	212 ±27 318 ±35	65 ±7% 50 ±7	100.79 ±1.38	6.42 ±0.26
L. hesperus(TB) N = 3	314 ±35 448 ±41	43 ±15% 36 ±14%	118.42±	4.31±12

silk type is that fibers are only spun during prey wrapping and egg case fabrication and cannot be forcibly silked. Instead, they are harvested as bundles by inducing spiders of the *Argiope* genus to “attack wrap” by inducing vibrations in the orb webs where small template cards are inserted to collect the silk during this process. This behavioral characteristic allows us to prepare samples of wrapping silk bundles as validated by SEM. As previously noted, there appears to be another source of AC silk fibers which

are apparently spun as a minor component during egg case production (**Figure 1b**). Spiders spin these thin fibers into the inner lining of their egg sacs and SEM confirms the presence of small diameter fibers within the egg sacs. A few tests from the inner lining of black widow egg cases have given results comparable to prey wrap fibers, but the sample variation is extremely high and more work is necessary to standardize this technique[27].

Silk fiber diameters for AC silk were determined by taking a sample of fibers from each batch of fibers and mounting them to a piece of carbon adhesive tape, sputter coating with a 6 nm thick layer of platinum and imaging with SEM. The measured average was then used for all fibers of that batch. We opted for this method rather than optical microscopy because of the higher resolution of SEM and the negligible variation between fibers from a single batch. Aciniform fibers are consistently sub-micron in diameter, and while reports of optical measurements of this length scale using very high magnification and polarized filters have been reported, they are extremely difficult to replicate[34]. We found the variation between SEM and optical measurements acceptable for our purposes.

Obtaining AC single-fiber measurements is extremely challenging due to the fact that these bundles cannot be readily separated into single fibers by optical microscopy, since their diameters approach the wavelengths of the illumination source. For this reason, our approach has been to stretch bundles across two separate gaps in the template cards. We can then separate bundles as thinly as possible using the same technique developed for MA silk, perform mechanical testing on one of the pairs of the template, and then use SEM measurements from the second pair to confirm the number

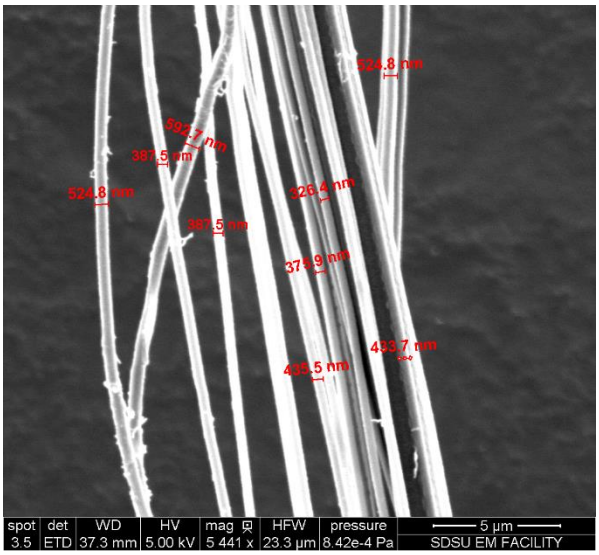


Figure 6 A collection of aciniform fibers highlighting their sub-micron diameters. Fibers are generally collected as bundles rather than individual fibers making individual fiber measurements challenging.

of fibers tested. In certain instances, this results in single fibers being tested across a template. Ultimately, this technique will allow us to measure the scaling factor of using multiple fibers and allow back calculation of single fiber measurements. We have used this method to obtain stress-strain measurements of a small number of aciniform fibers which match the very limited literature (**Table 2**.)

We further intend to study the effect that water wetting has on the mechanical properties. We have already begun studying the molecular changes of AC silk in dry,

wet, and post-wet states [35]. Preliminary results shown in **Figure 6** demonstrate that real morphological changes are induced by water treatment. This validates SEM data showing AC fibers taken from previously wrapped prey as having been fused together as seen in **Figure 7**. The next logical step, we believe, involves correlating the mechanical and molecular approach to fully elucidate the fundamental mechanism of the tensile properties of wrapping silk and especially the dynamics between water and protein fiber.

Another property of MA silks that has generated considerable interest is the supercontraction. This refers to the behavior of MA silk fibers to contract in length by as much as 60% when exposed to water. The proposed function of this is to maintain the

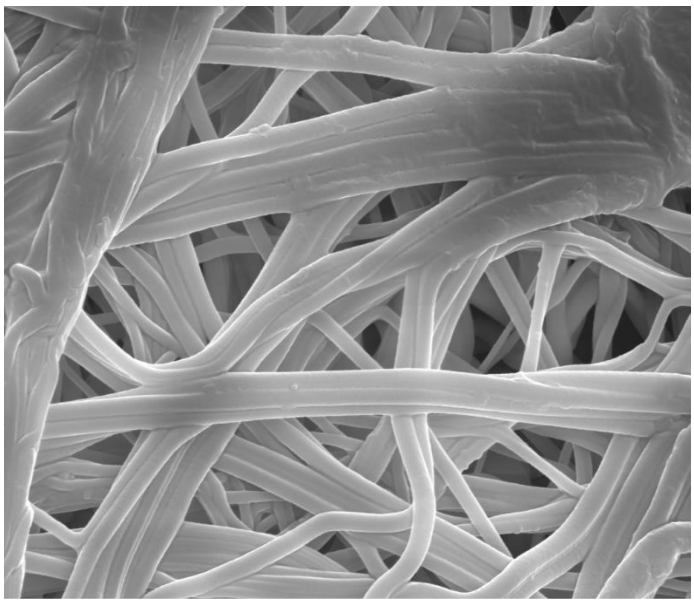


Figure 7 Electron micrograph of AC silk taken from a previously wrapped cricket. Individual fibers are still recognizable, however, clear signs of fusion can be seen. A combination of water and digestive enzymes from the spider change the physical appearance of this silk from a thread-like material to a matte-like sheet. The scale bar is 5 µm.

structural integrity of silk fibers when exposed to weather by preventing over-stretching of fibers, a type of self-repair[36-38]. This property is still under investigation, but appears

to vary considerably based on environment and species making reproducibility very challenging. **Figure 8** shows early testing done on MA fibers that have been wet, supercontracted and subsequently dried. More work remains to be done as

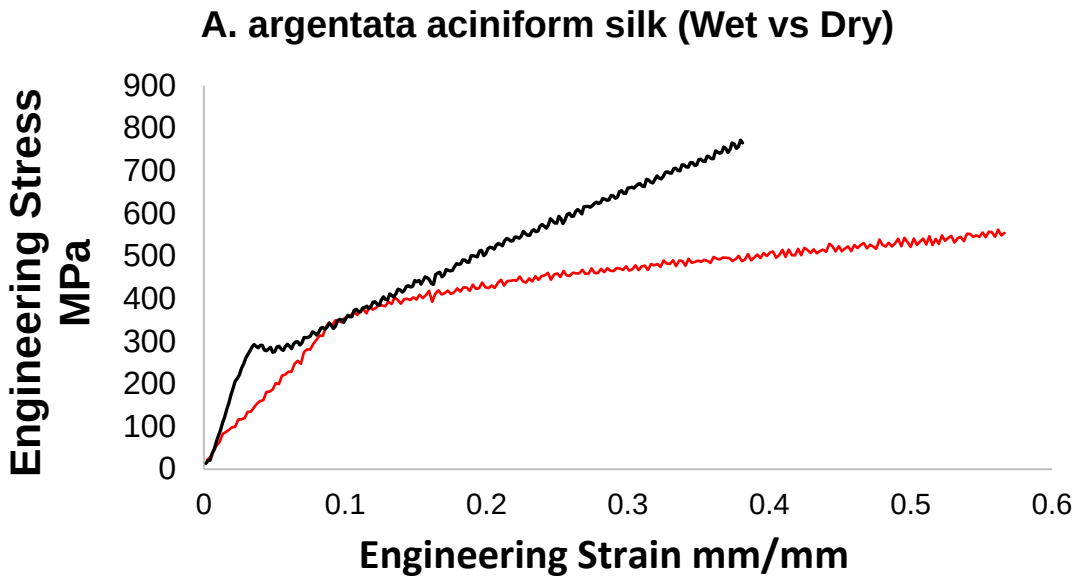


Figure 8 Single fiber MT tests of AC silk fibers. The absence of breaks in the curves before failure indicates that there were indeed single fiber tests. SEM and light microscopy have both been investigated as a means for confirming this as well. The black curve is a single naturally spun fiber while the red curve was water wet and then dried. The black curve has strength and extensibility of 700 MPa and 35% extensibility while the red curve drops to 550 MPa strength but extends out to 56%.

considerable sample variation is seen with even slight changes to testing environment. Nevertheless, the loss of strength and increase extensibility mirror the published values of super contracted MA silks.[37, 38]

The huge diversity of mechanical properties makes spider silk a highly versatile biopolymer with a multitude of potential industrial and medical applications. Unfortunately, spiders produce much less silk than silkworms and their territorial and cannibalistic temperament prevents large scale farming of these natural fibers. On the other hand, 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 dynamics and how these translate into mesoscale/bulk material properties. The challenge here lies in the fact that we do not fully understand the molecular and physical processes involved in the formation of silk from a soluble protein solution, or silk dope, to a nanocrystalline fiber that contain a broad range of secondary

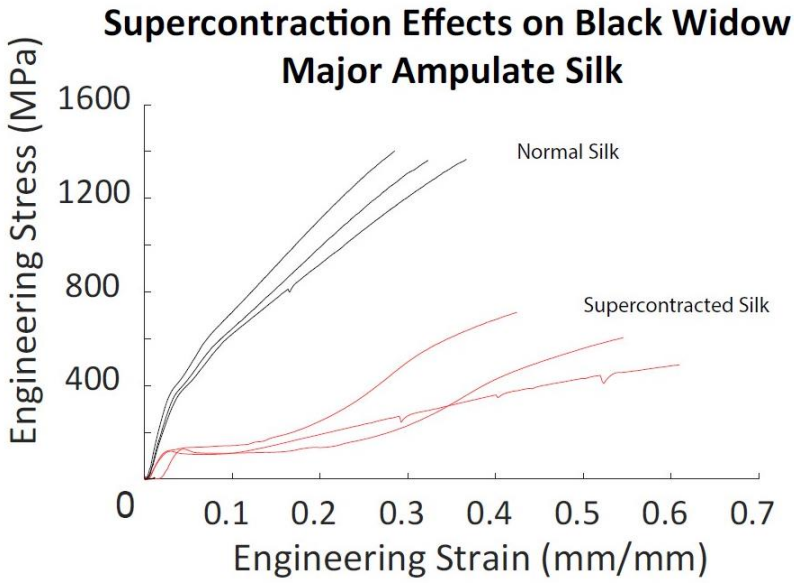


Figure 9 Supercontration of MA silk fibers from black widow maintains tension in the orb-webs of spiders but comes at the cost of overall strength of the fibers. The teste above shows fibers that were tested while still wet, or in their native dry state.

structures and a combination of highly ordered, intermediately ordered and highly disordered domains across molecular and nano length scales. We address this challenge in the subsequent chapters of this thesis but first we must discuss the molecular level description of spider silks which, after all, are the critical underpinnings for the differing bulk macroscopic properties.

Molecular Structure of Spider Silk Proteins

The structure of spider silk proteins in fibers has been studied by a variety of methods including solid-state NMR[39-42], X-ray diffraction[43, 44] and Raman spectroscopy.[45] As a result, a simplistic model of the hierarchical structure of silk proteins in fibers has been developed and well accepted in the silk community.[46] For MA silk, this model of silk fiber is a semi-crystalline polymer-type structure containing nanocrystallite regions embedded within an amorphous matrix. More correctly, the amorphous region can be better described as a semi-amorphous or glassy disordered structures as it contains both a disordered and an partially oriented domain (along the fiber axis).The fiber core is surrounded by two layers: an outer lipid layer and an inner "skin" layer which maintains the pressure exerted by the amorphous region[47, 48]. The inner core is composed of smaller fibrils oriented along the fiber's main axis. **Figure 9** summarizes the main points of this model. The glycine-rich Gly-Gly-X motifs according

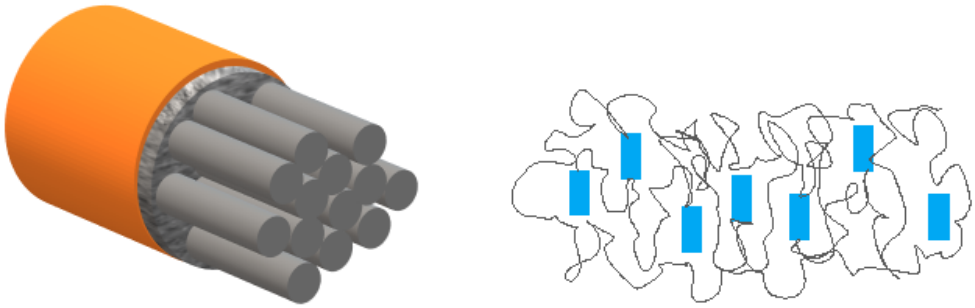


Figure 10 A summary of the model of spider silk fibers based on EM and molecular studies. The fiber core is surrounded by an outer covering (left) consisting of glycoproteins and lipids. This layer maintains the equilibrium pressure of the inner semi-amorphous regions and the fibrils (grey). The structure of the fibers themselves are shown on the right with blue rectangles symbolizing the beta-crystalline domains and the lines in between showing the loose coils that surround them.

to SSNMR exist primarily in coils, loose helical structures, β-turns and some disordered β-sheets.[49] The loose helical structures are believed to serve as intermediates between

the two types of secondary structure while the coils exist primarily in the disordered matrix. Proline rich regions, which are common is MaSp2 exist in β-turn type structures reminiscent of those observed in elastin. And finally, the nanocrystalline components are made up of poly-alanine rich domains. Studies continue to clarify how these nanocrystalites are arranged in the core region of the fiber.

There is far less known about the structure of AC silk, but recent analysis of recombinant silks have provided some insight. Recombinant expression of one repeating unit and subsequent analysis by solution NMR led to the first structural model of a single repeat of AC silk protein in solution.[50] In aqueous solution they appear as bundles of 5 or 6 α-helices, separated by loose coils and terminated by unstructured tail regions containing mainly serine rich regions of the repeat.[51] Recombinant proteins have even been pulled into fibers, albeit with inferior mechanical properties compared to natural full-length AC silks.[52] Nevertheless these studies have shown that AC silk, which is primarily α-helical in solution, maintains significant α-helical character even after spinning recombinant fibers. The first SSNMR structural analysis of AC silk fibers is discussed in detail in Chapter 4.

The structure and dynamic state of spider silk proteins prior to spinning has been one of the most challenging aspects of spider silk investigations. Prior to spinning, and while the spidroins are contained in the storage region of the gland (**Figure 2B**), they remain essentially unstructured (random coils) except for the termini that have been investigated individually and shown by both NMR and circular dichroism to be well defined a-helical bundles.[53-55] Nevertheless, the speed at which silk is converted from soluble dope to fibrous product suggests that some hierarchical pre-assembly may be

present even in the solution state. Without local secondary structure this order is expected to be supramolecular in nature. Furthermore, the spidroin proteins are maintained at extremely high concentrations between 30-50% wt.[55] The mechanism of storing such a highly concentrated dope of very large proteins is key to understanding how silk is stored in solution and on-demand spun into a fiber.

The current understanding of the spinning process is based on several principles which originated in the study of silkworm silk.[56] In silkworms, the large repetitive core domain of the silk protein is dominated by hydrophobic residues while the termini contain hydrophilic ones. A proposed mechanism for the assembly of silkworm silk, and further extended to spider silk, relies heavily on this distribution. It was hypothesized that the silk proteins form roughly spherical micelle-like assemblies in the storage solution.[56] 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. 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. 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. The stacking of these structures then would give rise to the final fiber.

This model of silk assembly has been extended to include spider silk as well despite significant differences in the sequence and a very different distribution of hydropathy along the protein. As with the mechanical work, a large chunk of the analysis to date has focused on MA silk formation. The hydropathy plot of *L. hesperus*

MaSp proteins show the more hydrophilic residues are localized in the termini and the more hydrophobic ones in the core but the core repeat region shows a greater distribution of hydrophilic residues than silkworms.[57] Nevertheless, shear force and an acidic pH gradient in the duct region (along with other biochemical triggers) is expected to occur in a similar way. This highly concentrated solution has also been described as a type of liquid crystal as the flow through the duct induces a tight packing order. With an understanding of the molecular and mechanical nature of the silk proteins, a brief discussion of the primary methods used to investigate silk in our lab will be given.[58, 59]

Introduction to Magnetic Resonance Techniques for Proteins

Nuclear magnetic resonance (NMR), first described in 1944 is a spectroscopic technique that is ubiquitously applied throughout all fields of chemistry, structural biology and materials science. It is grounded in an intrinsic, quantized property called nuclear spin, a property shared by the majority of nuclei. This nuclear spin gives rise to a quantized nuclear angular momentum which in turn gives rise to a nuclear magnetic moment. The proportionality constant between the two is called the gyromagnetic ratio. When a strong magnetic field is applied to an ensemble of nuclear spins, the nuclei undergo rotational movement known as “precession” at a frequency known as the Larmor frequency. This frequency is proportional to the energy associated with the interaction of the nuclear magnetic moments and the magnetic field. The *effective* magnetic field felt by the nuclei, however, is attenuated by the interaction of the field with the electronic environment around the nucleus, which also has magnetic properties. The different energy levels therefore, vary slightly as nuclei in a molecule or

protein exist in varying magnetic environments based on the localization of electrons (and other factors). Because electron distributions in organic and biochemistry are the result of bonding, molecular information is thus encoded into signals generated from NMR spectra in the form of a “chemical shift” value. The ensemble of nuclear magnetic moments is isotropic at equilibrium in the absence of a magnetic field, and is only slightly perturbed by the presence of an external, B_o, field. This slight inequality in the population of energy states is known as the “net magnetization”. A second oscillating magnetic field can be applied as a “pulse” to this distribution of spins to make this slight perturbation observable. Used in this simple 1-dimensional version, MR is of limited use for protein characterization. However, a number of experiments have been devised to expand the molecular information obtained by NMR into 2 and 3 dimensions. These experiments apply a “pulse sequence” which can transfer magnetization between nuclei and contain information on the structure and dynamics of proteins. Higher dimensional experiments (4 and 5D) are also possible but are more challenging and have more specific applications. They will be described later in the final chapter.

The first and most common 2D NMR experiment in protein NMR is known as the heteronuclear single quantum coherence experiment (HSQC). The pulse sequence for this experiment is composed of two blocks of the INEPT sequence (insensitive nuclei enhanced by polarization transfer)[60]. This transfers the magnetization of ^{1}H nuclei to nearby heteronuclei (typically ^{15}N or ^{13}C) via spin-spin interactions via J-coupling and then back again for detection on the ^{1}H channel. The resulting spectra shows one-bond correlations between heteronuclei and directly bound protons. If ^{15}N is the heteronuclei used, the spectra gives a broad picture of the overall protein chain backbone and its

conformation (secondary structure). An example of this spectrum for a well folded protein is shown in **Figure 10**. Furthermore, even though nmr-active isotopes of all biologically important nuclei exist in nature, they are often at very low natural abundance (^{13}C, ^{15}N). Multi-dimensional techniques such as HSQC enhance the magnetic properties of these low gamma nuclei and increase the sensitivity of the

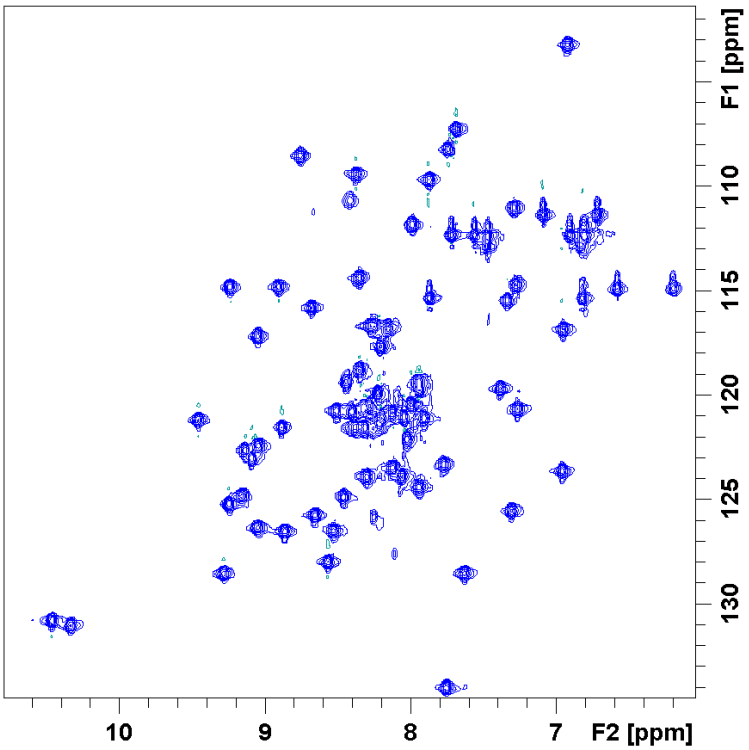


Figure 11 ^{15}N-HSQC of GB1. The proteins is fully C-13, N-15 labeled at all sites and serves as a good reference sample for other experiments. It is worth noting that GB1 is well folded and highly ordered, which provides the high degree of resonance dispersion. Due to its stability, it functions well to optimize parameters for multidimensional NMR experiments before switching over to less stable spidroin samples.

experiment and typically the ^{15}N and/or ^{13}C are isotopically enriched to further improve sensitivity. The enhancement due to much greater sensitivity for proton detection and can lead to better resolution. A number of modifications have been made to this sequence over the years which have improved sensitivity and reduced artifacts from

solvent.[61-63] This can serve as a map of the protein backbone, often with unique resonances for individual amino acid residues, as is the case for GB1 shown below.

However, this simple mapping technique is rarely sufficient to assign the HSQC peaks of anything but the simplest proteins. In order to link together resonances from sequential residues, another set of 3D experiments must be performed to transfer magnetization further along the backbone chain.

Backbone Assignment Strategy by NMR

The general approach for investigating large biomolecules like proteins begins with collecting a two dimensional NMR experiment such as a $^{1}H/^{15}N$ HSQC to generate a landscape of the protein backbone amides, followed by three-dimensional (3D) experiments to piece together the backbone connectivity and secondary structures[64]. This strategy relies on large chemical shift dispersion in order to resolve unique resonances corresponding to a single residue. Once the amides are resolved, 3D experiments identify amino acids based on correlation from the amide ^{15}N peaks to either the $^{13}C\alpha$ or the $^{13}C'$ and then $^{13}C\beta$ of the sidechain. In the ideal case a full "backbone walk" can be performed for the entire protein backbone. Furthermore, because secondary structure contributes to the chemical shielding in a predicable manner, the chemical shift information extracted from these experiments also informs on the secondary structure of the protein. An illustration of this is shown in **Figure 11** where a cube is used to project the peaks of a 3D experiments. Each face of the cube represents a 2D plan of the spectrum and the three nuclei are the three axis of the cube (^{1}H, ^{13}C, ^{15}N). The spectrum is never processed as a cube, but rather as a series of 2D

slices or “strips” from the full dataset. These strips are chosen where each of the 3D resonances have their maximum volume. We primarily focus on two 3D experiments which we use to identify large sections of a long peptide chain. The first of these three is called the HNCACB[65, 66] and is shown in the top part of **Figure 11**. Much like the HSQC, this experiment relies on the transfer of magnetization from one nucleus to another. The transfer pathway goes:

$$^{1}H \rightarrow ^{15}N \rightarrow ^{13}C\alpha / ^{13}Cb \rightarrow ^{15}N \rightarrow ^{1}H$$

The resulting resonances connect amide ^{15}N and ^{1}H cross peaks with the $^{13}C\alpha$ and $^{13}C\beta$ peaks of the same residue (the i^{th} residue). There is also another weak set of resonances coming from the preceding (i-1) residue. A second important feature of this

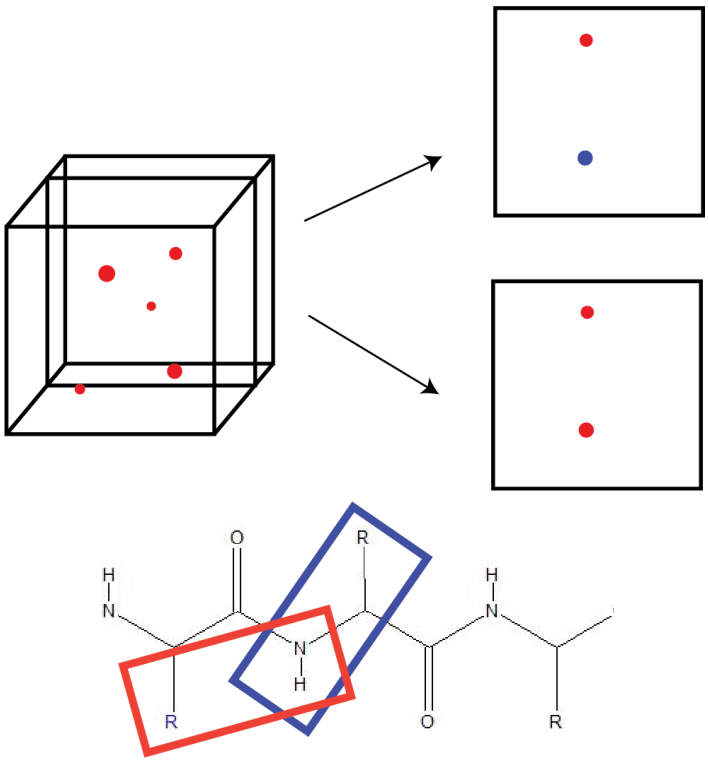


Figure 12 A graphical representation of the backbone walk strategy. The cube is a general representation, and the analysis is done on individual 2D slices as shown in the top figure. The blue outline highlights the residue revealed by the HNCACB experiments while, the red shows the CBCACONH experiments. Both correlate peaks to a common amide peak and the overlap is used to “walk” along the backbone.

spectrum is that Cβ and Cα peaks show up in opposite phase. The pulse sequence is shown in **Figure 12**. A second experiment, called the CBCACONH[67], contains co-incident carbon peaks with the first experiment, but correlates them to successive amide ^{15}N to identify a new amide that can be used to find the next residue in the HNCACB experiment.[68] **Figure 11** illustrates the direction of the experiments along the protein backbone.

^{1}H -> Cβ ->Cα -> (C=O) -> ^{15}N -> H

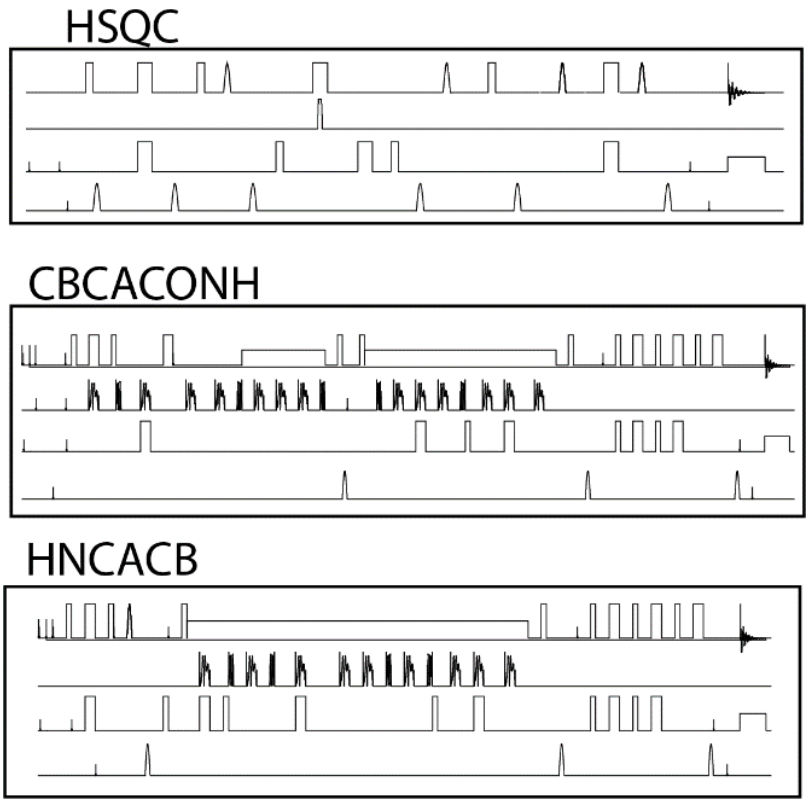


Figure 13 The pulse sequences of the three experiments used to follow the "backbone walk" strategy.

Dynamics of Proteins Measured by NMR

One of the unique capabilities of solution NMR is that it informs on dynamic processes such as local chain movement and flexibility, solvent exposure, proton

exchange processes and diffusion. The fibrilization of silk proteins being a highly dynamic process, these parameters are integral components to a complete fundamental understanding of spider silk assembly. Indeed, one of our goals is to track the dynamics over time as biochemical changes to the silk dope solution are introduced. Three of the most common dynamic parameters measured in proteins are spin-lattice (T_1) and spin-spin (T_2) relaxation and heteronuclear Overhauser enhancement (NOE) transfer. These processes involve a return to equilibrium and loss of coherence between excited spins and are tied to local chain movements and solvent environment. Diffusion of proteins is another type of motion that can be measured by solution NMR and can inform on intermolecular interactions or solutions containing multiple oligomers.

Diffusion NMR

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 performs a standard 90^o NMR pulse and then applies a non-uniform gradient field across the sample. In this way, spins in different parts of the sample will begin to oscillate at different frequencies, thereby becoming spatially labeled, or encoded. A 180^o pulse flips the transverse magnetization of the spins and then a gradient field in the opposite direction of the encoding pulse is applied. If the spins had been fixed in their original positions with little or no diffusion, the result would yield a signal with the same intensity as a standard Hahn echo. However, the spins are clearly not fixed and can diffuse at various rates to

other parts of the sample which have experienced different fields in the B_0 direction. If sufficient diffusion occurs the 180º pulse and second gradient pulse is unable to fully refocus the spins and a loss of coherence will result in a lower intensity signal proportional to both the strength of the magnetic fields applied and the rate of diffusion. **Figure 12** illustrates this. By repeating the experiment with increasing gradient strength, the diffusion coefficient (D) of the spins can be calculated from the decaying exponential using the formula developed by Stejskal and Tanner:

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

where γ is the gyromagnetic ratio of the nucleus, δ is the length of time the gradient

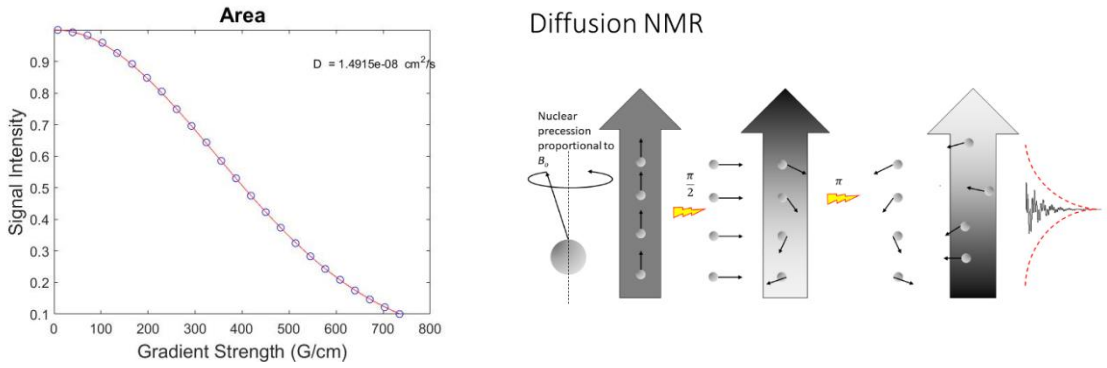


Figure 14 NMR PFG diffusion data and fitting (left) for the decaying signal intensity is shown for glycerol. The effect of the gradient strength is illustrated pictorially on the right, with dephased spins reducing the FID intensity if they diffused during the diffusion time, Δ.

pulse is applied for, g is the strength of the gradient pulse, Δ is the time in between the two pulses, and D is the self-diffusion coefficient. S and S_0 are the signal intensities at a given g value (S) or at a baseline value with no gradient (S_0). The shape of this curve is shown in **Figure 11** for glycerol, which is used to calibrate the strength of the gradient fields for slow diffusing molecules (such as very large proteins).

Modifications have been made to this pulse sequence in order to reduce the effects of eddy currents and loss of signal intensity due to T_2 relaxation. In this sequence, a pair of opposing, or biopolar, gradients (BPG) make up each of the encoding and decoding pulse sequences in order to reduce ADD. The experiment is also a modified spin echo, specifically a stimulated spin echo (STE), which prevents the loss of signal due to transverse decoherance (T2) by storing the magnetization along the z-direction T1 which is typically longer even for slowly diffusing molecules.

Relaxation Parameters

Two other informative measurements of dynamics of proteins involve the relaxation processes of the nuclei. As before these peaks rely on a two dimensional approach, from which resolution is possible and relaxation can be tracked in a third dimension. After assigning peaks to different residues of a proteins, the relaxation parameters can be used to describe the local chain dynamics for a given residue in a protein. Similar to the diffusion experiments, they are collected as a series of experiments with a certain “delay” inserted into the pulse sequence which results in a loss of signal intensity (or integral area). Again, this results in peaks changing intensity according to an exponential decay as:

$$T(1\ or\ 2) = I_o * e^{-\frac{t}{T(1\ or\ 2)}}$$

For proteins, this general description of relaxation is typically collected as a 2D HSQC experiment with a pseudo third dimension corresponding to the relaxation delay. The volume of the peak decreases over time and the resulting curves can be fit to the following formula from which T1 and T2 can be extracted. This results in better

resolution and more localized assignment [69, 70]. An example of this type of fit is shown in **Figure 12**.

Nuclei excited by an oscillating pulsed magnetic field are subsequently acted upon by two different forces. T1 describes the interaction with nuclei at the NMR Bo magnetic field strength or very close to the Larmor frequencies. T2 describes interactions between the nuclei themselves and results in decoherance of the spins and is the primary reason for the decaying nature of the FID. These parameters will be dealt with more in detail in chapter 2 where there relationship to the spidroins will be described and discussed.

Another dynamic parameter that can be used to detect interactions through space, rather than through bond, is the nuclear Overhausser effect (NOE). This experiment is collected by performing a two dimensional experiment once with proton

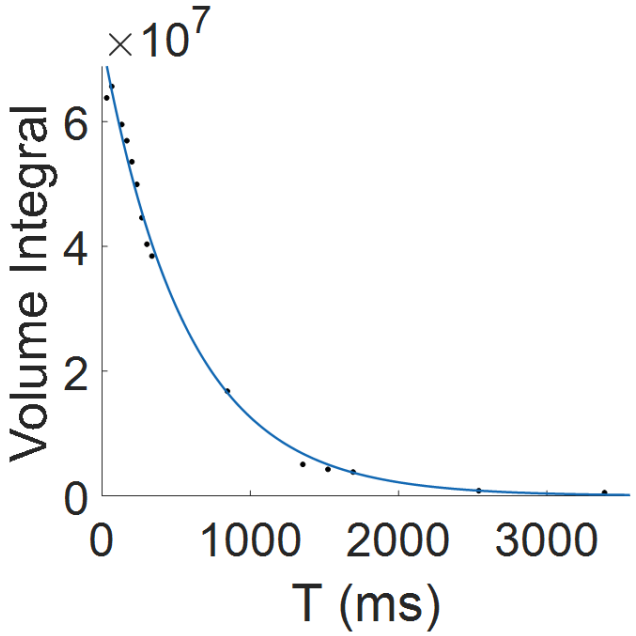


Figure 15 An example of the decay curve for T1 and T2 experiments.

saturation on and once with it off during the recycle delay. Nuclei that are close to each

other in space will see transfer via NOE and therefore will have significant NOE cross peaks visible once the two spectra are subtracted from each other. The equation for NOE is:

$$NOE = \frac{I(saturated)}{I(equilibrium)} - 1$$

The NOE observed in this case is the heteronuclear (^{15}N) NOE value which shows through-space transvers of energy transfer between nuclei. Again, this experiment and its specific use in spider silk analysis will be discussed in Chapter 2. What follows next is a brief introduction to higher dimensional NMR experiments which are planned to be performed and have been previously used to study intrinsically disordered proteins.

Higher-Dimensional NMR and Direct Detect Experiments

Besides the highly degenerate protein sequence, disorder and lack of secondary structure greatly reduces the dispersion of signals in ^{1}H dimension. In fact, the natively unfolded and unstructured character of spider silk proteins closely resembles IDP's. The problem of severe spectral overlap and reduced chemical shift dispersion has long been a challenge to studying IDPs. For these systems, higher dimensional (4D and 5D) ^{13}C-^{13}C, and ^{13}C-^{15}N experiments supplant ^{1}H-^{15}N correlations to provide all of the information used to assign resonances. A mentioned above, unlike their ^{1}H-detect counterparts, the ^{13}C-detect versions provide a means of observing proline residues. Proline is relatively common in MA silk proteins of black widows and also maps primarily to only one of the protein components. This gives ^{13}C-detect spectra the ability to directly inform on both inter- and intra-molecular interactions. Using a strategy similar to

the one used by Felli et al. we will develop a backbone assignment strategy optimized for spider silk protein motifs.

In brief this strategy involves a complex series magnetization transfer pathways which provide excellent resolution for assigning sequential amino acid residues. A 3D "basis" spectrum is collected to generate a landscape of amide resonances similar to the way the 2D HSQC provided the initialization step from the traditional methods outlined above. Each peak in this spectrum is then used as the basis for the 4D and 5D experiments, which are collected as 2D slices.

Generally, the highly efficient relaxation pathways of folded globular proteins make direct ^{13}C-detect experiments inefficient. However, IDPs and spider silk proteins in solution suffer from much less signal loss due to much longer relaxation times. Nevertheless, even with beneficial relaxation properties, the significantly lower gyromagnetic ratio of ^{13}C and lower labeling efficiency of spider silk proteins necessitates the use of a cryogenically cooled probe.

Despite a lower gyromagnetic ratio, ^{15}N direct-detect methods compliment the ^{13}C direct-detect suite of experiments and have several unique advantages. They avoid the splitting resulting from strong one-bond $J_{^{13}C-^{13}C}$ couplings, which require the collection of two interleaved spectra in the in-phase antiphase (IPAP) pulse sequence[71]. Nitrogen experiments, acquired as a single spectrum, are therefore often more straightforward to collect than ^{13}C experiments in a uniformly labeled sample. The chemical shift dispersion of ^{15}N is also slightly greater than the aliphatic or carbonyl carbon chemical shift regions. Presently, ^{15}N-detect versions of two of the most informative experiments, CON and CAN, have been demonstrated to have great value

in structural analysis of intrinsically disordered proteins[72]. These two experiments form the basis of a sequential assignment strategy similar to the HNCACB/CBCACONH one used in more traditional biomolecular NMR. In this modified version, the CAN experiment correlates backbone amide ^{15}N nuclei with both the sequential *i* Cα as well as the preceding *i*-1 Cα. This experiment exists as either protonless or proton-start versions (hCAN)[73]. These proton-start versions have the benefit of reducing the acquisition time by using a traditional INEPT sequence to transfer polarization from the Hα to the Cα. Taken all together, these direct-detect experiments form an elegant method for probing dynamic protein structures that elude nearly all other methods of direct investigation.

References

1. Seib, F. P., Self-assembling hydrogels from reverse-engineered silk. In *Self-assembling Biomaterials*, Elsevier: 2018; pp 27-47.

2. 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.

3. 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. *Journal of Biomedical Materials Research Part B: Applied Biomaterials* **2011,** *97B* (2), 381-387.

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

5. 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).

6. 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*, e11234.

7. 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). *Molecular Biology and Evolution* **2013,** *30*, 589-601.

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

9. Hu, X.; Lawrence, B.; Kohler, K.; Falick, A. M.; Moore, A. M. F.; McMullen, E.; Jones, P. R.; Vierra, C., Araneoid egg case silk: A fibroin with novel ensemble repeat units from the black widow spider, Latrodectus hesperus. *Biochemistry* **2005,** *44*, 10020-10027.

10. 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.

11. Hieber, C. S., The “insulation” layer in the cocoons of Argiope aurantia (Araneae: Araneidae). *Journal of Thermal Biology* **1985,** *10* (3), 171-175.

12. Hieber, C. S., Spider cocoons and their suspension systems as barriers to generalist and specialist predators. *Oecologia* **1992,** *91* (4), 530-535.

13. 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. *Molecular biology and evolution* **2004,** *21*, 1950-9.

14. Barbara A. Lawrence, †; Craig A. Vierra, a.; Moore‡, A. M. F., Molecular and Mechanical Properties of Major Ampullate Silk of the Black Widow Spider, Latrodectus hesperus. **2004**.

15. 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. *Journal of Proteome Research* **2015,** *14*, 4223-4231.

16. Collin, M. A.; Clarke, T. H.; 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. *International Journal of Biological Macromolecules* **2018,** *113*, 829-840.

17. 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. *Scientific Reports* **2019,** *9* (1), 8380.

18. Hagn, F.; Thamm, C.; Scheibel, T.; Kessler, H., pH-Dependent Dimerization and Salt-Dependent Stabilization of the N-terminal Domain of Spider Dragline Silk-Implications for Fiber Formation. *Angewandte Chemie International Edition* **2011,** *50*, 310-313.

19. Jaudzems, K.; Askarieh, G.; Landreh, M.; Nordling, K.; Hedhammar, M.; Jörnvall, 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. *Journal of Molecular Biology* **2012,** *422*, 477-487.

20. 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.

21. Brooks, A. E.; Steinkraus, H. B.; Nelson, S. R.; Lewis, R. V., An Investigation of the Divergence of Major Ampullate Silk Fibers from Nephila clavipes and Argiope aurantia. *Biomacromolecules* **2005,** *6* (6), 3095-3099.

22. Marhabaie, M.; Leeper, T. C.; Blackledge, T. A., Protein Composition Correlates with the Mechanical Properties of Spider (Argiope trifasciata) Dragline Silk. *Biomacromolecules* **2014,** *15* (1), 20-29.

23. Chaw, R. C.; Zhao, Y.; Wei, 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 evolutionary biology* **2014,** *14*, 31.

24. Blackledge, T. A.; Swindeman, J. E.; Hayashi, C. Y.; Savage, K. N., Quasistatic and continuous dynamic characterization of the mechanical properties of silk from the cobweb of the black widow spider Latrodectus hesperus. *The Journal of experimental biology* **2005,** *208*, 1937-49.

25. Gosline, J. M.; Guerette, P. A.; Ortlepp, C. S.; Savage, K. N., The mechanical design of spider silks: from fibroin sequence to mechanical function. *Journal of Experimental Biology* **1999,** *202*, 3295-3303.

26. Blackledge, T. A.; Hayashi, C. Y., Silken toolkits: biomechanics of silk fibers spun by the orb web spider Argiope argentata (Fabricius 1775). *The Journal of experimental biology* **2006,** *209*, 2452-61.

27. Vasanthavada, K.; Hu, X.; Falick, A. M.; La Mattina, C.; Moore, A. M. F.; 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. *Journal of Biological Chemistry* **2007,** *282*, 35088-35097.

28. Madsen, B.; Vollrath, F., Mechanics and Morphology of Silk Drawn from Anesthetized Spiders. *Naturwissenschaften* **2000,** *87*, 148-153.

29. Guinea, G. V.; Perez-Rigueiro, J.; Plaza, G. R.; Elices, M., Volume constancy during stretching of spider silk. *Biomacromolecules* **2006,** *7* (7), 2173-7.

30. Blackledge, T. A.; Zevenbergen, J. M., Condition-dependent spider web architecture in the western black widow, Latrodectus hesperus. *Animal Behaviour* **2007,** *73*, 855-864.

31. Madsen, B.; Shao, Z. Z.; Vollrath, F., Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. *International Journal of Biological Macromolecules* **1999,** *24*, 301-306.

32. Swanson, B. O.; Blackledge, T. A.; Beltran, J.; Hayashi, C. Y., Variation in the material properties of spider dragline silk across species. *Applied Physics A* **2006,** *82*, 213-218.

33. Madurga, R.; Plaza, G. R.; Blackledge, T. A.; Guinea, G. V.; Elices, M.; Pérez-Rigueiro, J., Material properties of evolutionary diverse spider silks described by variation in a single structural parameter. *Scientific Reports* **2016,** *6*, 18991.

34. Blackledge, T. A.; Cardullo, R. A.; Hayashi, C. Y., Polarized light microscopy, variability in spider silk diameters, and the mechanical characterization of spider silk. *Invertebrate Biology* **2005,** *124*, 165-173.

35. 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.

36. Guinea, G. V.; Elices, M.; Pérez-Rigueiro, J.; Plaza, G., Self-tightening of spider silk fibers induced by moisture. *Polymer* **2003,** *44* (19), 5785-5788.

37. 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. *Journal of Experimental Biology* **2009,** *212* (13), 1981.

38. Pérez-Rigueiro, J.; Elices, M.; Guinea, G. V., Controlled supercontraction tailors the tensile behaviour of spider silk. *Polymer* **2003,** *44*, 3733-3736.

39. Simmons, A.; Ray, E.; Jelinski, L. W., Solid-State 13C NMR of Nephila clavipes Dragline Silk Establishes Structure and Identity of Crystalline Regions. *Macromolecules* **1994,** *27* (18), 5235-5237.

40. Asakura, T.; Suzuki, Y.; Nakazawa, Y.; Holland, G. P.; Yarger, J. L., Elucidating silk structure using solid-state NMR. *Soft Matter* **2013,** *9*, 11440.

41. Asakura, T.; Suzuki, Y.; Nakazawa, Y.; Yazawa, K.; Holland, G. P.; Yarger, J. L., Silk structure studied with nuclear magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2013,** *69*, 23-68.

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. *Journal of the American Chemical Society* **2008,** *130*, 9871-9877.

43. Warwicker, J. O., Comparative studies of fibroins: II. The crystal structures of various fibroins. *Journal of Molecular Biology* **1960,** *2* (6), 350-IN1.

44. Sampath, S.; Isdebski, T.; Jenkins, J. E.; Ayon, J. V.; Henning, R. W.; Orgel, J. P. R. O.; Antipoa, O.; Yarger, J. L., X-ray diffraction study of nanocrystalline and amorphous structure within major and minor ampullate dragline spider silks. *Soft Matter* **2012,** *8* (25), 6713-6722.

45. Marie-Eve Rousseau; Thierry Lefèvre; Lilyane Beaulieu; Tetsuo Asakura, a.; Michel Pézolet*, Study of Protein Conformation and Orientation in Silkworm and Spider Silk Fibers Using Raman Microspectroscopy. **2004**.

46. Porter, D.; Vollrath, F.; Shao, Z., Predicting the mechanical properties of spider silk as a model nanostructured polymer. *The European Physical Journal E* **2005,** *16*, 199-206.

47. Papadopoulos, P.; Sölter, J.; Kremer, F., Hierarchies in the structural organization of spider silk—a quantitative model. *Colloid and Polymer Science* **2009,** *287* (2), 231-236.

48. 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-e998.

49. 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 β-turn structure in spider dragline silk. *Chem Commun* **2010,** *46*, 6714.

50. Xu, L.; Tremblay, M. L.; Meng, Q.; Liu, X. Q.; Rainey, J. K., 1H, 13C and 15N NMR assignments of the aciniform spidroin (AcSp1) repetitive domain of argiope trifasciata wrapping silk. *Biomolecular NMR Assignments* **2012,** *6*, 147-151.

51. Tremblay, M.-L.; Xu, L.; Lefèvre, T.; Sarker, M.; Orrell, K. E.; Leclerc, J.; Meng, Q.; Pézolet, M.; Auger, M.; Liu, X.-Q.; Rainey, J. K., Spider wrapping silk fibre architecture arising from its modular soluble protein precursor. *Scientific Reports* **2015,** *5*, 11502.

52. 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*, e50227.

53. 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.

54. Hijirida, D. H.; Do, K. G.; Michal, C.; Wong, S.; Zax, D.; Jelinski, L. W., 13C NMR of Nephila clavipes major ampullate silk gland. *Biophysical Journal* **1996,** *71* (6), 3442-3447.

55. 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*, 3879-3885.

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

57. 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*, e514.

58. Vollrath, F.; Knight, D. P., Liquid crystalline spinning of spider silk. *Nature* **2001,** *410*, 541-548.

59. Knight, D. P.; Vollrath, F., Liquid crystals and flow elongation in a spider's silk production line. *Proceedings of the Royal Society B: Biological Sciences* **1999,** *266*, 519-523.

60. Bodenhausen, G.; Ruben, D. J., Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **1980,** *69*, 185-189.

61. Mori, S.; Abeygunawardana, C.; Johnson, M. O.; van Zijl, P. C., Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J Magn Reson B* **1995,** *108* (1), 94-8.

62. Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V., Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity. *Journal of Magnetic Resonance, Series A* **1993,** *102* (2), 241-245.

63. Piotto, M.; Saudek, V.; Sklenar, V., Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **1992,** *2* (6), 661-5.

64. Meissner, A.; Sørensen, O. W., Sequential HNCACB and CBCANH Protein NMR Pulse Sequences. *J Magn Reson* **2001,** *151*, 328-331.

65. Wittekind, M.; Mueller, L., HNCACB, a 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, Series B* **1993,** *101* (2), 201-205.

66. Muhandiram, D. R.; Kay, L. E., Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *Journal of Magnetic Resonance, Series B* **1994,** *103* (3), 203-216.

67. Grzesiek, S.; 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* (16), 6291-6293.

68. Grzesiek, S.; Bax, A., Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *Journal of Biomolecular NMR* **1993,** *3* (2), 185-204.

69. Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A., Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry* **1992,** *31*, 5269-5278.

70. Kay, L. E.; Torchia, D. A.; Bax, A., Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **1989,** *28*, 8972-8979.

71. Takeuchi, K.; Heffron, G.; Sun, Z. Y.; Frueh, D. P.; Wagner, G., Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments. *J Biomol NMR* **2010,** *47* (4), 271-82.

72. Chhabra, S.; Fischer, P.; Takeuchi, K.; Dubey, A.; Ziarek, J. J.; Boeszoermenyi, A.; Mathieu, D.; Bermel, W.; Davey, N. E.; Wagner, G.; Arthanari, H., (15)N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins. *Proc Natl Acad Sci U S A* **2018,** *115* (8), E1710-E1719.

73. Gal, M.; Edmonds, K. A.; Milbradt, A. G.; Takeuchi, K.; Wagner, G., Speeding up direct (15)N detection: hCaN 2D NMR experiment. *J Biomol NMR* **2011,** *51* (4), 497-504.

Chapter 2 Hierarchical Spidroin Micellar Nanoparticles as the Fundamental Precursors of Spider Silks

Introduction

The ability to replicate the natural silk spinning process with an aqueous synthetic analogue at bulk scale, to truly mimic the properties of natural silks, holds tremendous promise for biomedical materials, architectural design, and civil and mechanical engineering. However, to date, no synthetic silks have been produced that exhibit the mechanical properties of the natural product.[5–8] Currently, we know the primary sequence of the natural spider silk spidroins, and we have significant information on the biochemical triggers of the spinning process and how to mimic it.[2, 4, 11, 12, 15–17] The gap in our knowledge involves the nanoscale processes at work within the silk gland where the highly-concentrated protein precursor is stored and then on-demand transformed into silk fibers.[2, 4, 11, 12, 18–20] Knowledge of this central part of the natural process is critical to the development of synthetic analogues and key to preparative methods including microfluidics for producing artificial fibers from recombinant proteins that exhibit the mechanical properties of native silks.[1, 4, 6–8, 21] Beyond spider silks, analogous mechanisms might be involved in the formation of protein fibers generally, such as the detrimental nucleation and growth of amyloid fibers from oligomers.[9, 22, 23] Critical insights gleaned for one fibrous protein system can potentially be translated more generally to further our understanding of how concentrated proteins are stored and then assembled to yield structurally organized 3D materials.

The dragline (major ampullate, MA) silk-precursor dope solution of *Latrodectus hesperus*, or Black Widow, spiders is predominantly composed of two large repetitive proteins (spidroins) that are stored in high concentration[24–26] (25-50 wt%) in the glands prior to extrusion, acidification and fiber formation in the duct.(2) These spidroins, *MaSp1* and *MaSp2*, are large, highly repetitive proteins (~300-400 kDa). In the gland environment, the repetitive core regions of the spidroins are predominantly unstructured [24–27] (random coil), while the termini exhibit pH-sensitive helical bundles.[12, 19, 28] The pH sensitivity of the termini has been implicated as an important characteristic for the assembly of spider silk.[12, 19, 28, 29] The hydropathy plots for the spidroins are roughly sinusoidal in form, with rapidly alternating hydrophilic-hydrophobic units (*c.a.* +0.3 to -0.6, **Figure. 21**).[16] The C- and N-terminal domains contain the units with the highest hydrophilicity in each oligomer, while the long, central, repetitive domain is generally neutral and hydrophobic, giving these proteins an amphiphilic character analogous to a complex block copolymer amphiphile.[30, 31] This amphiphilicity suggests, according to the micelle theory of silk assembly,[11, 12] that the concentrated protein dope solution is composed of nanoscale assemblies;[32] the essential starting structures needed for the formation of robust, microscale silk fibers.[11, 12] Microscopic characterization of synthetically formed silkworm fibroin/PEO fibers has shown evidence of spherical micelle-like structures on the micron scale,[11] Similarly, scanning electron microscopy (SEM) and atomic force microscopy (AFM) imaging of the fully and partially dried silk gland dope from *Nephilia clavata* spiders has found the existence of micron-sized granule particles. Intriguingly, these appear to be present within in the hierarchical structure of the spiders' silk fibers themselves.[33] However, no direct experimental

evidence from the hydrated native protein dope solution has been provided in support of the micelle theory. The true physical form of the protein assemblies when stored in the gland and the process by which this dope solution is transformed into a hierarchical polycrystalline structure when spun through the duct remains largely unknown.

Here we analyze the native protein dope from the silk glands of Black Widow spiders to determine the physical form of the liquid precursor of natural silk fibers, and test the validity of the existing micelle theory.[11, 12] We use a combination of indirect and direction observation methods for nanoscale characterization, applying solution NMR spectroscopy and cryo-TEM tomographic imaging. Measuring the diffusion behavior of the spidroins in the natural silk gland dope by NMR reveals that the spidroin proteins in the silk precursor are predominantly confined within entangled volumes of several-hundred nm diameters, suggesting the existence of tightly-packed spherical micelles, as postulated in the existing micelle theory.[11, 12] However, 3D imaging of these nanostructures by cryo-TEM tomography shows that the spidroin micellar assemblies are far more morphologically complex, existing as hierarchical micellar nanoparticles (several hundred nm diameters) composed of networks of flake-like subdomains. When the native dope is physically sheared, mimicking the extrusion process occurring through a spider's spinning duct, the subdomains within the hierarchical micelles transform, become narrowed and elongated fibrils that remain aggregated as interwoven networks.

Methods

Dissection of *L. hesperus* (Black Widow) Spider and Extraction of MA Silk Glands. Large Western Black Widows (*Latrodectus hesperus*) were fed approximately

100 µl of saturated U-^{13}C-^{15}N Alanine solution to label the silk proteins (*MaSp1* and *MaSp2*) with NMR-accessible nuclei for heteronuclear solution NMR experiments. Labeling was done three times a week for two weeks. While feeding, the spiders were forcibly silked to deplete the native, unlabeled spidroin supply and encourage the synthesis of labeled proteins. At the end of two weeks the spiders were dissected and the major ampullate glands were extracted in deionized water. For the studies on native silk solution, major ampullate glands were rinsed with DI water and transferred directly to a 5 mm Shigemi tube filled with 90:10 H_2O:D_2O. For the urea studies, glands were carefully removed, rinsed, and cleaned of the outer, insoluble membrane. This silk dope was then diluted with a solution of 4M urea with 1% sodium azide, and allowed to incubate at 4 ˚C for 1 hour. The final solution was then either transferred to 5 mm NMR tube for analysis, or prepared for cryo-TEM or DLS.

Cryo-TEM Sample Preparation. *L. hesperus* spiders were asphyxiated with N_2-gas and quickly sacrificed with a sharp blade. Spiders were then dissected to remove the major ampullate gland (see dissection methods for details) and to manually remove the outer membrane and clean the gland (done at ambient conditions, ~27˚ C). This process of dissection and gland removal and cleaning took ~35 min for the gland sample used in **Figure 17** and **Figure 18** (these two are from the same gland sample prepared on the same day), but only took ~15 min for the gland sample in **Figure 23** (for this gland dissection and cryo-TEM sample preparation was done using a different spider than for **Figures 17, 18**, and on a different day). For both glands, the surface of the gland was cleaned very thoroughly of membrane without perturbing or rupturing the dope. The cleaned gland was immediately added to a vial (100 µL) of urea denaturant

solution, and 10 min later the gland-urea solution was transferred to a 4 ˚C refrigerator. During this time, the mass of the solution with the gland was measured to determine approximate solution concentration (~1 wt%). For the gland used in **Figure 17** and **Figure 18**, the gland-urea solution was incubated at 4 ˚C for ~240 min. For the gland used in **Figure 23**, the gland-urea solution was incubated at 4˚ C for ~175 min. After this incubation time at 4 ˚C, gland-urea solution was immediately transferred to a 4 ˚C (cold room) cryo-TEM grid preparation lab.

For all samples (**Figures 17, 18** and **Figure 23**), graphene-oxide coated holey-carbon (Quantifoil R2/2, 400 mesh Cu support) TEM grids (Structure Probe, Inc., West Chester, PA) were used. The graphene-oxide coating was applied according to refs. [41, 42] using a 0.15 mg/mL aqueous solution (drop-cast 4 µL onto plasma-cleaned grid), and the coating was prepared several days prior to the grids being used for cryo-TEM sample prep.

After incubation, cryo-TEM sample preparation and plunging was done in a zero-humidity controlled cold room (4 ˚C). For the preparation of the “Native MA protein” samples (**Figure 17** and **Figure 23**) no micropipette was used, as would be done for conventional cryo-TEM sample prep. Instead, a borosilicate Pasteur pipette (#13-678-20B, Fischer Scientific International, Inc., Hampton, NH) was used to transfer a small volume of the sample solution from the vial to the surface of the GO-coated TEM grid. By eye, the volume of the droplet applied was ~4-5 µL of solution, applied to the coated surface of the grid while the grid was held aloft in self-clamping tweezers. The tweezers were then quickly, but carefully, attached into a home-made manual plunger, with a small liquid-ethane bath set below the plunging apparatus. Using filter paper, the grid

with the droplet on its surface was manually blotted for a 4 s count, applying the filter paper to the coated-side of the grid (the side with the droplet). The grid was immediately plunged into the *L*-ethane bath after the 4 s blot, and the grid was then transferred into a *L*-N_2 bath where it was placed into a storage box. The storage box was transferred into a large *L*-N_2 Dewar storage container and stored.

For the preparation of the “Native” MA protein after vigorous micropipette pumping sample (**Figure 18**), conventional preparation was done using a micropipette. However, prior to drop-casting the droplet of solution onto the GO-coated surface of the grid, the micropipette (20 µL volume micropipette, set to 4 µL fill) was used to vigorously pump in-and-out on the sample solution 40 times. A 4 µL droplet was then applied to the coated surface of the grid while the grid was held aloft in self-clamping tweezers. The grid was left to sit undisturbed for ~5 min, and then the tweezers were carefully attached into the home-made manual plunger, with a small liquid-ethane bath set below the plunging apparatus. Using filter paper, the grid with the droplet on its surface was manually blotted for a 4 s count, applying the filter paper to the coated-side of the grid (the side with the droplet). The grid was immediately plunged into the *L*-ethane bath after the 4 s blot, and the grid was then transferred into a *L*-N_2 bath where it was placed into a storage box. The storage box was transferred into a large *L*-N_2 Dewar storage container, and stored.

Cryo-TEM Imaging and Tomography Tilt Series Acquisition. All cryo-TEM imaging (**Figures 16, 17** and **Figures 23, 25, 26**) was done using a FEI (FEI Company, Hillsboro, OR) Polara cryo-TEM operating at 200 keV equipped with a Gatan (Roper Technologies, Sarasota, FL) K2 “direct electron” detector (FEG extraction: 4350 V, FEG

emission: 110 µA, spot size 5, 70 µm OL aperture, 70 µm CL aperture). The Polara cryo-TEM is a cartridge-loaded microscope, where the column is continually maintained at *L*-N_2 temperature (< -175 ˚C). The grids that were prepared previously and stored in *L*-N_2, were loaded into the microscope's cartridge while submerged in *L*-N_2, and the cartridge with grids was inserted into the microscope. The Polara was aligned for low-dose imaging, measuring the dose rate on the K2 detector through vacuum (no grid inserted). The dose rate used was 1.25 $e^-/Å^2s$ for high-magnification (50 kx) images (4.096 Å pixel size, 3708 x 3838 pixels, **Figure 17*B-E***, **Figure 18*B-E***, and **Figure 23*B-D***). The dose rate was not measured for the low-mag images (**Figure 17*A***, **Figure 18*A***, and **Figure 23*A***), as the dose rate was orders of magnitude lower for those images (1 s exposure time, 10.1034 nm pixel size, 1852 x 1918 pixels, binning 2). The high-mag images associated with tomography tilt series (**Figure 17*C-E*** and **Figure 18*C-E***) were acquired using a 2 s exposure time (cumulative dose: 2.5 $e^-/Å^2$ per image), and all other high-mag images (not part of tilt-series, **Figure 17*B***, **Figure 18*B***, and **Figure. 23*B-D***) were acquired using a 3 s exposure time (cumulative dose: 3.75 $e^-/Å^2$ per image). All imaging and tomography tilt-series acquisition was done using SerialEM software (http://bio3d.colorado.edu/SerialEM/), which applies auto focusing on adjacent regions of the grid to minimize dose on the sample, and uses an automated cross-correlation drift/shift correction during acquisition of tomography tilt series. Raw cryo-TEM images were saved a MRC stacks.

All cryo-TEM tomography tilt-series were acquired using 3˚ tilt-intervals. For the tilt-series in **Figure 17*C-H***, the range covered was +49˚ to -59 (37 images in the series), and the total cumulative dose experienced was ~92.5 $e^-/Å^2$. For the tilt-series in **Figure**

18*C-H*, the range covered was +38˚ to -58 (33 images in the series), and the total cumulative dose experienced was ~82.5 $e^-/Å^2$.

Tomography Reconstruction of cryo-TEM Tilt-Series and 3D Rendering/Visualization. The raw cryo-TEM tilt series images (MRC stack) were manually aligned **Figure 24**) using IMOD software (http://bio3d.colorado.edu/imod/).[43] The negative log of the TEM intensities were used to form a pseudo dark-field image and convert the intensity of the object into a positive value, with the background intensity close to zero. The TVR-DART tomography reconstruction algorithm [37, 44] was downloaded from GitHub (https://github.com/astra-toolbox/ContributedTools), and was implemented to generate the reconstructions of the protein nanostructures in **Figure 17*F-H*** and **Figure 18*F-H***. A value of 50 was used for the "lambda" parameter to reduce noise contributions and 100 iterations of the SIRT (simultaneous iterative reconstruction technique) algorithm was used with three grey levels (object, background and vacuum). After reconstruction, a 3D Gaussian filter *was* applied to further reduce noise and remove isolated voxels and the objects were globally thresholded and visualized using FEI Avizo 9.0 software (FEI Company, Hillsboro, OR). For **Figure 17**: "sigma rel." of 1, kernel size of 7, and threshold value of 80 after the Gaussian filter was applied. For **Figure 18**: "sigma rel." of 1, kernel size of 5, and threshold value of 100 after the Gaussian filter was applied. Threshold values were applied using the magic wand tool, which only selects connected pixels. **Figure 17*F*** and **Figure 18*F*** are isosurface renderings of the reconstructions, and **Figure 17*G,H*** and **Figure 18*G,H*** are 1-voxel thick orthoslice images (0.8 nm thick) from the reconstruction at various *z*-heights.

Magnetic Resonance Imaging (MRI) of Black Widow Spider. Gradient echo (GRE)-MRI images of a Black Widow spider (**Figure 16*A***) were collected using an 18.8T magnet (800 MHz) with a Varian VNMRS console and Doty Scientific 12 mm ID triple axis MRI gradient probe capable of producing up to 300 Gs/cm PFG along each axis. GRE-MRI images with 40 µm x 40 µm in plane resolution and 300 µm slice thickness were collected using a 4 ms echo time and 50 ms repetition time. Localized spectra utilized the STEAM2 localization scheme, TE = 2.3 ms, TM = 4.5 ms, and TR = 2 s. Additionally, outer voxel suppression (OVS) pulses and VAPOR water suppression was employed.[45] A 1.6 uL volume of interest (VOI) was selected in the interior of the MA gland using pilot gradient echo images.

PFG-STE NMR Acquisition and Analysis. NMR diffusion experiments (**Figure 17** and **Figures. 21**, **27**) were conducted using the same magnet as used for GRE-MRI imaging (18.8T), and with an 800 MHz Agilent spectrometer and a Doty diffusion probe using a stimulated echo (STE) pulse sequence with bipolar pulse-field gradient pulses lasting 2 ms (δ) with a 50 ms Δ diffusion delay and a 10 s d1 recycle delay.[46, 47] 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. The diffusion experiments were collected by incrementing the gradient magnetic field strength in ten steps. The probe gradients were calibrated using anhydrous glycerol sample. The calibrated temperature was 26 ˚C and the corresponding diffusion coefficient for glycerol is 2.50×10^{-12} m^2/s.[46] The gradient pulse (δ) was set to 2 ms and the delay (Δ) between two bipolar gradient pulses was 50 ms. 20 experiments were collected with an array of field strengths from 6.4 Gs/cm to 1273 Gs/cm.

For ^{1}H diffusion measurements on silk proteins, the experiments were carried out at 23 ˚C. The general experimental parameters used were: spectrum width of 9615 Hz, 32 transients, 10 s of recycle delay, 2.45 s of acquisition time, δ = 1 ms, Δ = 50 ms and gradient field strengths arrayed from 7 Gs/cm to 600 Gs/cm. When arraying the diffusion delay, Δ, all other parameters were kept constant with Δ ranging from 25 ms to 500 ms with 25 ms increments. The diffusion coefficient change with time in 4M urea was tracked for 36 hours with 2 hour sampling intervals and following one weeks' time. One dimensional (1D) ^{1}H NMR spectra from diffusion experiments were assigned based on previous results [26, 27] and the ^{1}H peaks from Ala H$ß$, Gln Hγ, Gly Hα and Tyr Hδ were selected for diffusion calculations as they can be easily distinguished from other peaks. The diffusion coefficient was determined using the following equation:

$$ln[S(2\tau)] = ln[S(0)] - \gamma^2\delta^2 D(\Delta - \delta/3)G^2$$

where $S(2\tau)$ is the signal intensity after the second bipolar gradient pulse and $S(0)$ is the theoretical signal intensity before applying the first bipolar gradient pulse, γ is the gyromagnetic ratio of ^{1}H, D is the translational diffusion coefficient and G is the gradient pulse strength.[46] From a plot of the natural log of the peak integral ($ln[S(2\tau)]$) vs. the corresponding gradient strength squared (G^2), the diffusion coefficient (D) can be determined from the slope of the linear regression. The error in the determined *D* is +/- 0.25 x 10^{-8} cm^2/s. Single component and multicomponent fits were attempted (**Figure 27**). NMR spectra were processed using Vnmrj 4.2 and exported as ASCII data format. The ASCII data was then analyzed using home-made scripts in Matlab 2013b.

The hydrodynamic radius of the silk protein in 4 M urea was estimated using the Stokes-Einstein equation, $D = kT/6\pi\eta r_H$, where k is the Boltzmann constant, η is the

viscosity of 4 M urea (1.085×10^{-4} Pa· s) and r_H is the hydrodynamic radius.[48] The hydrodynamic radius was also estimated from a semi-empirical equation based on residue number, $r_H = (2.21 \pm 1.07)N^{0.57\pm0.02}$.[36] The mean square displacement (MSD) is estimated from $\sqrt{2D\Delta}$.

2D ^{1}H/^{15}N heteronuclear single quantum correlation (HSQC) NMR experiments (**Figures. 22, 28**) were carried out with a 600 MHz spectrometer using a Bruker TXI probe controlled by a Bruker Avance III console at room temperature. 2k points were collected in F2, zero-filling up to 4k, and 128 points up to 512 used for F1. 32 scans were collected per slice. A line broadening function of 0.3 Hz was applied to the spectrum in the F2 dimension. Phasing and data processing was carried out in Topspin 3.5. All chemical shifts are referenced to DSS.

Dynamic Light Scattering (DLS). DLS (**Figure 24**) 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 viscosity of 4M urea was previously determined to be 1.0854cP with a refractive index of 1.364 determined with a refractometer. The data was fitted using the protein analysis model for data processing.

Hydropathy Plot Generation (**Figure 20**). The hydropathy plots for the MaSp1 (**Figure 20*A***) and MaSp2 (**Figure 20*B***) spidroins were created using the respective protein sequences in ref. [16]. The hydropathy index was plotted for each individual amino acid using the ExPASy web resource (www.expasy.org).

Results

Solution NMR Indicates Gland-Stored Spidroin Diffusion Confinement and Entanglement within ~300 nm Diameter Volumes. Using pulsed field gradient stimulated-echo nuclear magnetic resonance (PFG-STE NMR) (**Figure 16, Figure 21**), we measured the Einstein self-diffusion coefficient (*D*) and the observation time (Δ, inter-pulse delay) dependence of native MA silk gland dope from *L. hesperus* spiders (**Figure 16*A***). This NMR technique probes the average temporal displacement of an ensemble of molecules, from which *D* can be determined (*see Methods*). The measured displacement and *D* of a heterogeneous sample, for example, a solution containing micelles, provides information about the physical constraints on motion, distances and metrics, arrangement of molecules, and structural permeability.[34, 35] The diffusion of the native gland protein shows a strong Δ dependency, evidenced by the initial decay in **Figure 16*B***, behavior that indicates restricted diffusion of the silk protein in the gland. Assuming bulk diffusion, the mean squared displacement (MSD) that encloses the volume of restricted diffusion is ~300 nm, suggesting the proteins are confined/entangled within structural elements on this length scale. Once denatured, the diffusion of the protein (48 hr in urea) no longer exhibits a Δ dependence (**Figure 16*C***), indicating that the denatured solution is homogeneous and free of entanglement or diffusion restriction. By sampling multiple protein concentrations in urea and extrapolating to infinite dilution (**Figure 16*D***), we find a self-diffusion coefficient of ~10.7 x 10^{-8} cm^2/s corresponding to a hydrodynamic radius (r_H) of ~19 nm (Methods). This experimental r_H is consistent with the theoretical values for the *MaSp1* protein (~22 nm) based on number of residues and chain length,[36] indicating that the diffusion restriction was removed following treatment in denaturant after 48 hr. These diffusion NMR

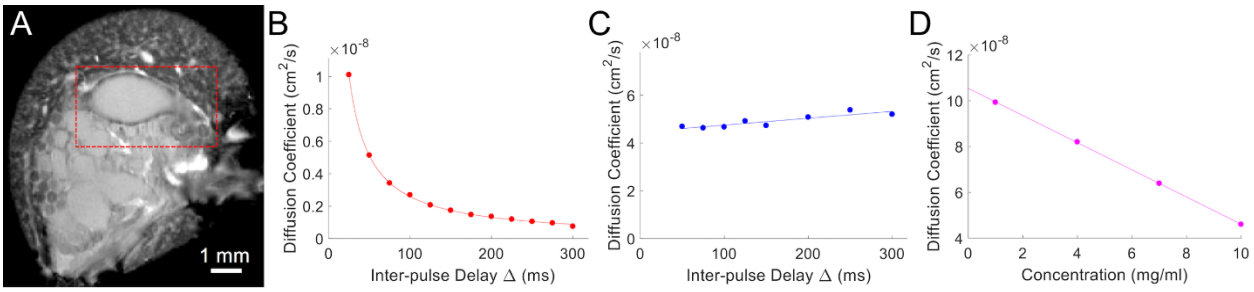


Figure 16 Native *L. hesperus* MA silk protein diffusion measured by PFG-STE NMR. (A) Gradient echo magnetic resonance image (GRE-MRI, 18.8 T) of 300 µm thick cross-section slice of the L. hesperus spider abdomen at the coronal cross-section orientation. Red box indicates one of the two MA glands, which were removed by dissection from *L. hesperus* spiders for all subsequent NMR, DLS, and cryo-TEM analysis. (B-D) PFG-STE NMR data. (B) Self-diffusion coefficient (D) vs inter-pulse delay (Δ) for MA silk protein at native gland conditions (conc. ~ 35 wt%, pH ~ 7). Strong Δ dependence is observed, and the decreasing D as a function of Δ illustrates that diffusion is restricted at native conditions (MSD ~300 nm). (C) D vs Δ for the MA silk protein dope following dilution in 4M urea (conc. ~ 10 wt%, 48 hr in urea). Δ dependence is no longer observed, indicating diffusion is not restricted following solubilizing in urea. (D) MA Silk protein D measured by PFG-STE NMR as a function of protein concentration in 4 M urea (48 hr). The extracted self-diffusion coefficient at infinite dilution is 10.7 x 10-8 cm2/s (rH ~19 nm

measurements provide strong evidence that under native conditions in the MA gland, spidroin proteins exist as entangled assemblies several hundred nanometers in diameter. These structures are disrupted and broken down to freely diffusing proteins that remain unstructured (**Figure. 22**) when fully denatured in urea for 48 hrs. We note that for short urea incubation times on the order of 4-10 hrs (**Figure 21**), restricted diffusion similar to the native dope is found.

Direct Morphological Characterization of the Entangled Spidroin Nanostructures by Cryo-TEM Tomography Reveals Hierarchical Micellar Nanoparticles. We aimed to directly image and characterize the morphology of the entangled spidroin protein nanostructures in the native silk dope using cryogenic transmission electron microscopy (cryo-TEM, **Figure 17**). To prepare cryo-TEM grids (see Methods), the native dope was incubated in urea for 4 hrs to lower sample viscosity and concentration. At low magnification (**Figure 17*A***), an abundance of roughly spherical structures was

observed, with similar results found for multiple spiders and multiple samples (**Figure. 23**). The largest structures were ~800 nm in diameter (small population), and the predominant population of particles were ~200-400 nm in diameter, with similar sized structures implicated by the results from NMR diffusion and dynamic light scattering (DLS) analysis of the MA protein dope following brief urea incubation (**Figure 24**).

At high magnification (**Figure 17*B***), we found that the spherical structures are

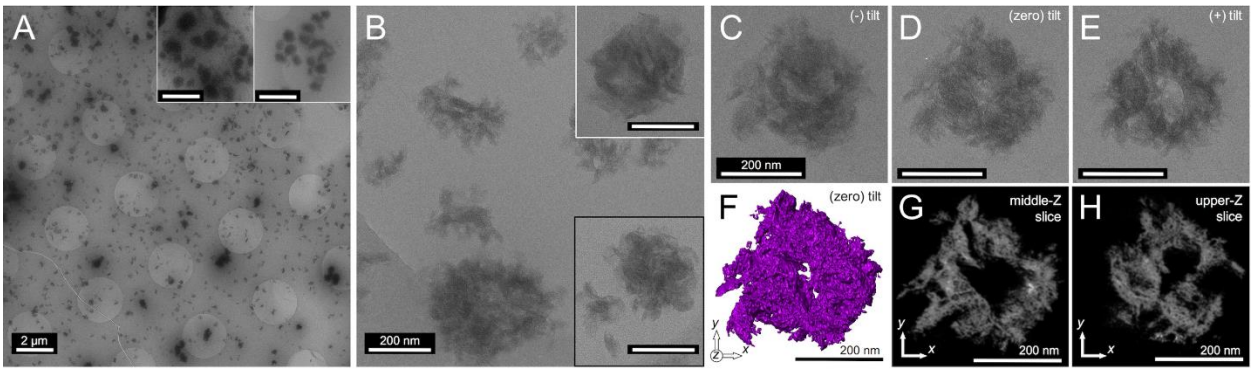


Figure 17 Fig. 10. Native L. hesperus MA silk protein imaged by cryo-TEM (4 hr incubation in 4M urea). (*A*) Low-magnification image showing abundance of generally spherical (200-400 nm) micellar protein assemblies. Insets (2 µm scale bar) highlight the small population of the largest assemblies found (~500-800 nm). (*B*) High-magnification images of the spidroin hierarchical assemblies, representative of the highest populations; spherical micellar assemblies composed of disordered "flake-like" sub-domains (200 nm scale bar for insets). Black boxed inset shows similar structures found in a sample prepared from a separate spider following the same procedure (**Figure 23**). (*C-E*) Cryo-TEM images from a tomography tilt-series of one selected hierarchical protein nano-assembly. (*F*) 3D isosurface rendering of the assembly in *C-E*. (*G,H*) ~0.8 nm thick z-slices extracted at different z-heights from the tomography reconstruction in F of the assembly in C-E

micellar assemblies having hierarchical architectures, composed of flake-like subdomains, ~50 nm laterally and ~25 nm in thickness. Many of these protein assemblies appear to contain significant solvent-filled voids within their internal volumes. The three-dimensional intricacy of these architectures, evident from initial 2D-transmission observations, led to our acquisition of cryo-TEM tomographic tilt series (**Figure 17*C-E***, **Figure 25**) to allow the reconstruction and rendering of the 3D volume of individual assemblies (**Figure 17*F-H***). Using 3˚ tilt intervals to minimize cumulative

dose, a 37 image-stack of a selected assembly was acquired, manually aligned, and 3D-reconstructed using a discrete tomography algorithm (Methods).[37] The 3D isosurface rendering (**Figure 17*F***) of the hierarchical assembly in **Figure 17*C-E***, captures the empty internal volume and disordered subdomains of that structure, features that are convoluted in the 2D transmission images. The 1-voxel thick (~0.8 nm) *z*-slices (**Figure 17*G,H***) of the reconstruction show the interconnected, and disordered nature of the flake-like subdomains within each assembly. These cryo-TEM results provide direct evidence for the existence of nanoscale micellar protein assemblies in the native silk dope, assemblies that are hierarchical structures composed of disordered subdomains, more akin to complex compound micelles than the spherical micelles previously postulated.[11, 12]

The progressive extrusion (shearing and liquid crystalline alignment) and acidification that occurs as the initial native protein dope travels down the spinning duct of the spider is believed to trigger and control the transformation and formation processes that yield the robust, polycrystalline silk fibers.[4, 11, 12, 18, 19] To probe the initial nanoscale steps of transformation that occur in the native spidroin hierarchical assemblies in response to shear, we prepared a separate set of samples for cryo-TEM analysis where the dope solution was vigorously micropipetted just prior to deposition on cryo-TEM grids (**Figure 18**). Similar to the findings for the unperturbed native dope sample (**Figure 17A**), at low magnification, an abundance of spherical, submicron spherical structures are also found for the sheared dope sample (**Figure 18A**). However, close inspection of the cryo-TEM data at higher magnifications, reveals subtle but significant differences in the architecture of the hierarchical assemblies, specifically

at the subdomain level (**Figure 18B**). The subdomains following micropipette shearing transformed from the initial flake-like morphologies (**Figure 17**) into elongated and narrow fibrillar morphologies (~50-100 nm in length and ~5-15 nm in width).

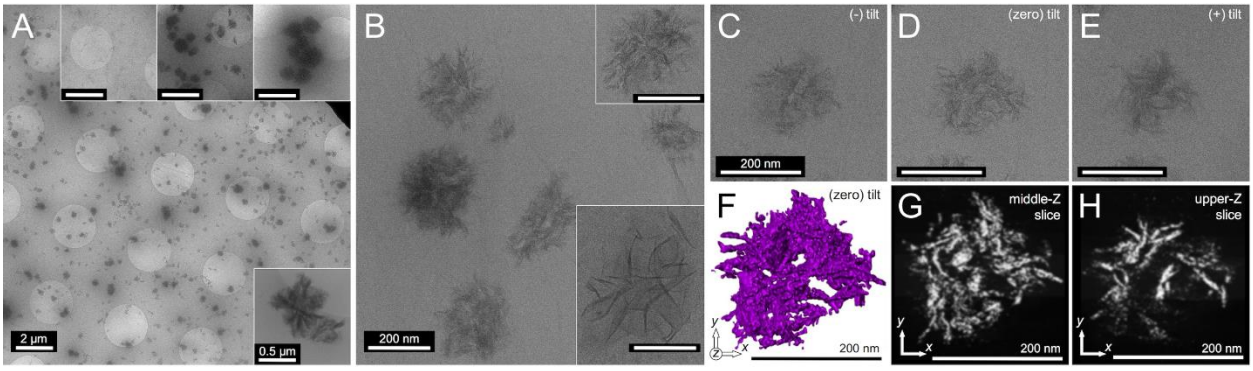


Figure 18 Native *L. hesperus* MA silk protein imaged by Cryo-TEM (4 hr incubation in 4M urea) after shear (vigorous micropipette-pumping). (*A*) Low-magnification images showing the abundance of generally spherical assemblies (100-300 nm). Insets (2 µm scale bar) highlight the small population of the largest assemblies found (~500-800 nm) and a region found with isolated fibrils (left). (*B*) High-magnification images of the spidroin assemblies, representative of the highest populations; spherical hierarchical assemblies composed of interwoven fibrillar sub-domains. Inset (200 nm scale bar) shows the smaller population of loosely-packed fibrillar aggregates. (*C-E*) Cryo-TEM images from a tilt-series of one selected hierarchical nano-assembly. (*F*) 3D Tomography rendering of the assembly in *C-E*. (*G,H*) ~0.8 nm thick z-slices extracted at different *z*-heights from the tomography reconstruction in *F* of the assembly in *C-E*.

We acquired cryo-TEM tomographic tilt series (**Figure 18*C-E***, **Figure 26**) and created 3D reconstructions and volume renderings (**Figure 18*F-H***) of the hierarchical assemblies in the sheared dope sample for comparison to those found in the native dope sample (**Figure 17**). The 3D isosurface rendering in **Figure 18*F*** shows the disordered and interwoven three-dimensional arrangement of the fibril subdomains within each spherical assembly, which also contain significant internal void volume (solvent filled). From the 1-voxel thick (~0.8 nm) *z*-slices (**Figure 18*G,H***) of the 3D reconstruction, the narrow (~10 nm) fibrillar structure of the interlaced subdomains is

visualized clearly, showing the differences between the sheared subdomain morphologies and flake-like subdomains in the hierarchical assemblies of the native [unperturbed] dope (**Figure 17*G,H***). These cryo-TEM results suggest that the initial transformations occurring as the gland-stored spidroin protein dope is extruded along

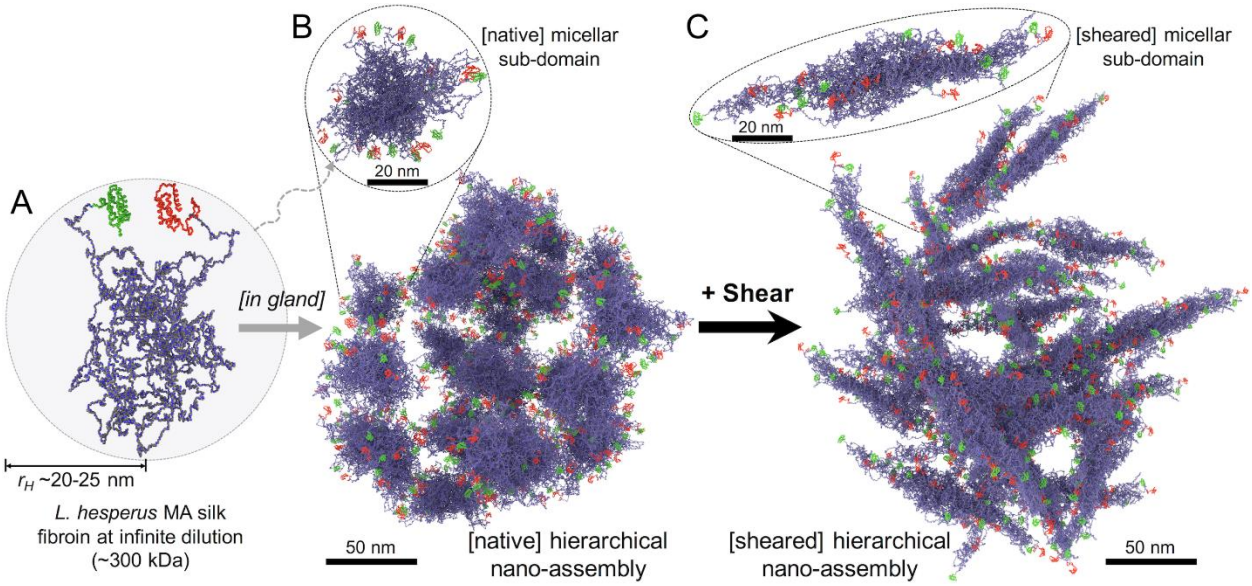


Figure 19 Graphical interpretation of the cryo-TEM and diffusion NMR data; illustrative scheme of the modified micelle theory. (*A*) Spidroin oligomer at infinite dilution (the building-blocks for the nano-assemblies that form at native gland conditions (*B*). The C- and N-terminus domains (green and red) form helical structures and the long, central domain (blue and grey) is in the molten-globule state (r_H ~20-25 nm).(12, 24–27) (*B*) At neutral pH and native concentration within the MA gland, the spidroin oligomers assemble into small, flake-like subdomains (here shown composed of 12 oligomers) that hierarchically assemble into larger, spherical micellar structures (several hundred-nm diam.). (*C*) When sheared, the initial transformations occur at the subdomain level. Individual subdomains transform into narrow, elongated fibrils that remain assembled (disordered and interwoven network) within the spherical hierarchical assemblies.

the spinning duct occur at the subdomain level, prior to transitions at the micellar level or at higher orders (**Figure 19**). Immediately following shear, the general morphologies of the micellar assemblies remain roughly spherical, of the same overall size (several hundred-nm diameters) and with disordered arrangements of subdomains. However, the subdomains themselves transition from flake-like morphologies to narrow fibril morphologies. This suggests that the complex, hierarchical architecture of native

micellar assemblies (**Figure 17**) is integral to the formation of natural-quality spider silk fibers.

Discussion

From these NMR and cryo-TEM data, we revise the existing "micelle theory" [11, 12] and propose instead a "hierarchical micelle theory" for the native spidroin storage and initial transformation, shown in **Figure 19**. At very low concentrations, the free spidroin oligomers pack into hydrodynamic radii of ~20-25 nm, balancing their hydrophobic/hydrophilic interactions with the local solvent (**Figure 4*A***). At native concentration (when stored in the MA gland), the spidroins assemble into hierarchical, spherical architectures composed of a disordered and interconnected network of flake-like subdomains with interspersed internal voids (**Figure 19*B***). The physical dimensions of the subdomains from the cryo-TEM data suggest that individual subdomains are composed of *ca.* 10-15 spidroin monomers. We propose that these hierarchical micellar assemblies are the essential starting structures critical for the formation of natural silk fibers. Further, when sheared, the individual subdomains transform into narrow and elongated fibrils that remain disordered and interwoven within the hierarchical spherical assemblies (**Figure 19*C***).

These data imply that the structural transformations (fibrillization) at the subdomain level within the hierarchical micelles are critical first steps during silk fiber formation in the natural silk spinning system, prior to transitions, liquid crystalline organization, and/or assembly/disassembly events at the micellar level. Nonclassical two-step growth processes have been identified for many natural materials, from amyloid fiber formation [22, 38] to $CaCO_3$ biomineralization [39, 40], where the initial formation

of specific “seed” nanostructures is essential for the progression of the second growth/transition step and the ultimate formation of the final material. Intriguingly, synthetically reproducing the quality of fibers that are naturally spun by spiders could be intrinsically dependent on our ability to start the spinning process with the same types of precursors (seeds), not only from a biological and chemical perspective, but also in their nanoscale morphologies. Indeed, Rising, Johannson and coworkers recently developed methods to spin artificial spider silk fibers using recombinant spidroins, starting by first forming small (*ca.* 20-50 nm) spherical micelles (non-hierarchical).[7, 8] However, the artificial fibers exhibited significantly inferior mechanical properties compared to natural spider silks although they are among the best synthetic fibers reported to date.[7, 8] In this work we have shown that larger, hierarchical micellar nano-assemblies are the initial protein structures in the native dope; the seeds from which natural silk fibers are spun. Achieving similar synthetically produced hierarchical nano-assemblies (several hundred nanometers in diameter) will be critical for bulk-producing robust synthetic silk fibers from recombinant spidroin proteins.

Supplemental Figures

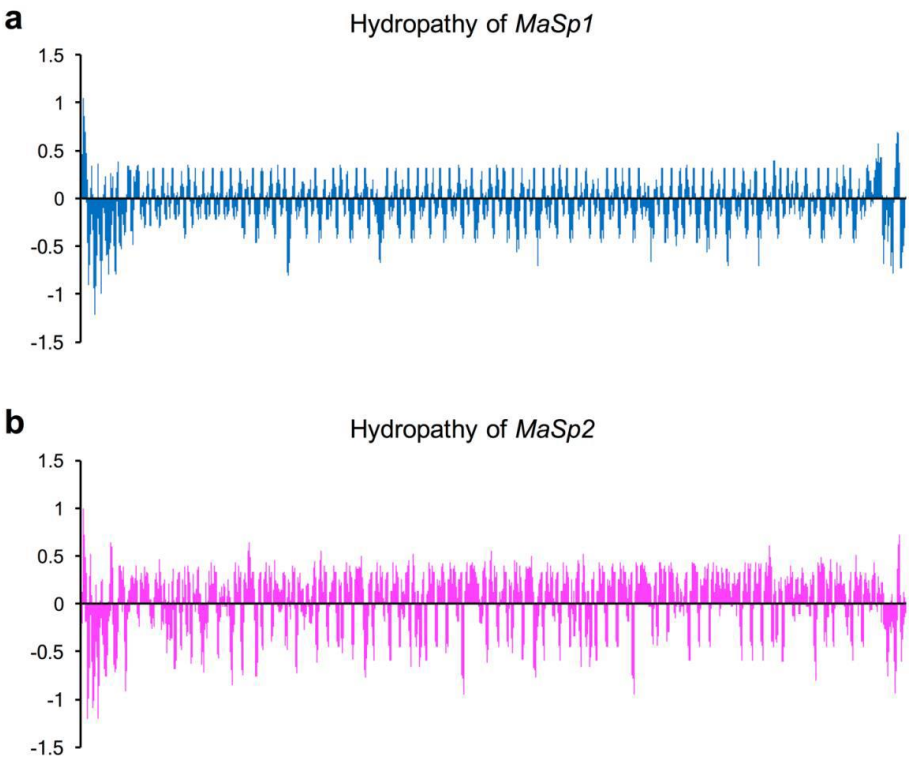


Figure 20 Hydropathy (Roseman)[7] plots for MaSp1 and MaSp2 spidroin proteins of L. hesperus MA silk. (A) Hydropathy plot for MaSp1. (B) Hydropathy plot for MaSp2. Hydropathy plots are generated from the protein sequences in ref. (1).

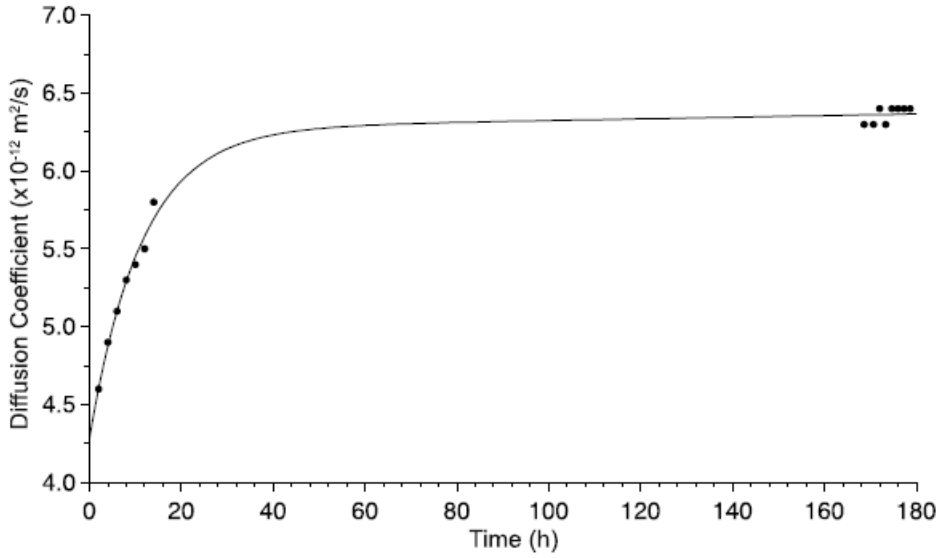


Figure 21 Diffusion of *L. hesperus* MA silk protein dope as a function of incubation time in 4M urea measured by PFG-STE NMR.A strong ∆dependence is observed for MA silk protein when the urea incubation time is < 15 hrs(**Figure. 1B**). Following incubation in urea for 24-48 hrs only free (not restricted) diffusion is observed with a constant diffusion coefficient as a function of ∆(**Figure 1C**)

A

	δ ^{1}H (F2)		δ ^{15}N (F1)	
Residue	Native	4M Urea	Native	4M Urea
A**A1**	8.30	8.29	123.7	123.9
AA2	8.15	8.18	123.5	123.8
A**G1**	8.48	8.46	108.9	108.9
AG1	8.29	8.31	108.4	108.6
G**A1**	8.31	8.30	124.6	124.6
GA1	8.15	8.14	125.0	124.8
G**G1**	8.44	8.44	111.6	111.6
GG2	8.32	8.31	109.6	109.5
GG3	8.23	8.25	109.3	109.4
G**Q1**	8.40	8.40	120.6	120.6
GQ2	8.30	8.30	120.6	120.6
Q-NH_2	7.57	7.52	112.7	112.9
Q**G**	8.55	8.54	110.7	110.7

B

	^{1}H FWHM (Hz)	
Residue	Native	4M Urea
A**A2**	43	30
A**G1**	33	27
G**A2**	40	24
G**G1**	33	15
G**Q1**	25	17
G**Q2**	42	33
Q-NH_2	26	13
Q**G**	45	35

Figure 22 Amide chemical shift and linewidth data extracted from $_1$H-$_{15}$N HSQC NMR spectra of the native *L. hesperus* MA silk protein and the protein diluted in urea. (*A*) ^{1}H and ^{15}N chemical shifts display negligible shifts when the silk protein is diluted to 5 wt% in 4 M urea. This indicates the urea denaturant does not cause any changes to the secondary structure of the silk proteins that maintain a random coil conformation. (*B*) The $_1$H full-width-at-half maximum (FWHM) decreases following dilution in the 4 M urea indicating an increase in local molecular dynamics. As expected, dilution in 4M urea does not change the random coil structure of the silk protein in the gland with an increase in dynamics observed as evidence by narrower linewidths.

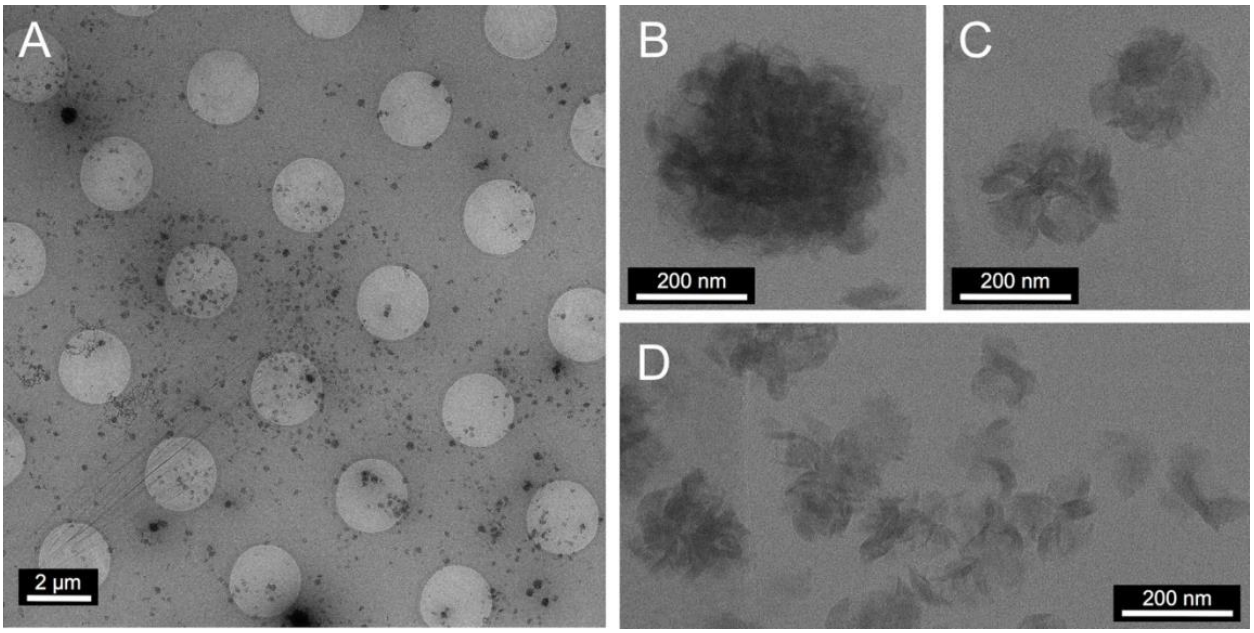


Figure 23 Native L. hesperus MA silk protein imaged by Cryo-TEM (4 hr incubation in 4M urea) from a different spider than **Figure 2**. (A) Low-magnification image showing abundance of generally spherical (200-500 nm) silk protein assemblies. (B-D), High-magnification images of the micellar protein assemblies, representative of the highest populations; predominantly spherical hierarchical assemblies composed of disordered subdomains, and some smaller hierarchical assemblies that have less-spherical morphologies also composed of similar subdomains. Note the similarity between images in **Figure 2** and the images shown here. Samples imaged in **Figure 2** and **Figure 23** were prepared by the same procedure from two different spiders to confirm the reproducibility of the results (hierarchical micelles).

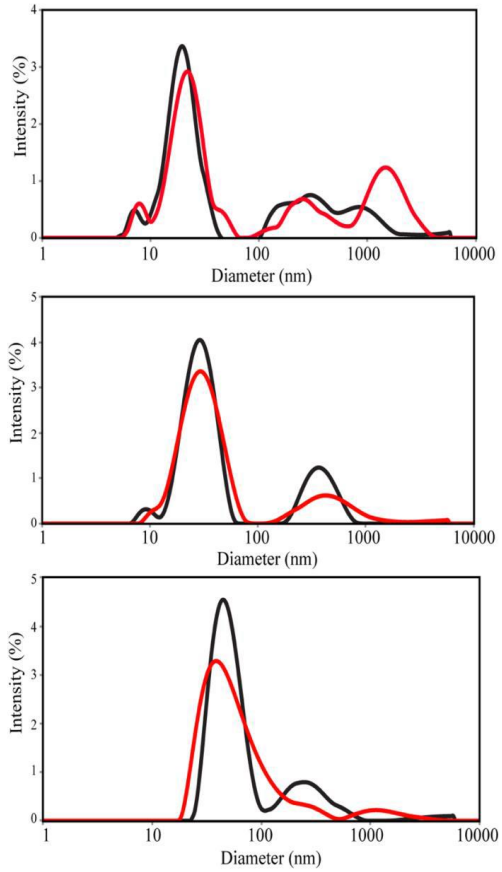


Figure 24 Dynamic light scattering (DLS) data of native *L. hesperus* MA silk protein diluted in 4 M urea to various concentrations. (*A*) 0.7 wt%, (*B*) 0.5 wt%, and (*C*) 0.06 wt%. DLS data was collected following 2 hr (black) and 24 hr (red) incubation in 4 M urea. Multiple sizes were detected at all concentrations. A small, predominate silk protein component is observed at 24, 26 and 47 nm for the three concentrations, respectively. At all concentrations, larger silk protein assemblies in the 100's of nm range were observed. In some cases (*A* and *C*) there was evidence for larger silk protein structures >1 μm following a 24 hr incubation period. A monomer of MaSp1 silk protein is expected to have a hydrodynamic diameter of 44 nm based on the molecular weight and number of residues.

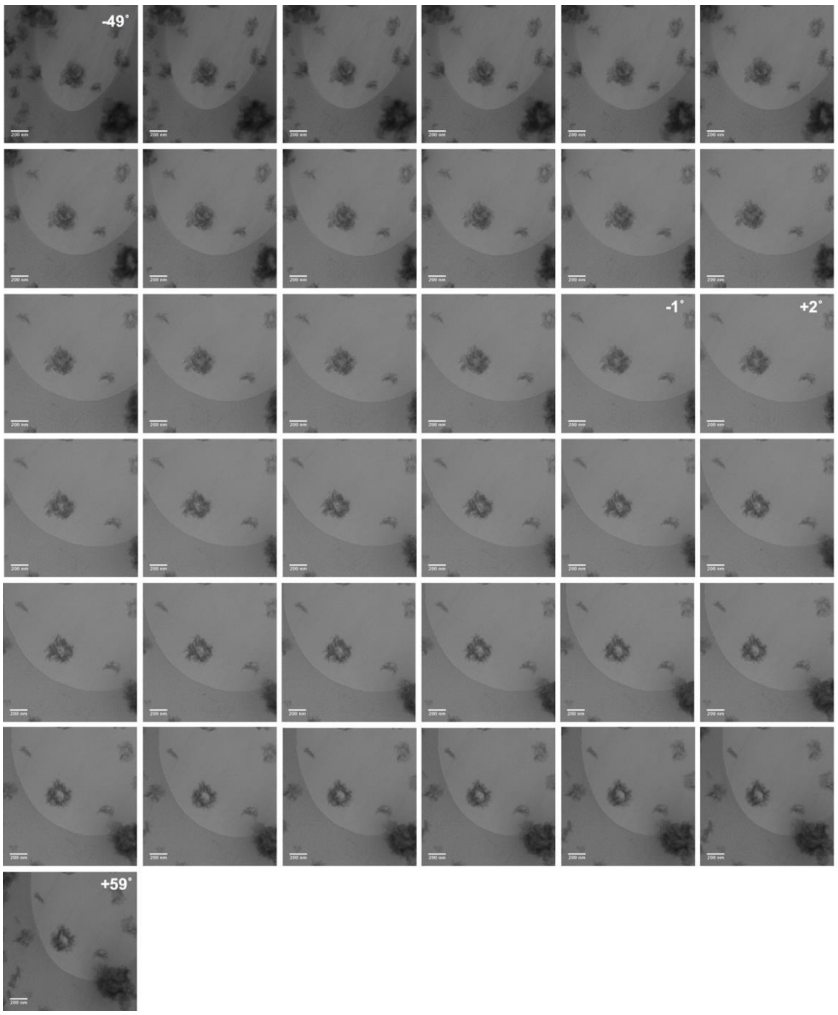


Figure 25 Complete image stack from the cryo-TEM tomography tilt-series of the native *L. hesperus* MA silk protein sample in **Figure 2*C-H*** (4 hr incubation in 4M urea). 37 images were acquired (dose: 2.5 e-/Å$_2$ per image, total cumulative dose: ~92.5 e-/Å$_2$) at 3˚ tilt intervals. This image stack was aligned and used to make the 3D reconstruction and renderings in **Figure 2*F-H*.**

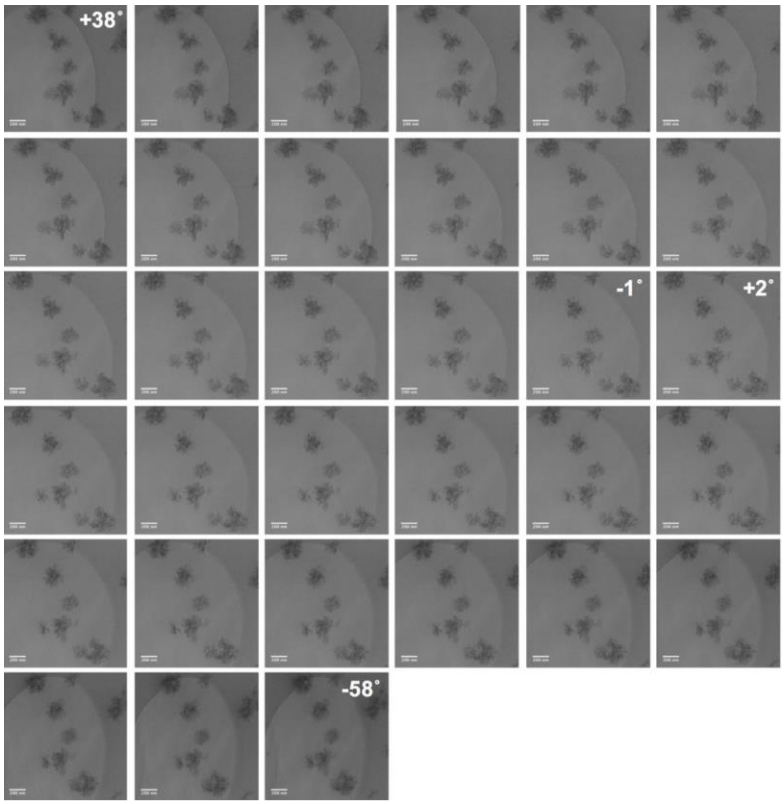


Figure 26 Complete image stack from the cryo-TEM tomography tilt-series of the vigorously micropipetted *L. hesperus* MA silk protein sample in **Figure 3*C-H*** (4 hr incubation in 4M urea). 33 images were acquired (dose: 2.5 $e^-/Å^2$ per image, total cumulative dose: ~82.5 $e^-/Å^2$) at 3˚ tilt intervals. This image stack was aligned and used to make the 3D reconstruction and renderings in **Figure 3*F-H***.

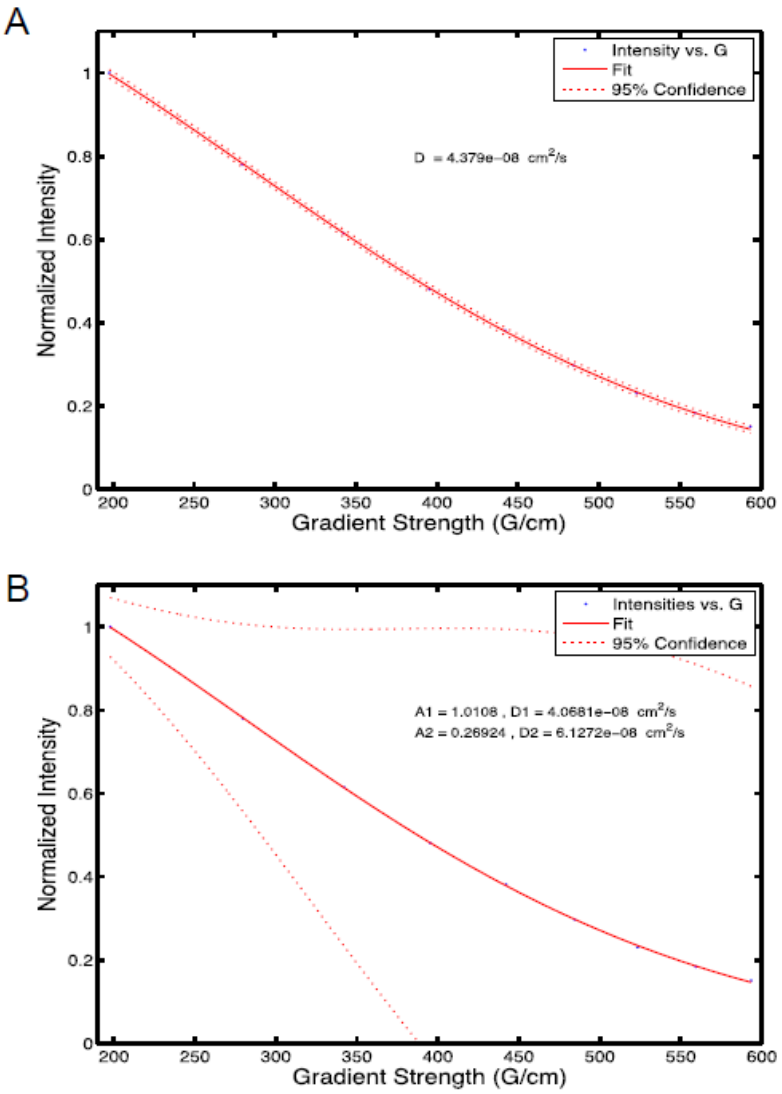


Figure 27 Native *L. hesperus* MA silk protein Einstein self-diffusion coefficient, *D*, determined by PFG-STE NMR. Fitting the raw PFG-STE NMR diffusion data with, a, one component or, b, two components. Sample concentration was 0.2 wt%, data collected 1 hour following sample preparation in 4 M urea. (*A*) Diffusion coefficients were extracted from the decays by fitting the data to the Stejskal-Tanner equation using a single component. (*B*) The same data fit to two diffusing components. Data points are in blue, the resulting fit in solid red line, and 95% confidence intervals with dotted red line. While there may be two diffusing species (*e.g.* a major monomer species and a minor dimer species), we cannot reliably extract out two components.

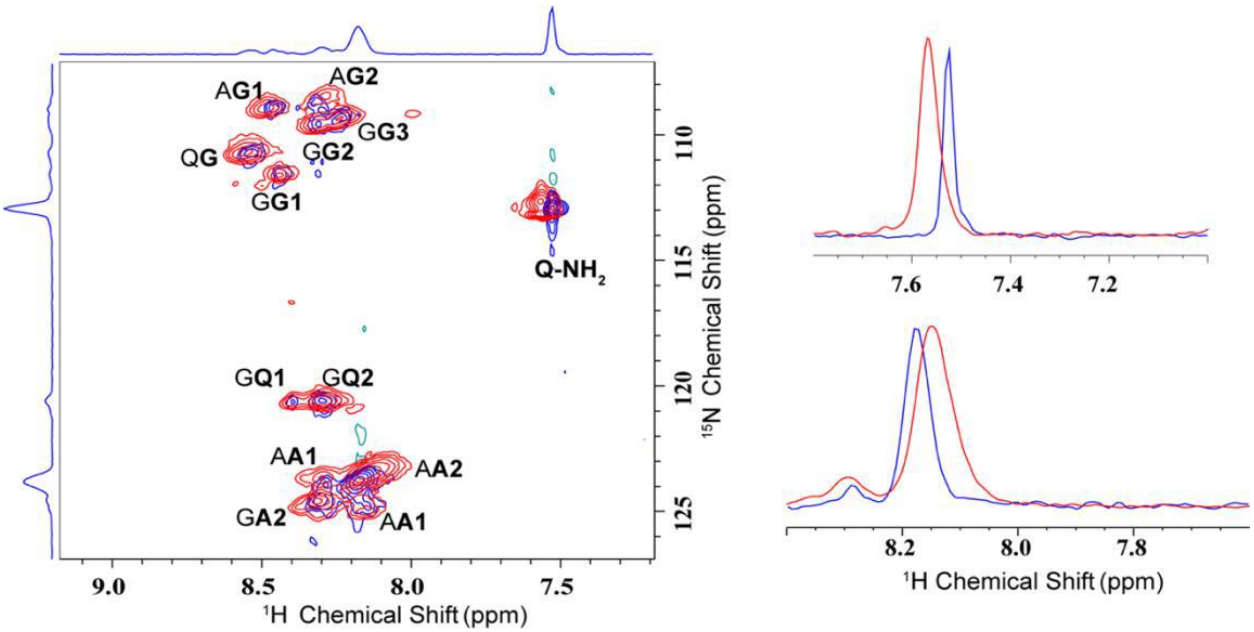


Figure 28 $_{1}$H-$_{15}$N HSQC NMR of native *L. hesperus* MA silk protein. (*A*) HSQC collected on native (c ~35 wt%, pH ~ 7), intact glands (red) and following dilution to ~5 wt% in 4 M urea (pH ~ 9) for 2 hrs (blue). (*B*) ^{1}H slices extracted for the Q-NH$_2$ (top) and AA2 (bottom) resonances. Negligible chemical shift perturbations were observed following dilution in 4 M urea indicating the denaturant did not cause any major structural changes to the repetitive core of the spider silk proteins that maintain a predominantly random coil conformation (see **Figure 22*A***). The NMR linewidths exhibited a marked sharpening as expected following dilution in the denaturant due to an increase in local molecular dynamics (see **Figure 22*B***).

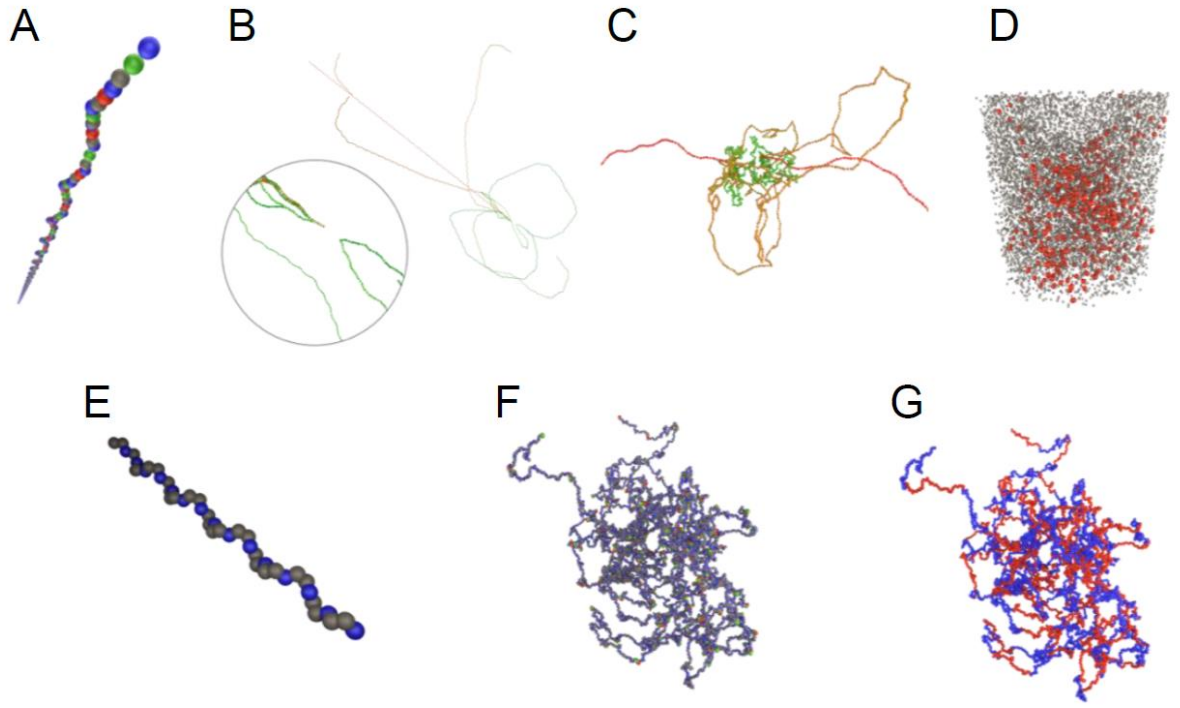


Figure 29 Details on the underlying computation used to construct the graphics (maquette) in **Figure 19**. (*A*) Successive N-X-C triplets define spherocylinders (linear sphere-swept volumes) that are concatenated via well-defined dihedrals, bond angles and bond-lengths at each C-N junction between the coarse grains. C termini are grey, N termini are blue and X (red or green) refers to the hydrophobic or hydrophilic identity of the segments. (*B*) The N terminus is placed at the first anchor point and the free end is rotated about the axis formed by a randomly picked C-N junction. If the resulting structure reduces the distance between the free end and a second anchor point and no spherocylinders intersect the structure is kept. If the distance increases the structure is kept with a probability that is proportional to the Boltzman factor associated with the change in energy of an imaginary spring between the end-point and the second anchor. Green structures have the free end closer to the anchor point. (*C*) Crankshaft moves are performed in which interior segments of chain are rotated as rigid bodies about an axis defined between pairs of C atoms or pairs of N atoms. (*D*) A biased diffusion involving crankshaft folding moves that do not lead to self-intersection eventually finds a structure which is wholly contained inside any given envelope, depicted here by the locus of the transparent spheres. (*E*) Once a pathway of non-intersecting spherocylinders is found each segment is populated with a peptide backbone consisting of the integer part of the mean number of residues for that segment type from Table S1. The short sequence of N-C-C triplets is connected between the end points of the spherocylinder and folded within a cylindrical envelope with 2 Å diameter using a method similar to that found in panels *B*-*D*. The CA atoms in adjacent residues define the dihedral twist axis for connecting the short peptide to the coarse grain end point and the crankshaft moves use randomly picked CA atoms to define the axis of rotation. (*F*) The fully populated structure in which the ends of each coarse grain segment are highlighted with red and green points. (*G*) Each segment forms a hydrophobic or hydrophilic segment of the randomly coiled peptide chain in which no self-intersection of the polymer backbone occurs.

Acknowledgements

Gregory Holland, Jeffrey Yarger and Nathan Gianneschi conceived and led the project. Lucas Parent and David Onofrei contributed equally to the work. David Onofrei and Dian Xu performed the PFG-STE NMR experiments with Brian Cherry and Samrat Amin and prepared **Figure 16**. Brian Cherry collected and processed the MRI data. Dillan Stengal collected and analyzed the DLS data. Lucas Parent performed the cryo-TEM characterization, prepared **Figure 17** and **Figure 18**, and wrote the manuscript. John Roelhling performed the tomographic reconstructions and 3D renderings/visualizations from the cryo-TEM data. Christopher Foreman created the graphics in **Figure 19** and Lucas Parent prepared **Figure 19**. All co-authors provided critical feedback on data processing and analysis and helped shape the manuscript. We thank Prof. Baker for access to the instrument, and we thank Dr. James Bower for his assistance with the cryo-TEM experiments. The DLS work was performed with the assistance of David Pullman at SDSU.

References

1. Lazaris A, et al. (2002) Spider Silk Fibers Spun from Soluble Recombinant Silk Produced in Mammalian Cells. *Science* 295:472–476.

2. Yarger JL, Cherry BR, van der Vaart A (2018) Uncovering the structure–function relationship in spider silk. *Nat Rev Mater* 3:18008–11.

3. Koski KJ, Akhenblit P, McKiernan K, Yarger JL (2013) Non-invasive determination of the complete elastic moduli of spider silks. *Nat Mater* 12:262–267.

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

5. Hardy JG, Romer LM, Scheibel TR (2008) Polymeric materials based on silk proteins. *Polymer (Guildf)* 49:4309–4327.

6. Heidebrecht A, et al. (2015) Biomimetic Fibers Made of Recombinant Spidroins with the Same Toughness as Natural Spider Silk. *Adv Mater* 27:2189–2194.

7. Andersson M, et al. (2017) Biomimetic spinning of artificial spider silk from a chimeric minispidroin. *Nat Chmical Biol* 13:262–264.

8. Kronqvist N, et al. (2017) Efficient protein production inspired by how spiders make silk. *Nat Commun* 8:15504–15.

9. Vollrath F, Knight DP (2001) Liquid crystalline spinning of spider silk. *Nature* 410:541–548.

10. Swanson BO, Blackledge TA, Beltran J, Hayashi CY (2006) Variation in the material properties of spider dragline silk across species. *Appl Phys A* 82:213–218.

11. Jin H-J, Kaplan DL (2003) Mechanism of silk processing in insects and spiders. *Nature* 424:1057–1061.

12. Hagn F, et al. (2010) A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature* 465:239–242.

13. Willcox PJ, Gido SP (1996) Evidence of a Cholesteric Liquid Crystalline Phase in Natural Silk Spinning Processes. *Macromolecules* 29:5106–5110.

14. Kerkam K, Viney C, Kaplan D, Lombardi S (1991) Liquid Crystallinity of Natural Silk Secretions. *Nature* 349:596–598.

15. Jenkins JE, et al. (2013) Characterizing the Secondary Protein Structure of Black Widow Dragline Silk Using Solid-State NMR and X-ray Diffraction. *Biomacromolecules* 14:3472–3483.

16. Ayoub NA, Garb JE, Tinghitella RM, Collin MA, Hayashi CY (2007) Blueprint for a High-Performance Biomaterial: Full-Length Spider Dragline Silk Genes. *PLoS One* 6:e514-13.

17. van Beek JD, Hess S, Vollrath F, Meier BH (2002) The molecular structure of spider dragline silk: Folding and orientation of the protein backbone. *Proc Natl Acad Sci U S A* 99:10266–10271.

18. Römer L, Scheibel T (2008) The elaborate structure of spider silk. *Prion* 2:154–161.

19. Askarieh G, et al. (2010) Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature* 465:236–238.

20. Eisoldt L, Hardy JG, Heim M, Scheibel TR (2010) The role of salt and shear on the storage and assembly of spider silk proteins. *J Struct Biol* 170:413–419.

21. Rising A, Widhe M, Johansson J, Hedhammar M (2011) Spider silk proteins: recent advances in recombinant production , structure–function relationships and biomedical applications. *Cellular Mol Life Sci* 68:169–184.

22. Saric A, Chebaro YC, Knowles TPJ, Frenkel D (2014) Crucial role of nonspecific interactions in amyloid nucleation. *Proc Natl Acad Sci* 111:17869–17874.

23. Kenney JM, Knight D, Wise MJ, Vollrath F (2002) Amyloidogenic nature of spider silk. *Eur J Biochem* 269:4159–4163.

24. Hijirida DH, et al. (1996) 13C NMR of Nephila clavipes Major Ampullate Silk Gland. *Biophys J* 71:3442–3447.

25. Hronska M, van Beek JD, Williamson PTF, Vollrath F, Meier BH (2004) NMR Characterization of Native Liquid Spider Dragline Silk from Nephilia edulis. *Biomacromolecules* 5:834–839.

26. Jenkins JE, Holland GP, Yarger JL (2012) High resolution magic angle spinning NMR investigation of silk protein structure within major ampullate glands of orb weaving spiders. *Soft Matter* 8:1947–1954.

27. Xu D, Yarger JL, Holland GP (2014) Exploring the backbone dynamics of native spider silk proteins in Black Widow silk glands with solution-state NMR spectroscopy. *Polymer (Guildf)* 55:3879–3885.

28. Andersson M, et al. (2014) Carbonic Anhydrase Generates CO2 and H+ That Drive Spider Silk Formation Via Opposite Effects on the Terminal Domains. *PLOS Biol* 12:e1001921-14.

29. Xu D, Guo C, Holland GP (2015) Probing the Impact of Acidification on Spider Silk Assembly Kinetics. *Biomacromolecules* 16:2072–2079.

30. Alexandridis P, Lindman B (2000) *Amphiphilic Block Copolymers: Self-Assembly and Applications* eds Alexandridis P, Lindman B (Elsevier Science B.V., Amsterdam). First Edit.

31. Alexandridis P (1996) Amphiphilic copolymers and their applications. *Curr Opin Colloid Interface Sci* 1:490–501.

32. Israelachvili JN, Mitchell DJ, Ninham BW (1975) Theory of Self-Assembly of Hydrocarbon Amphiphiles into Micelles and Bilayers. *J Chem Soc Faraday Trans 2* 72:1525–1568.

33. Lin T-Y, et al. (2017) Liquid Crystalline Granules Align in a Hierarchical Structure To Produce Spider Dragline Microfibrils. *Biomacromolecules* 18:1350–1355.

34. van Dusschoten D, de Jager PA, Van As H (1995) Extracting Diffusion Constants from Echo-Time-Dependent PFG NMR Data Using Relaxation-Time Inforamtion. *J Magn Reson Ser A* 116:22–28.

35. Sorland GH (2014) *Dynamic Pulsed-Field-Gradient NMR* ed Sorland GH (Springer-Verlag Berlin Heidelberg). XIII doi:10.1007/978-3-662-44500-6.

36. Wilkins DK, et al. (1999) Hydrodynamic Radii of Native and Denatured Proteins Measured by Pulse Field Gradient NMR Techniques. *Biochemistry* 38:16424–16431.

37. Zhuge X, Palenstijn WJ, Batenburg KJ (2016) TVR-DART: A More Robust Algorithm for Discrete Tomography From Limited Projection Data With Automated Gray Value Estimation. *IEEE Trans Image Process* 25:455–468.

38. Castello F, Casares S, Ruedas-Rama MJ, Orte A (2015) The First Step of Amyloidogenic Aggregation. *J Phys Chem B* 119:8260–8267.

39. Nielsen MH, Aloni S, De Yoreo JJ (2014) In situ TEM imaging of CaCO3 nucleation reveals coexistence of direct and indirect pathways. *Science* 345:1158–1162.

40. De Yoreo JJ, et al. (2015) Crystallization by particle attachment in synthetic, biogenic, and geologic environments. *Science* 349:aaa6760.

41. Pantelic RS, Meyer JC, Kaiser U, Baumeister W, Plitzko JM (2010) Graphene oxide: A substrate for optimizing preparations of frozen-hydrated samples. *J Struct Biol* 170(1):152–156.

42. Patterson JP, et al. (2012) A simple approach to characterizing block copolymer assemblies: graphene oxide supports for high contrast multi-technique imaging. *Soft Matter* 8(12):3322–3328.

43. Kremer JR, Mastronarde DN, McIntosh JR (1996) Computer Visualization of Three-Dimensional Image Data Using IMOD. *J Struct Biol* 116:71–76.

44. Zhuge X, et al. (2017) Automated discrete electron tomography–Towards routine high-fidelity reconstruction of nanomaterials. *Ultramicroscopy* 175:87–96.

45. Tká I, Gruetter R (2005) Methodology of 1H NMR Spectroscopy of the Human Brain at Very High Magnetic Fields. *Appl Magn Reson* 29:139–157.

46. Stejskal EO, Tanner JE (1965) Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *J Chem Phys* 42:288–292.

47. Wu D, Chen A, Johnson CS (1995) An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses. *J Magn Reson Ser A* 115:260–264.

48. Kawahara K, Tanford C (1966) Viscosity and Density of Aqueous Solutions of Urea and Guanidine Hydrochloride. *J Biol Chem* 241:3228–3232.

Chapter 3 Investigating the Atomic and Mesoscale Interactions that Facilitate Native Spider Silk Protein Pre-assembly

Abstract

Black widow spider dragline silk is one of nature's strongest materials, far exceeding man-made materials by weight such as high tensile steel. In order to replicate the mechanical properties of these incredible biological fibers, we must understand, first, how the pre-spun silk protein-rich solution known as dope is synthesized, stored and organized within the gland and, second, the spinning process by which this dope is rapidly converted into high performance fibers. Major ampullate spidroin 1 (MaSp1) and MaSp2 are the two major protein components of dragline silk, or Ma silk fibers. These proteins are stored as primarily unstructured, random coils within intact Ma glands, although recent studies have shown that silk proteins are stored in part, as pre-assembled, 200-300 nm complex micelle-like protein superstructures. In this work we use DLS and three dimensional (3D) protein solution NMR to probe the size of the protein assemblies and identify which domains of the MaSp proteins facilitate pre-assembly. This was accomplished by disrupting these supramolecular protein structures with a moderate denaturant (4M Urea) and using a combination of light scattering and 3D solution NMR. DLS illustrates that these 200-300 nm silk protein assemblies persist even at dilute concentration and decrease in size as function of time in the urea denaturant. 3D solution NMR on isotopically ($^{13}C/^{15}N$) enriched native spider silk proteins provides evidence that the Gly-Ala and Ala-Ala-Gly regions flanking the poly(Ala) runs that comprise the β-sheet core in fibers are perturbed by the disassembly process in urea suggesting that these regions may be important for silk protein

superstructure stabilization. The poly(Ala) runs and Gly-Gly-X repeating motifs show essentially no chemical shift perturbations during the disassembly process. Lastly, NMR relaxation measurements at different silk protein concentrations in urea illustrate the presence of two distinct components for T_1 which are initially ascribed to silk protein monomer and superstructures that are observed in solution by DLS.

Introduction

With a single spider capable of producing up to seven different silk varieties, spider silks have evolved to be one of nature's most versatile biopolymers, exhibiting a broad range of extensibilities, strength and toughness that surpasses both steel and Kevlar®.[2-4] Each of these unique fibers is synthesized by a specialized gland within the spider's abdomen before being used in combination to create function-specific silks for applications such as building webs, wrapping prey and constructing egg cases. Resourceful and robust, this adaptable biomaterial has been sought after for a host of applications, ranging from defense, to biomedical and textile[5-8]; however, these industries would require silk to be produced at a much larger scale than spiders are capable of delivering. As such, many studies have focused on producing a synthetic silk that mimics the properties of native fibers.[7, 9-11] Accomplishing this feat requires a deeper knowledge of the silk fiber hierarchical structure and the spinning process that converts the silk protein dope to fibers.

MA silk, the most widely studied type of silk, is used in constructing the radii of the orb weavers web and most of cobwebs of cob-weavers.[4] The MA silk is composed primarily of two silk fibroin proteins, known as spidroins, as well as a number of minor components.[12] These two prominent spidroins have been coined major ampullate

spidroin 1 and 2, (MaSp1 and MaSp2), named for the gland from which they are synthesized and stored prior to fiber formation. Though these have long been considered the main components that make up the MA fibers, recent genomic analysis has revealed that MA silk may have at least one other protein component (MaSp3).[13-15] In fact, a fourth MaSp protein has been identified in *Caerostris darwini* (MaSp4), though it remains unclear whether a paralog of this protein exists for Ma silk from *L. hesperus*.[16] Expression of each spidroin is generally localized to a single gland in the spider, although some spidroins may be expressed from multiple glands.[15, 17-19] These very large (~300kDa) proteins have in common short, highly ordered termini[20, 21] flanking very long repetitive core regions. MaSp1 and Masp2 core regions are composed of blocks of similar repeat units varying mainly in their number of short motifs. Typical repeating motifs include poly(Ala), Gly-Gly-X (where X is either Gln, Ala, or Tyr), and poly(Gly-X) (where X is either Arg or Gln). In the case of MaSp2, Gly-Pro-Gly[22] is also a common motif not found in MaSp1 and has been shown to form an elastin-like type II β-turn that is believed to contribute to spider silk's extensibility.[23] Using a range of protein solution NMR structural and dynamic methods, the repetitive regions of these spidroins have been shown to be primarily intrinsically disordered random coils as stored in the MA gland.[1, 24]

Despite knowing the full protein sequence of several spidroins, it is still unclear how they are solubilized and stored in the MA gland at very high concentrations (25-50 wt.%) before being converted into solid fibers rich in secondary structure. The native solution spidroin structure is particularly difficult to probe with solution NMR due to its highly repetitive sequence, dynamic, random coil nature and propensity for

aggregation.[25] To circumvent these challenges, many studies have focused on recombinant proteins of various smaller lengths which are readily studied as they are often soluble in water and considerably lower molecular weight.[11, 26] Previously, we avoided the solubility issues altogether by analyzing intact glands with solution NMR, and found that the native spidroins as stored in MA glands are maintained in a predominantly unfolded, random coil state.[1] More recently, we discovered with diffusion NMR that large nanoscale (100's of nm) pre-assembled, micelle-like complexes are present within in the MA gland. Using cryogenic transmission electron microscopy (cryo-TEM) on glands solubilized in 4 M urea we discovered that these structures continue to exist even under dilute conditions.[27] These prefibrillar, pre-assembled protein superstructures are crucial for proper assembly of silk fibers with native mechanical properties and are becoming increasingly appreciated as important for synthetic spider silk assembly.[28-30] They are also believed to be sensitive to the gland conditions including pH, salt, and protein concentration, each of which can inherently be investigated by the described solution NMR approaches and are continuing to be investigated.[31-42] Recombinant fibers which do not fully replicate the native tensile mechanical properties have also been observed to form similar structures, albeit significantly smaller in size (20-30 nm), but there is not yet a clear understanding of the atomic and molecular level details leading to their assembly, the impact of physiochemical conditions in the duct, nor their overall significance in forming better fibers.[43] Without these key pieces of information of how the different variables impact the atomic and molecular level structure of the spidroins, there is little hope of being able to devise 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 further how these structures change when exposed to the physiochemical conditions in the duct prior to 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, nanoscale complexes. Intentional denaturation is often used for purification of recombinantly produced spidroins[44, 45]; to date, however, only a few studies have looked at the structure and dynamics of spidroins under denaturing conditions.[24, 46] By intentionally disrupting the spidroin superstructures in denaturant, we are able to determine the subtle but crucial interactions that facilitate their stabilization. In the present work, we characterize the spidroin protein pre-assemblies by disrupting them in a modest 4M urea denaturant and using DLS to investigate mesoscale assembly size and solution-state NMR to interrogate the atomic level intermolecular interactions.

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 of MA silk during these feedings to drive metabolic turnover in the silk proteins and direct incorporation of isotope labeled amino acids into the silk proteins. Previous work has shown that this labeling scheme gives significant labeling for Ala, Gly and Gln residues.[47] Ala and Gly make up over 60% of the protein sequence of the two major spidroins of black widow MA silk.[20] 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 4M 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. Several samples were prepared at difference concentrations. For NMR relaxation analysis we prepared samples in two regimes: dilute (~1 wt%) and concentrated (~ 5-7% wt.) using 6 and 18 MA glands respectively. Solutions prepared for DLS were at lower concentrations as discussed below.

Dynamic Light Scattering Samples were freshly prepared as described above and then sent to U. Mass. for DLS analysis. Samples arrived within 24 hrs and were run immediately. DLS measures the intensity-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,

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

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,

$$g_1(q,t) = A_0 + \sum A_i \exp(-\Gamma_i t) \qquad (2)$$

where A_0 is the baseline, and A_i is the intensity weighting fraction of ith mode, each having a characteristic decay rate Γ_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 39**.

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 and 23° C. $^{1}H/^{15}N$-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 experiments. $^{1}H/^{13}C$-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 the dynamics (T_1, T_2, and NOE) experiments. Sample concentrations were kept similar for all samples through the addition of 4M 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.[48] 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. We have observed very slight disturbances of the gel-like dope of a sample to induce aggregation and render samples unusable. For this reason NMR samples have only been used for 48 hours. Fitting of relaxation parameters was done in the Dynamics package of TopSpin using one and two-component exponential decay functions.

Results

Dynamic light scattering measures the scattering of light by particles in solution and can be used to determine the size of individual proteins monomers and larger structures such as protein assemblies or superstructures. Previously we have shown by

Table 2 Hydrodynamic radius determined from DLS data for native MA silk proteins in 4M urea as a function of dilution. The smaller component around 22nm agrees with NMR and empirical measures of the size of a single monomer of MaSp1. The larger component agrees, at least in general length scale, with the larger pre-assemblies previously observed by both NMR and Cryo-TEM. Both monomers and large protein assemblies are always observed.

Dilution Factor	D_1 (cm2/s)	R_{h1} (nm)	D_2 (cm2/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

routine DLS that dilute solutions of native silk dope dissolved in denaturant contain both a spidroin monomer (~19 nm) and larger protein assemblies on the order of 200-300 nm.[27] 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 30**(left)). 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 and protein concentrations well below what can be observed by NMR can be explored. We observe a decrease in the size of the larger assemblies as a function of both time and concentration (**Figure 30).** Stability is reached after a few days, which suggests that some fraction of the larger assemblies are long lived.

Spiders fed U-^{13}C-^{15}N-Ala resulted in MA silk proteins enriched with ^{13}C/^{15}N-Ala, ^{13}C/^{15}N-Gly and ^{13}C/^{15}N-Gln similar to previous observations.[49] In the ^{1}H/^{15}N-HSQC, we observe 12 resonances which correspond to three separate regions: Gly-rich, Gln-rich and Ala-rich (**Figure 31A**). These same 12 resonances have been observed in solution

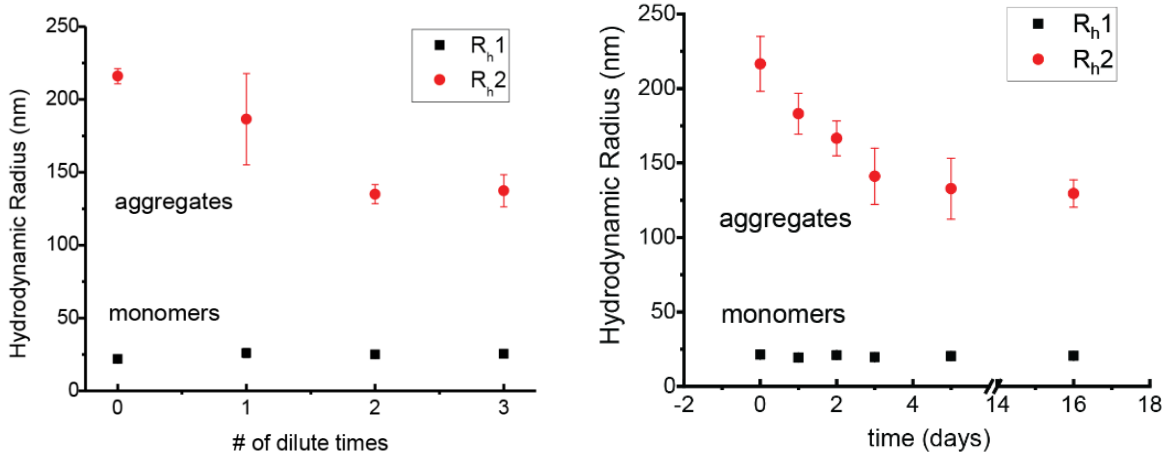


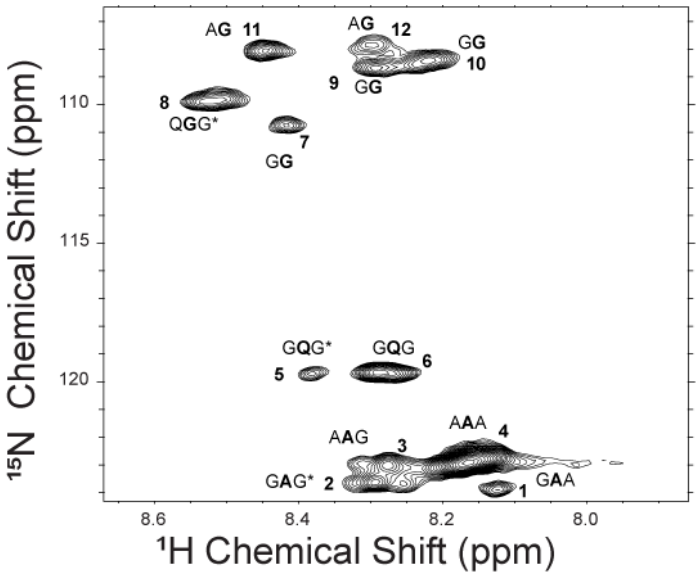
Figure 30 Hydrodynamic radius determined from DLS data for native MA silk proteins in 4M urea as a function of dilution. Both monomers and large protein aggregates are observed. A decrease in the size of the larger assemblies appears to stabilize after a few days. Monomer sizes remain roughly the same 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.

NMR spectra of intact glands and we are able to directly compare the two sets of resonances.[1] Similar to intact glands, very narrow chemical shift dispersion in the ^{1}H dimension is observed and indicates unstructured, highly disordered random coil proteins, reminiscent of intrinsically disordered proteins.[50] The very long chain length, sequence redundancy and poor chemical shift dispersion due to the random coil structure of the proteins make unique amino acid assignment for each of the thousands of amino acid residues extremely challenging; nevertheless, by using 3D solution NMR experiments we can still resolve small 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 32**). This strategy initially permits two-residue assignments for each resonance, which agree very well with the correlated assignment in the intact glands from previous work (Table 2) indicating highly random coil proteins.

In the HNCACB experiment shown in **Figure 32**, 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. 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 32**). Each pair of strips from the CBCAcoNH and HNCACB experiments correlates to an identical ^{15}N slice from the ^{1}H/^{15}N HSQC. For some of the spectra, considerable spectral overlap exists making it challenging to confidently make more than two-residue assignments. However, in a few cases, unique ^{13}Cα/β resonances for a single ^{15}N chemical shift can be found in both spectra, allowing for three residue assignments by comparing both 3D data sets. The chemical shifts of all identified residues correspond very closely to those of random coils, but in a few

cases, we observed significant perturbations between the intact gland and 4M urea spectra, as discussed below.[51] The perturbations from the shifts observed in native silk glands are summarized in **Figure 33**. 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 the previous studies. For example,



...AGAGQGGQGGYGQGGNGQGGAGQGGAAAAAAAAG- MaSp1
GAGQGGYGR...

...GPGGYGQGPSGYGPSGSGGQGYGQGGSGAAAAAA... MaSp2

Figure 31 $^{1}H/^{15}N$-HSQC collected at 600 MHz showing 12 regions identified for MA silk gland proteins dissolved in 4M urea. The triple residue assignments shown are possible with the use of 3D experiments shown in Figure 3. An example of the repetitive core regions of both spidroins is shown below the HSQC spectrum illustrating the degeneracy of the sequence as well the as the 3-residue motifs.

the GQG motif is very common in the sequence of MaSp1 and MaSp2 although it generally localizes to MaSp1**.**

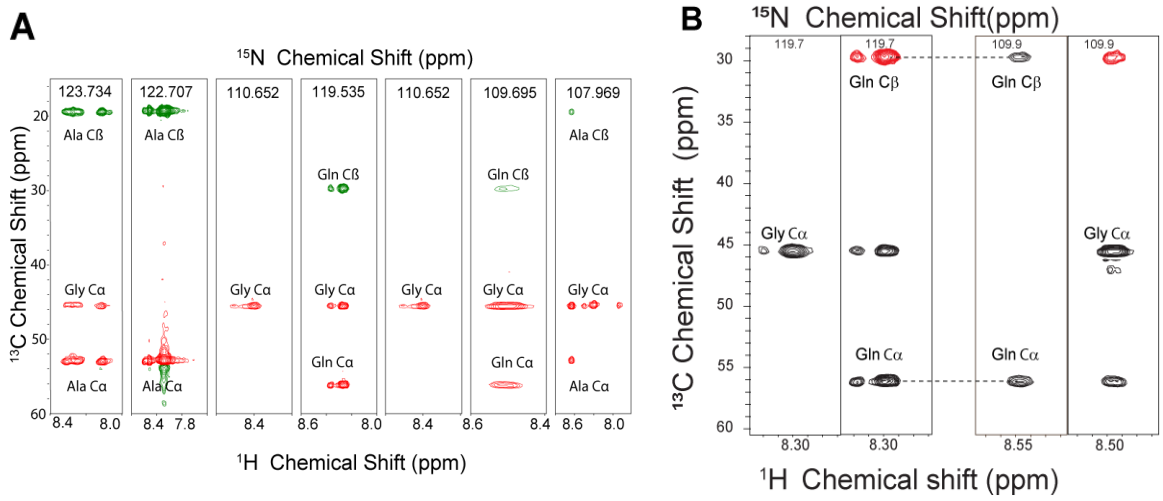


Figure 32 Strip plots from HNCACB experiments showing 2-residue correlations (top). Two residue assignments can be made from this single experiment, but the majority of repeating motifs in the silk sequences are three-residues in length. For this reason we applied a second 3D experiment, CBCACONH, which ties together two amide resonances to a common set of carbon resonances. For some cases, a unique pair could not be assigned due to large amounts of overlap. But in a several cases (example of GQG shown) unique carbon signals could be matched. It is important to note that these are not unique residues, but rather a collection of repeating motifs that have the same triple-residue pattern. In two cases, the published sequence was used to assign sequences which could not be unambiguously assigned to a single resonance type.

We next measured the MA silk spidroin backbone relaxation parameters T_1, T_2, and heteronuclear NOEs for the backbone amide nitrogen nuclei. These values report on global tumbling of proteins and local chain dynamics and serve as an indicator of the level of order along the protein backbone. Not all 12 resonances in the original $^{1}H/^{15}N$ HSQC were sufficiently resolved in the relaxation measurements to be adequately fit to a decay curve but most of them were. The significant increase in the dynamics of the proteins relative to intact glands further suggests that they exhibit a similar disordered random coil state in 4M urea. However, unlike the structural experiments, direct comparison between intact glands, which are highly concentrated (25-50% wt), and

samples dissolved in 4M urea are not necessarily possible. The increased dynamics can be readily correlated to the concentration of the sample, which agrees with diffusion data published by us on this system.[27] This data clearly shows that the bulk relaxation

Table 3 Chemical shift comparison between native[1] and denatured Ma spider silk dope in 4M Urea. Differences refer to difference between the intact values from the literatures and the urea measured values in this paper.

*	Cα	Cß	N H	15N		
Ala	52.67±1.28	19.03±1.27	8.11±0.68	123.52±3.51		
Gln	55.94±1.83	28.67±1.73	8.25±0.75	119.73±385		
Gly	45.34±1.17	-	8.34±0.83	109.94±4.09		
	Cα (Urea)	**Cα (Intact)**	**difference**	**Cß (Urea)**	**Cß (Intact)**	**difference**
G A A G A G	53.06, 52.81	52.93, 5295	0.13, 0.14	19.52, 19.52	18.77, 19.36	0.75, 0.16
A A G, A A A	53.81, 52.81	53.03, 52.99	0.78, 0.18	19.27, 19.27	19.12, 19.11	0.15, 0.16
G Q G	56.34, 56.09	56.09, 56.04	0.25, 0.05	29.61, 29.61	29.51, 29.50	0.10, 0.11
G G	45.50	45.41 - 45.50	0.00			
Q G G	45 .50	45.50	0.00			
A G	45.50, 45.50	45.41, 45.37	0.09, 0.13			
	NH (Urea)	**NH (Intact)**	**difference**	**^{15}N (Urea)**	**^{15}N (Intact)**	
G A	8.11, 8.28	8.10, 8.27	0.01, 0.01	123.9, 123.6	123.99, 123.75	0.1, 0.2
A A G, A A A	8.25, 8.14	8.25, 8.10	0.00, 0.04	123.1, 122.9	122.75, 122.62	- 0.3, - 0.3
G Q G	8.36, 8.27	8.35, 8.26	0.01, 0.01	119.8, 119.8	119.72, 119.68	- 0.1, - 0.1
G G	8.21 - 8.40	8.19 - 8.40	0.00 - 0.02	108.6 - 110.8	108.49 110.68	- 0.1
Q G G	8.52	8.51		109.9	109.85	- 0.1
A G	8.43, 8.28	8.44, 8.24	0.01, 0.04	108.2, 107.9	10 8.08, 107.62	- 0.1, - 0.3
	Hα (Urea)	**Hα (Intact)**	**difference**	**Hß (Urea)**	**Hß (Intact)**	**difference**
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			

profiles of spidroins in 4M urea are very different from their native state in intact glands but, as expected show an increase in dynamics at lower protein concentration.

Because spin-lattice, T_1 relaxation is field dependent, and in order to investigate the role of concentration at the same field strength, we performed this experiment in two concentration regimes at 600 MHz. The T_1 and T_2 parameters are extracted from a decaying exponential fitting function which can contain one or multiple components. The T_2 times extracted from these fits could be represented by a single decaying

exponential (as measured by Monte Carlo simulations to estimate error to the 95% confidence level). Interestingly, the T_1 times were clearly best fit to a two-component decaying curve. A fast and a slow component for the T_1 relaxation times is thus, reported in **Figure 34**.

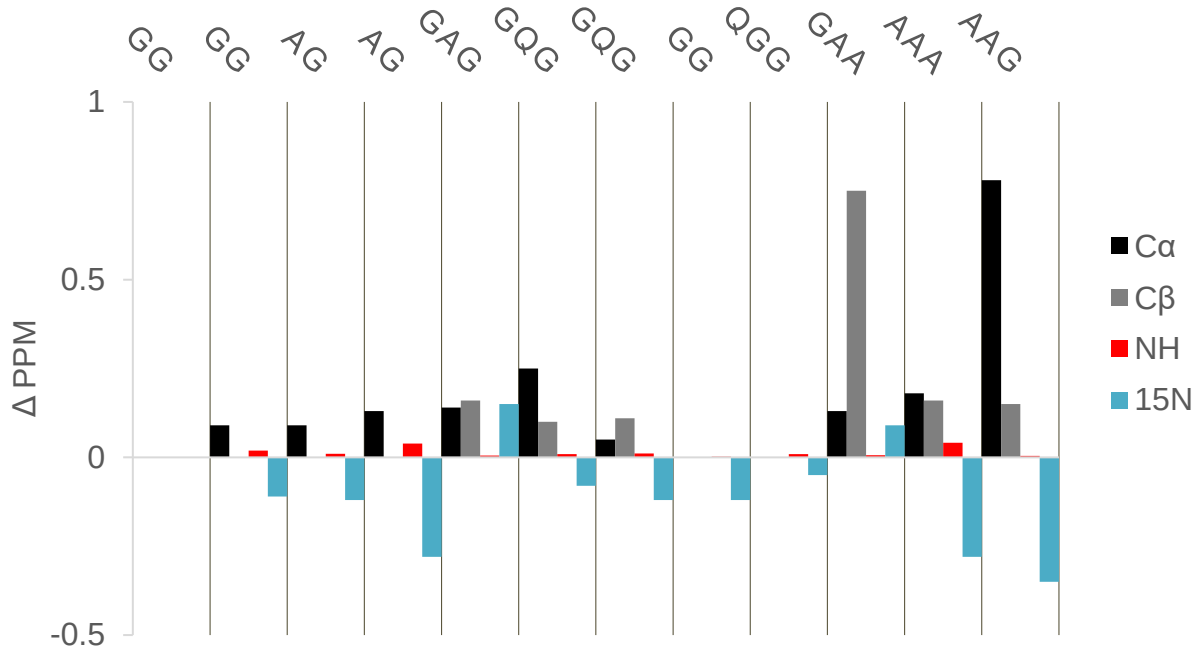


Figure 33 Changes in the chemical shift between the native intact glands and the diluted (~1 wt%) samples are summarized in this chart. The chart is arranged roughly in the order of direction the sequence proceeds so that the glycine-rich regions come before the polyalanine regions.

Heteronuclear NOE values are another valuable indicator of protein dynamics (**Figure 36**). Values for all resolvable resonances are negative, including the AAA peak, 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.

Discussion

Understanding the nature of spider silk proteins in solution, prior to fiber formation, is crucial to elucidating the spinning process, which in turn is a requirement to successfully produce synthetic spider silk fibers with native properties. Solution NMR is a powerful tool for probing atomic and molecular level interactions both in terms of structure and dynamics. Recent work by our group has shown that spidroins assume a hierarchical superstructure within the MA glands that is 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. By diluting the silk protein in 4M urea, we are able to solubilize the native silk dope with all associated components, maintain its stability for at least 48 hours and monitor the deconstruction of these protein superstructures.

Previously, we used diffusion NMR to show that in a dilute solution of 4M urea, the primarily observed resonances arise from freely diffusing monomers following incubations for a few days. In order to observe if a small population of larger protein assemblies is maintained following dilution in 4M urea, we performed DLS on 4M urea samples. DLS can be utilized at considerably lower concentrations compared to NMR. Two peaks roughly matching the predicted sizes of both the monomeric proteins and larger assemblies were observed consistent with our previously published work.

By applying all the typical ^{1}H-detected 2D and 3D protein solution NMR approaches to silk protein residue assignment and have made three residue assignments for the majority of resonances observed in the ^{1}H/^{15}N HSQC spectrum utilizing 3D HNCACB and CBCACONH solution NMR methods.[52, 53] This has allowed us to assign >80% of the repetitive motifs found in the MaSp1 spider silk protein. This is a

tremendous step towards the complete NMR structural and dynamic characterization of the primary spider silk protein and opens the door to continuing investigation under various biochemical conditions (pH, salts, and denaturants).

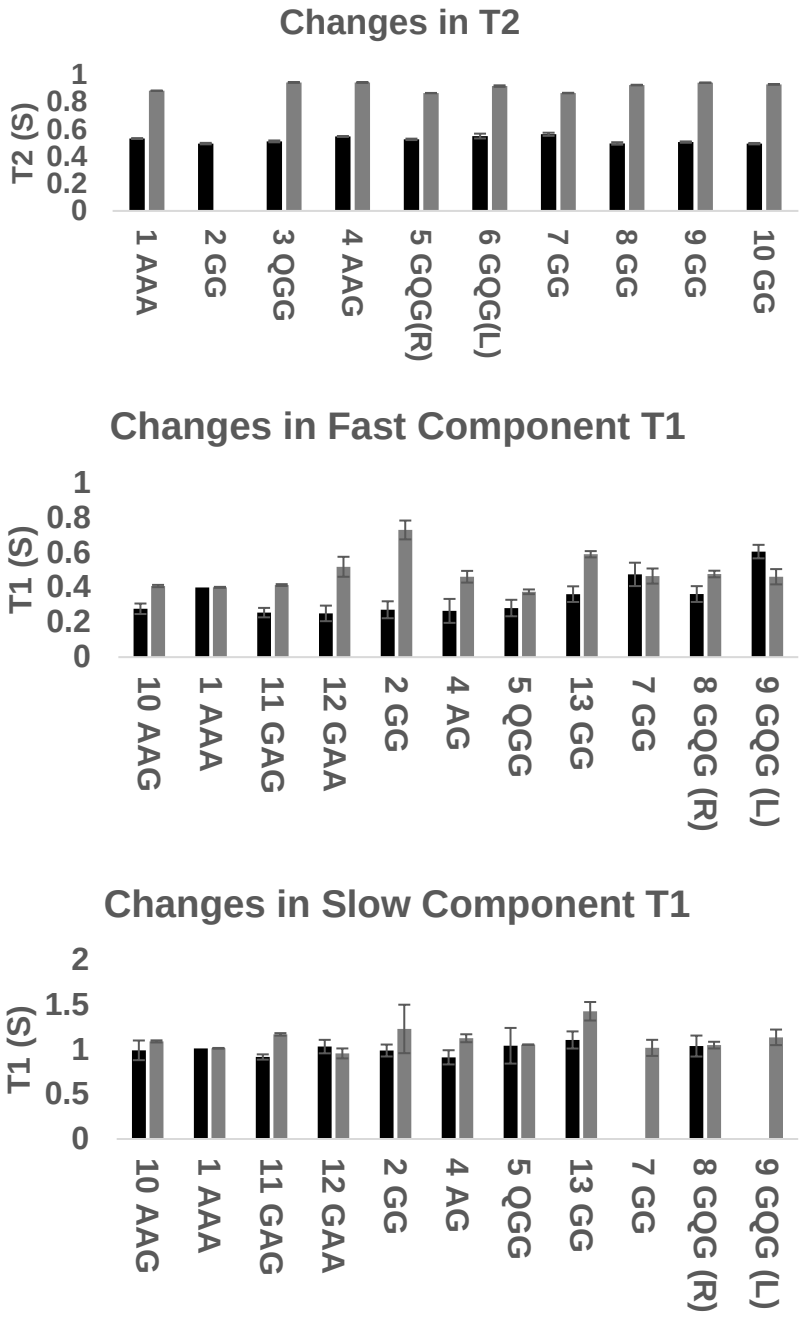


Figure 34 Longitudinal relaxation T_1 (A), transverse relaxation T_2 (B) heteronuclear NOEs (C). T2 values are significantly longer than previously published average values for relaxation on intact glands[1].

With 3D solution NMR, we have identified regions of the protein that correspond to the most common three-residue motifs and report their chemical shifts. This provides greater localization of chemical shift analysis than previous two-residue experiments

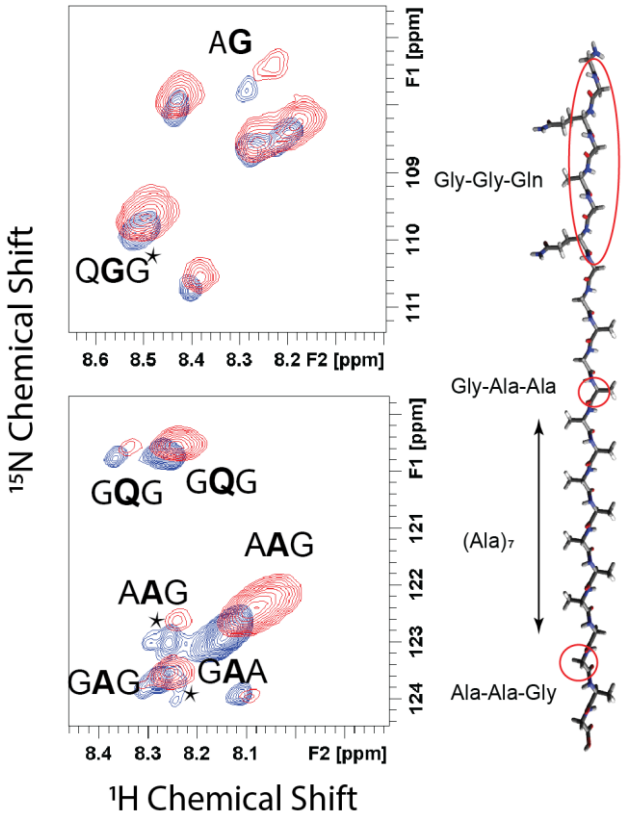


Figure 35 The $^{1}H/^{15}N$ HSQC NMR spectra of $^{13}C/^{15}N$-Ala-labeled MA spider silk protein within intact glands excised from L. hesperus spiders. The Gly and Ala region of the spectrum are expanded in the top and bottom spectra, respectively. The silk dope within intact glands at native concentration (~35 wt%) is shown in red and dope dissolved in 4M urea (~5 wt% protein) in blue. MaSp1 backbone chain is shown to the right. Large chemical shift perturbations are observed for Gly-Ala-Ala and Ala-Ala-Gy domains and the Gly-Gly-Gln and $(Ala)_n$ regions only display negligible or minor changes.

(ref).[1] In fact, the three-residue assignments covers 82% of the full sequence of MaSp1 and 50% of the sequence of MaSp2 (**Figure 38**). However, we have identified regions of the proteins that *do* show significant (~1 ppm) changes and found that these regions flank the polyalanine domains of the core repetitive domains. Spider silk is a semi-

crystalline polymer, with small β-sheet domains interspersed throughout an otherwise mostly amorphous matrix.[54, 55] The β-crystalline domain consists of the poly-Ala regions of varying length.[56] We rationalize the large chemical shift changes at the termini of these regions as disrupting collapse of these domains into their native secondary structure. Hydropathy of the poly-Ala regions dictates they would preferentially be buried within a globular domain. Denaturant might cause these regions to be more exposed to solvent and therefore be unable to form native contacts, which also helps to explain the solubility of these large aggregation-prone proteins. Because no other regions exhibit large chemical shift disruptions, we hypothesize that poly-Ala collapse is one of the first changes that occurs in the formations of spider silk.

Dynamic experiments show that dilute solutions in 4M urea have increased T1, and T2 times, corresponding to highly flexible unfolded random coils with increased chain dynamics. Heteronuclear NOE values also indicate increased dynamics in solution. Based on previous results showing primarily monomer-sized populations in dilute solutions (~1% wt) we believe that our experiments primarily probe this population. DLS nevertheless confirms the presence of larger assemblies even in dilute conditions. At higher concentration our data clearly shows that relaxation parameters are better fit to two-component exponential decay functions than one-component. We interpret this to mean that the larger micelle-like hierarchal assemblies remain stable at a spectroscopically relevant concentration. Further analysis of concentration dependence 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 synthetic dope.

Supplemental Figures.

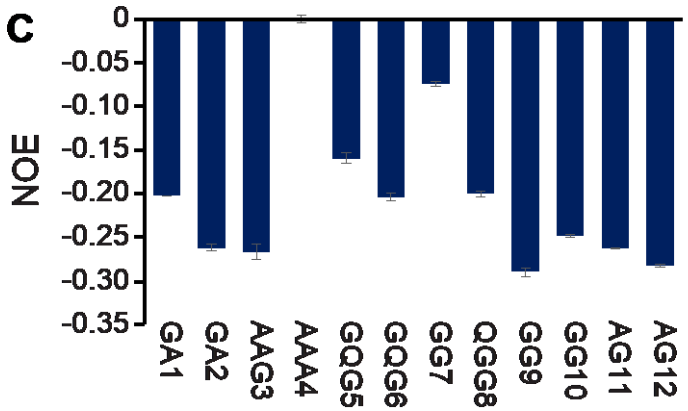


Figure 36 Heteronuclear NOE values for samples in 4M urea are negative for all assigned resonances including the AAA resonance which is relatively small compared to the others.

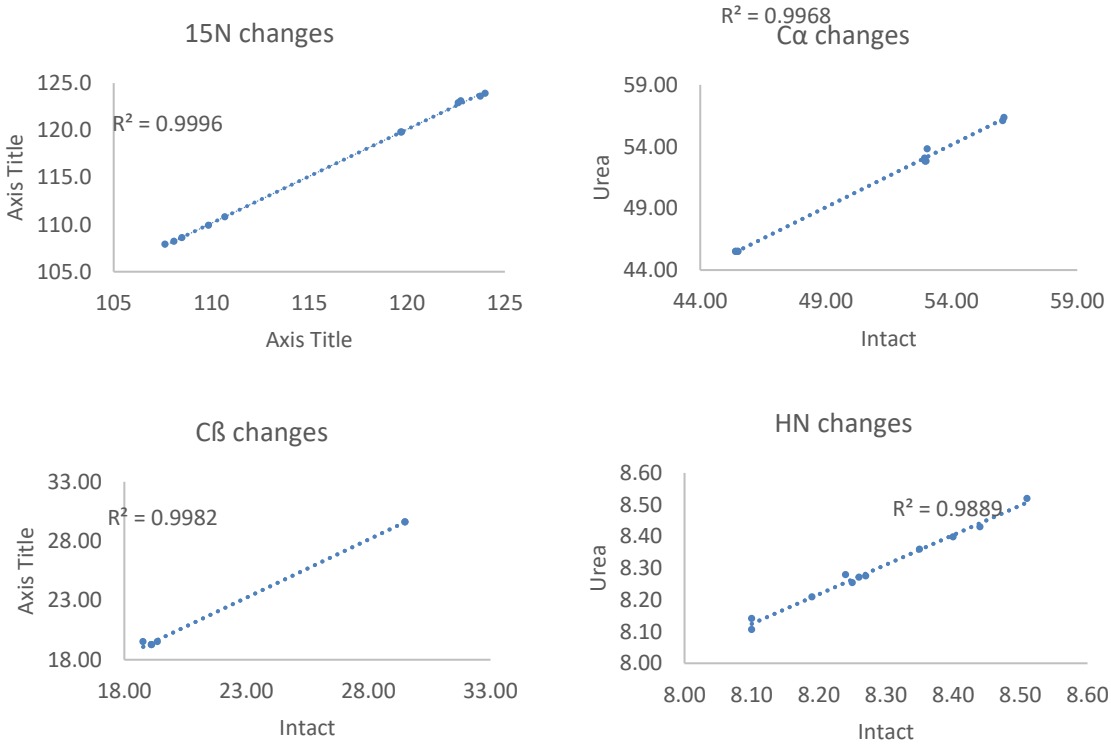


Figure 37 Comparison of the chemical shifts of the assign 12 peaks in urea (y-axis) compared to the literature values in the intact glands (x-axis).

Figure 38 The complete sequence of the two major spidroin components of black widow MA silk, MaSp1, and MaSp2. The highlighted regions show all motifs that can be assigned from the 3D NMR experiments. Underlined regions show regions show sequences that were partially ambiguous from the 3D NMR and were assigned based on sequence prevalence.

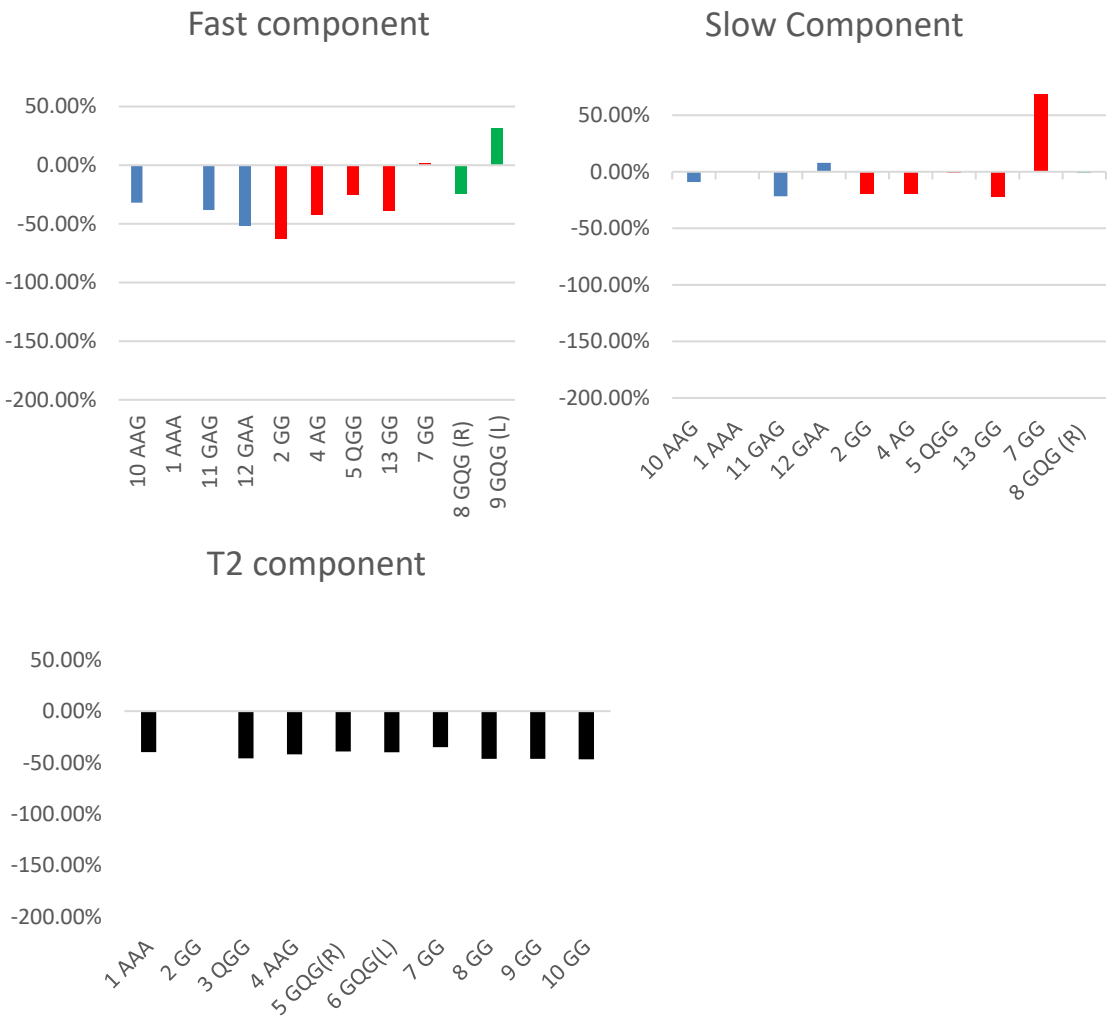


Figure 39 Differences in the relaxation parameters between the two concentration regimes in this study. Top graphs show the slow and fast relaxing components of T1 while the bottom graph shows the T2 component. All samples are dissolved in 4M urea.

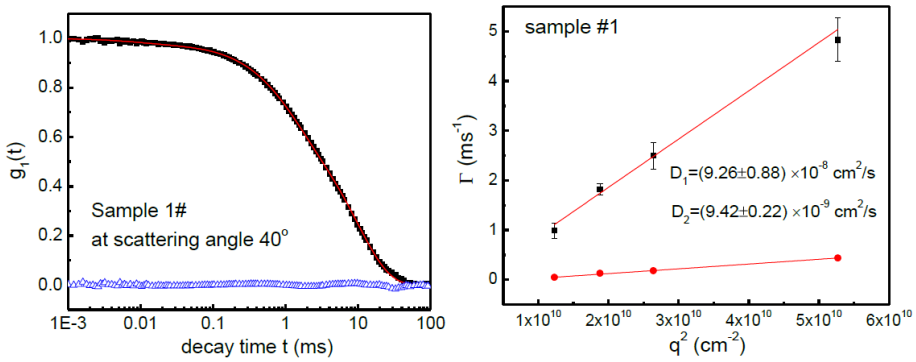


Figure 40 Examples of Fitting curves for DLS experiments.

Acknowledgements:

This work is currently being prepared for publication. David Onofrei, is the primary author and wrote the manuscript. Co contributors to this work are Dillan Stengel, Hannah R. Johnson, Jia Di, S. Trescott, A. Soni, M. Muthukumar, who all provided various data on included in the figures and text and Gregory P. Holland who organized and led the project.

References

1. 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*, 3879-3885.

2. 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). *Molecular Biology and Evolution* **2013,** *30*, 589-601.

3. 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*, e11234.

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

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

6. Ai-Chun Zhao, §; Tian-Fu Zhao; Koichi Nakagaki; Yuan-Song Zhang; Yang-Hu SiMa, ‖; Yun-Gen Miao, ⊥; Kunihiro Shiomi; Zenta Kajiura; Yoko Nagata; Masayuki Takadera, a.; Masao Nakagaki*, Novel Molecular and Mechanical Properties of Egg Case Silk from Wasp Spider, Argiope bruennichi†. **2006**.

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

8. 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.

9. 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*, e50227.

10. Hardy, J. G.; Romer, L. M.; Scheibel, T. R., Polymeric materials based on silk proteins. *Polymer* **2008,** *49*, 4309-4327.

11. 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. *Advanced Materials* **2015,** *27*, 2189-2194.

12. Guinea, G. V.; Elices, M.; Plaza, G. R.; Perea, G. B.; Daza, R.; Riekel, C.; Agulló-Rueda, F.; Hayashi, C.; Zhao, Y.; Pérez-Rigueiro, J., Minor Ampullate Silks from Nephila and Argiope Spiders: Tensile Properties and Microstructural Characterization. *Biomacromolecules* **2012,** *13*, 2087-2098.

13. Collin, M. A.; Clarke, T. H.; 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. *International Journal of Biological Macromolecules* **2018,** *113*, 829-840.

14. 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. *Scientific Reports* **2019,** *9* (1), 8380.

15. 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. *Journal of Proteome Research* **2015,** *14*, 4223-4231.

16. Garb, J. E.; Haney, R. A.; Schwager, E. E.; Gregorič, M.; Kuntner, M.; Agnarsson, I.; Blackledge, T. A., The transcriptome of Darwin's bark spider silk glands predicts proteins contributing to dragline silk toughness. *Communications Biology* **2019,** *2* (1), 275.

17. Larracas, C.; Hekman, R.; Dyrness, S.; Arata, A.; Williams, C.; Crawford, T.; Vierra, C. A., Comprehensive Proteomic Analysis of Spider Dragline Silk from Black

Widows: A Recipe to Build Synthetic Silk Fibers. *International journal of molecular sciences* **2016,** *17*.

18. Blamires, S. J.; Kasumovic, M. M.; Tso, I. M.; Martens, P. J.; Hook, J. M.; Rawal, A., Evidence of decoupling protein structure from spidroin expression in spider dragline silks. *International Journal of Molecular Sciences* **2016,** *17*, 1-13.

19. Clarke, T. H.; Garb, J. E.; Haney, R. A.; Chaw, R. C.; Hayashi, C. Y.; Ayoub, N. A., Evolutionary shifts in gene expression decoupled from gene duplication across functionally distinct spider silk glands. *Scientific Reports* **2017,** *7*, 8393.

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*, e514.

21. Dagmara Motriuk-Smith, †; Alyson Smith; Cheryl Y. Hayashi, a.; Lewis†, R. V., Analysis of the Conserved N-Terminal Domains in Major Ampullate Spider Silk Proteins. **2005**.

22. Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V., Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *International Journal of Biological Macromolecules* **1999,** *24*, 271-275.

23. Shi, X.; Yarger, J. L.; Holland, G. P., Elucidating proline dynamics in spider dragline silk fibre using ² H– ¹³ C HETCOR MAS NMR. *Chem. Commun.* **2014,** *50*, 4856-4859.

24. 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.

25. Hijirida, D. H.; Do, K. G.; Michal, C.; Wong, S.; Zax, D.; Jelinski, L. W., 13C NMR of Nephila clavipes major ampullate silk gland. *Biophysical Journal* **1996,** *71* (6), 3442-3447.

26. Rising, A.; Johansson, J., Toward spinning artificial spider silk. *Nature Chemical Biology* **2015,** *11*, 309-315.

27. 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.

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

29. 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.

30. 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.

31. Knight, D. P.; Vollrath, F., Changes in element composition along the spinning duct in a *Nephila* spider. *Naturwissenschaften* **2001,** *88*, 179-182.

32. 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.

33. Zhou, P.; Xie, X.; Knight, D. P.; Zong, X. H.; Deng, F.; Yao, W. H., Effects of pH and Calcium Ions on the Conformational Transitions in Silk Fibroin Using 2D Raman Correlation Spectroscopy and ^{13}C Solid-State NMR. *Biochemistry* **2004,** *43* (35), 11302-11311.

34. Peng, X.; Shao, Z.; Chen, X.; Knight, D. P.; Wu, P.; Vollrath, F., Further Investigation on Potassium-Induced Conformation Transition of *Nephila* Spidroin Film with Two-Dimensional Infrared Correlation Spectroscopy. . *Biomacromolecules* **2005,** *6* (1), 302-308.

35. Foo, C. W. P.; Bini, E.; Hensman, J.; Knight, D. P.; Lewis, R. V.; Kaplan, D. L., Role of pH and charge on silk protein assembly in insects and spiders. *Applied Physics A: Materials Science & Processing* **2006,** *82* (2), 223-233.

36. Askarieh, G.; Hedmammar, 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*, 236-238.

37. 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. *Journal of Structural Biology* **2010,** *170*, 413-419.

38. Hagn, F.; Thamm, C.; Scheibel, T.; Kessler, H., pH-Dependent Dimerization and Salt-Dependent Stabilization of the N-terminal Domain of Spider Dragline Silk-Implications for Fiber Formation. *Angewandte Chemie International Edition* **2011,** *50*, 310-313.

39. Schwarze, S.; Zwettler, F. U.; Johnson, C. M.; Neuweiler, H., The N-terminal domains of spider silk proteins assemble ultrafast and protected from charge screening. *Nat Commun* **2013,** *4*, 2815.

40. Kronqvist, N.; Otikovs, M.; Chmyrov, V.; Chen, G.; Andersson, M.; Nordling, K.; Landreh, M.; Sarr, M.; Jornvall, H.; Wennmalm, S.; Widengren, J.; Meng, Q.; Rising, A.; Otzen, D.; Knight, S. D.; Jaudzems, K.; Johansson, J., Sequential pH-driven dimerization and stabilization of the N-terminal domain enables rapid spider silk formation. *Nat Commun* **2014,** *5*, 3254.

41. Xu, D.; Guo, C.; Holland, G. P., Probing the Impact of Acidification on Spider Silk Assembly Kinetics. *Biomacromolecules* **2015**.

42. Bauer, J.; Scheibel, T., Dimerization of the Conserved N-Terminal Domain of a Spider Silk Protein Controls the Self-Assembly of the Repetitive Core Domain. *Biomacromolecules* **2017,** *18* (8), 2521-2528.

43. Andersson, M.; Jia, Q.; Abella, A.; Lee, X.-Y. 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. *Nature Chemical Biology* **2017,** *13*, 262-264.

44. Wong Po Foo, C.; 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.

45. 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.

46. 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.

47. Alexandrescu, A. T.; Shortle, D., Backbone dynamics of a highly disordered 131 residue fragment of staphylococcal nuclease. *Journal of Molecular Biology* **1994**.

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

49. Yarger, J. L.; Holland, G. P.; Shi, X.; Yarger, J. L.; Holland, G. P., Probing site-specific 13C/15N-isotope enrichment of spider silk with liquid-state NMR spectroscopy. *Analytical and Bioanalytical Chemistry* **2013,** *405*, 3997-4008.

50. 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. *Journal of Molecular Graphics and Modelling* **2001,** *19*, 26-59.

51. Wang, Y.; Jardetzky, O., Probability-based protein secondary structure identification using combined NMR chemical-shift data. *Protein science : a publication of the Protein Society* **2002,** *11*, 852-61.

52. Grzesiek, S.; Bax, A., Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR. *J. Am. Chem. Soc.* **1990,** *29*, 4659-4667.

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

54. 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.

55. 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,** *0*, 5568-5570.

56. Sintya, E.; Alam, P., Self-assembled semi-crystallinity at parallel β-sheet nanocrystal interfaces in clustered MaSp1 (spider silk) proteins. *Materials Science and Engineering: C* **2016,** *58*, 366-371.

Chapter 4 Aciniform prey-wrapping silk as a coiled-coil / β-Sheet hybrid

Introduction

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 fibers 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 structure conformation of the toughest natural biopolymer: aciniform prey-wrapping silks.

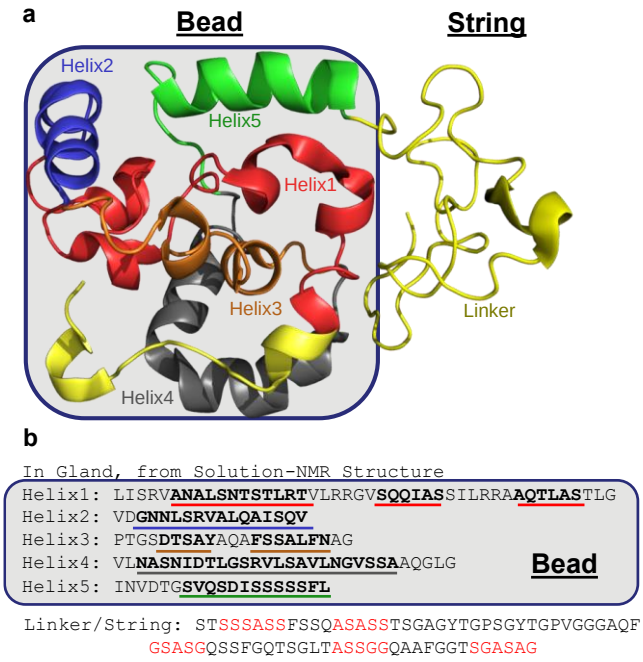


Figure 41 (a) Solution NMR structure of a single recombinant *Argiope trifaciata* prey-wrapping silk repeat unit taken from PDB code 2MU3 demonstrating the "beads on a string" structure.[7] 1 (b) Primary-protein sequence for *Argiope argentata* AcSp1 repeat unit. Regions assumed to be helical in the gland are bolded and color-underlined based on structure homology between *A. trifaciata* and *A. argentata* (**Figure 46**). Proposed β-sheet-forming motifs from the linker region are shown with red lettering.

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 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 (**Figure 41a**). 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 41b**, 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 spidron 1 from *Argiope trifaciata* spiders (**Figure 41a**).[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 41a**, 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-

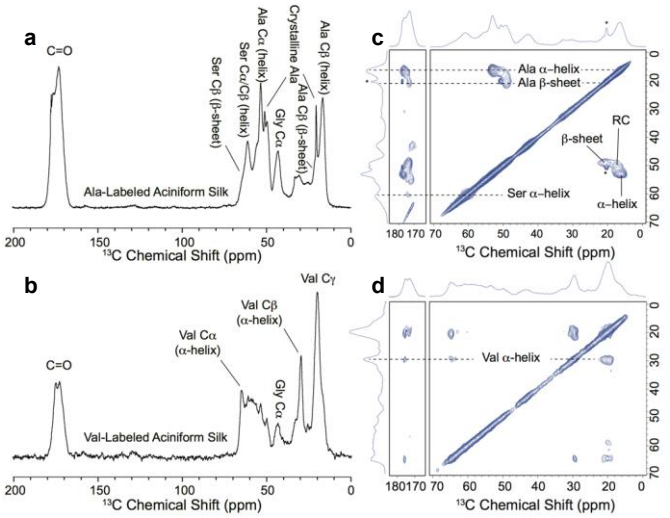


Figure 42 ^{1}H-^{13}C CP-MAS spectra (a, b, 30 kHz MAS) and 2D ^{13}C-^{13}C DARR (c, d, 14 kHz MAS, 100 ms DARR mixing time) spectra of freshly-spun *Argiope argentata* aciniform silks that were isotopically enriched with either ^{13}C-Ala (a, c) or ^{13}C-Val (b, d). Signal from crystalline ^{13}C-Ala contaminant is indicated with an asterisks.

labelled alanine (serine is also labelled through metabolism of alanine), and/or B) ^{13}C-labelled 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 41**); 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 (**Figure 45**) but not in helix 3.

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 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 42c**).[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 (Table SX). 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 43**).‡

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 non-β-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.

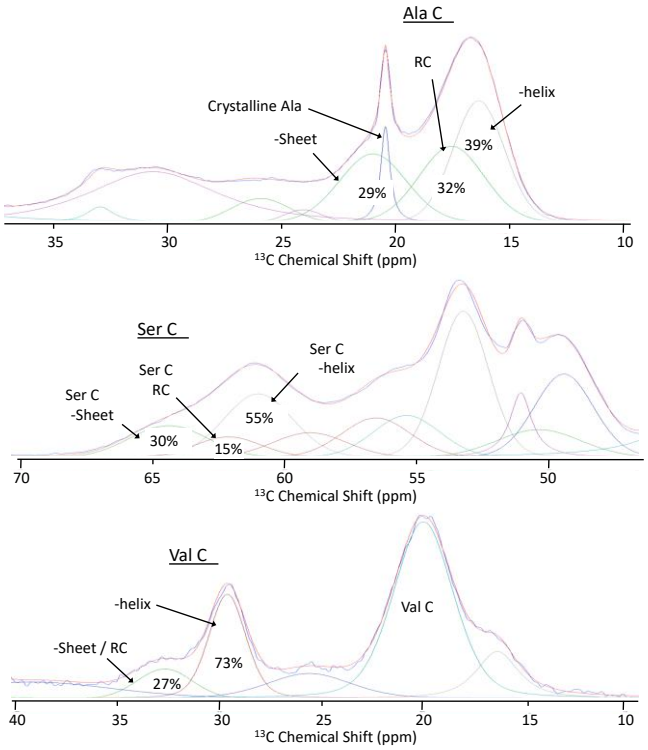


Figure 43 Spectral deconvolutions of Ala, Ser and Val Cβ resonances from ^{1}H-^{13}C CP-MAS data collected on isotopically-enriched freshly-spun *Argiope argentata* aciniform silks.

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 42**). Again, we extracted exact ^{13}C chemical shifts for the dominant Val Cα, Cβ, and CO resonances (Table SX), which as anticipated, align very clearly with α-helical secondary structure. We utilized the Val Cβ resonance to quantify helical versus non-helical structures because it is the most well resolved (**Figure 43c**). Peak fitting results suggest that at least 73% of all Val residues adopt an α-helical environment, with a minor RC and/or α-

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 valine environment is assumed to be a combination of random-coil and/or β-sheet structures, although we expect RC structures to outweigh β-sheets.

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 fibrilization, 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 (**Figure 41**) 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. Considering our 73% observation, it is clear that the loosely structured Val residues on the edges of helices are enticed to

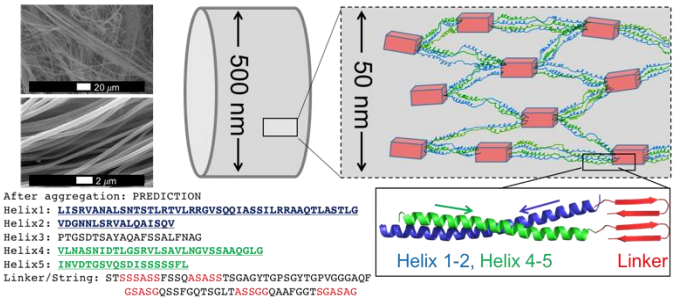


Figure 44 Proposed hierarchical molecular protein structure of aciniform prey-wrapping silks as a hybrid coiled-coil and nanocrystalline β-sheet fibroin. Helix 1 and 2 form a helical coil (blue), Helix 3 is unstructured, acting as a possible turn, helices 4 and 5 form another single coil (green), and the Ser / Ala-rich regions of the linker form pleated β-sheet (red) subunits.

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 45**). With this data we can hypothesize which specific regions in the AcSp1 primary protein sequence adopt α-helical 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. 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 silks: SSSASS, ASASS, GSASG, ASSGG, SGASAG, and SSSSS. Indeed, these motifs (highlighted with red lettering in

Figure 44) show a high propensity for β-sheet aggregation in our preliminary Monte-Carlo Molecular Dynamics simulations. 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. 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. 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 44**). 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. This model is illustrated in **Figure 44**. Significant work remains to support or modify the proposed subunit.

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, [22] 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.

Supplemental

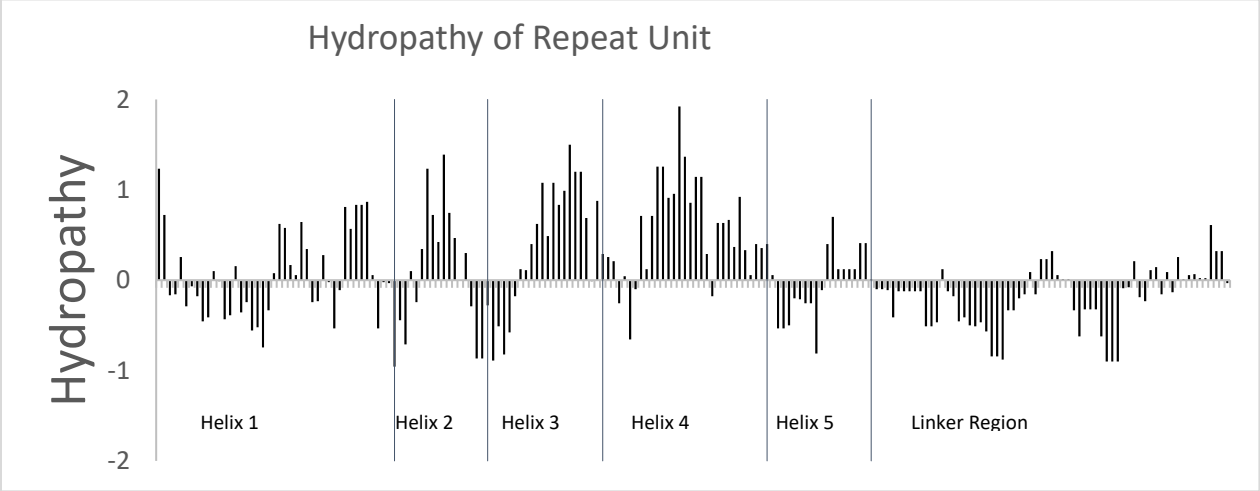


Figure 45 Kyte-Doolittle hydropathy plot of repeat unit of AC silk. A distribution of hydrophilic amino acids are found in the first half of the sequence in which the helical residues are found. The linker region near the end is primarily hydrophobic.

Figure 46 Sequence alignment between two different *argiope* spp. showing how well conserved the repeat unit is and allowing for direct comparison between species.

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.

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. S. Wang, W. Huang and D. Yang, Journal of biomolecular NMR, 2012.

9. 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.

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

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

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

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

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

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

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

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

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

19. 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.

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

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

22. 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.

The natively unfolded and unstructured character of spider silk proteins as stored in the MA gland (pre-fibrillar) closely resembles IDP polymers such as elastin [1, 2] and collagen [3] with little to no evidence of secondary structure. Previous work has shown

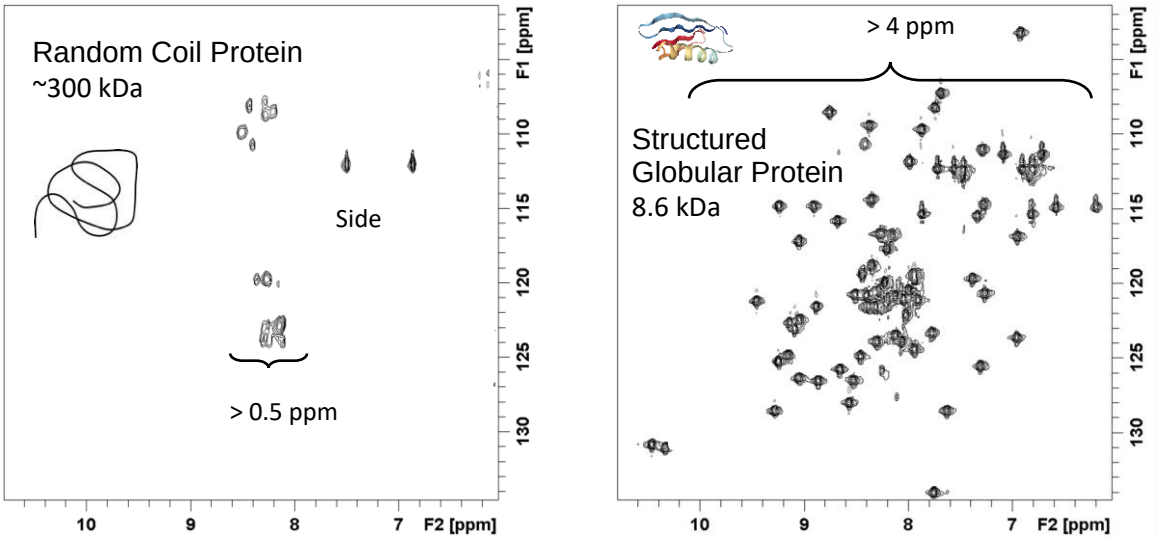


Figure 47 $^{1}H/^{15}N$-HSQC 2D NMR spectra of MA silk proteins in solution (left) and GB1 (right). The differences in chemical shift dispersion are remarkable considering the size difference of the two proteins. This serves to highlight the challenge of studying large IDP-type polymers with traditional protein solution NMR experiments.

that within intact MA glands, spidroins assume a primarily random coil structure and exhibit extremely rapid sub-nanosecond backbone dynamics illustrating a high degree of protein chain flexibility similar to a typical IDP [4-7]. Besides having a highly degenerative primary sequence, disorder and lack of secondary structure greatly reduces the ^{1}H chemical shift dispersion of in conventional protein solution NMR (e.g. $^{1}H/^{15}N$ HSQC). **Figure 47** shows a comparison between a solution of native Ma spider silk dope dissolved in 4 M urea and the beta subunit of protein G (~6 kDa, GB1 PDB:2J52). The considerably larger spider silk proteins (300-350 kDa) have a ^{1}H

chemical shift dispersion < 0.5 ppm (not including side chain resonances) while the comparatively tiny GB1 (~6 kDa) has a chemical shift dispersion > 4 ppm indicative of a well-ordered globular protein with well-defined secondary and tertiary structure. This narrow range of amide ^{1}H chemical shifts observed for spider silk proteins results in crowding and overlap in NMR spectra traditionally used to interrogate protein secondary and tertiary structure. Without greater resolution, it becomes very difficult to assign resonances to individual amino acid residues and therefore understand precise local chemical environments and dynamics. This problem of severe spectral overlap has long been a challenge to studying IDPs by NMR [1, 2, 8].

Despite this inherent lack of secondary structure, we have shown that prior to fiber formation, spidroins undergo a supramolecular pre-assembly as stored in the Ma gland.[9] These structures are crucial to proper formation of silk fibers with native mechanical properties.[10] They are also believed to be sensitive to the gland condition including pH, salt, and protein concentration. Recombinant fibers, which do not fully replicate the tensile properties of native spider silk have also been observed to form similar structures but there is not yet a clear understanding of the molecular interactions leading to their formation nor their significance in forming better fibers.[11] Without this key piece of information of how the different variables affect the molecular level structure of the spidroins, there is little hope of being able to devise a spinning system that successfully mimics the spiders' evolutionarily optimized silk fibers. It is critical then to understand how the individual amino acid residues interact with each other and the solvent. These molecular contacts and dynamical processes can be accessed with

advanced NMR techniques not commonly used to investigate structured globular proteins in solution.

The first step to increased resolution in the NMR spectra of IDPs is to observe nuclei other than protons. Direct detection experiments using heteronuclei (^{13}C and ^{15}N) take advantage of the greater chemical shift dispersion for the ^{13}C and ^{15}N dimensions.

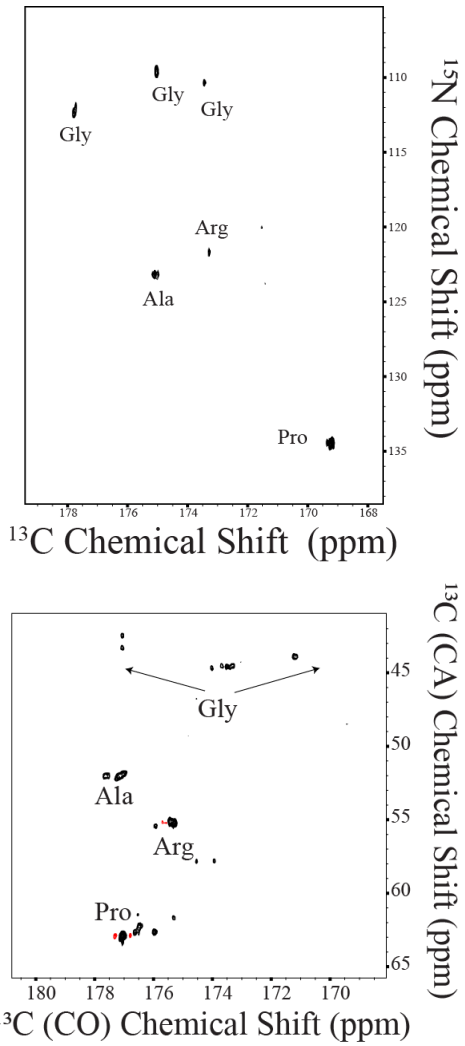


Figure 48 Direct detect solution NMR experiments for isotopically enriched, proline labeled *A. argentata.* The top spectrum shows the 13C and 15N correlations with a directly detected 13C dimension. Observation of proline is clearly evident from proline's unique 15N chemical shift. The bottom spectrum shows 13C-13C correlation experiment which provides a way to investigate the CO regions of the proteins.

^{13}C-^{13}C, and ^{13}C-^{15}N experiments are often used in support of, or in some cases instead

of their ^{1}H-detect counterparts to provide all of the information used for resonance assignment [12]. A crucial difference is that, unlike their ^{1}H-detect counterparts, the ^{13}C-detect versions provide a means of observing Pro residues. Pro, common in the Ma silk proteins of certain spider species, is only present in MaSp2 [13] rather than MaSp1 and is known to impact a number of spider silk properties including a dramatic impact on mechanical properties where higher extensibility is observed for Ma silk containing higher Pro content [14-17]. This gives ^{13}C-detect spectra the ability to directly inform on both inter- and intra-molecular interactions within and potentially between the two primary spidroin protein components. **Figure 48** shows a comparison between two different ^{13}C detect experiments, CACO and CON. These experiments were performed on a solution of native *A. argentata* Ma glands dissolved in 4 M urea. Previously the spiders had been fed a solution of ^{13}C-^{15}N proline to isotopically label this residue. Proline resonances are easily identifiable due to their unique ^{13}Cα and ^{15}N chemical shifts.

A broad stretch of Gly-rich resonances is also identifiable and informs on some conformation variability for the Gly-rich motifs in the spidroins. The Gly-Pro-Gly motif displays a unique resonance because it is forced to kink forming a β-turn. These X-detect experiments are extremely promising for expanding our ability to characterize both proteins. And because Pro is only in MaSp2, when we observe and monitor these resonances we are characterizing the MaSp2 protein exclusively. This should allow us to further understand the role of MaSp2 in spider silk assembly. In addition, Leu is only present in MaSp1, so we can label Leu and Pro to look for MaSp1-MaSp2 protein-protein interactions through ^{13}C-^{13}C NOESY type experiments. Thus, it is possible to

truly explore with solution NMR at an atomic/molecular level the two different proteins, their structural/dynamic differences and potentially interactions between proteins in the superstructure and following introduction of the pertinent biochemical conditions as discussed above.

Methods

Argiope spiders were fed a solution of U-^{13}C,^{15}N Proline three times a week for two weeks. During each silking spiders were forcibly silked in order to deplete protein contents of their major ampullate gland and facilitate incorporation of the proline label into the MaSp2 spidroin. At the end of the two weeks, spiders were sacrificed, dissected, and their major ampullate glands were removed. The outer membrane covering was removed from the glands and they were dissolved into a 4M urea solution. Protein concentration was estimated at around 1 wt%, %, ~ 30-50 times lower than native concentration in the gland.

NMR Experiments were performed on a Bruker Neo 800 MHz instrument using a cryogenically-cooled triple resonance probe (TCI) with sensitivity enhancement on the carbon and nitrogen channels. For the CACO experiments 2048 points were processed out of 1024 collected, with the balance used for zero-filling. 256 scans were taken for 128 slices using a using in-phase/antiphase virtual decoupling to remove carbon-carbon coupling. Carbonyl sweep width was 40 ppm centered around 174 ppm. The second carbon dimension was CON experiments were performed using the same number of points but using 1024 scans over 64 increments.

Future Directions

Using a strategy similar to the one described by Felli et al [18] we are developing an X-detect backbone assignment strategy optimized for spider silk proteins. In brief, this strategy involves a complex series of magnetization transfer pathways that provide excellent resolution for assigning sequential amino acid residues. A 3D “basis” spectrum is collected to generate a landscape of amide resonances similar to the way the 2D HSQC provides the initialization step for traditional ^{1}H-detect protein NMR methods. Each peak in this spectrum is then used as the basis for 4D and 5D experiments, which are collected as 2D slices. This “protonless” strategy can further be complimented by “proton-start” experiments which take advantage of signal-to-noise enhancement by using ^{1}H as the initial source of magnetization. Despite a lower gyromagnetic ratio, ^{15}N direct-detect methods compliment the ^{13}C direct-detect suite of experiments and have several unique advantages. They avoid the splitting resulting from strong one-bond $J_{13_C-13_C}$ couplings, which require the collection of two interleaved spectra in the in-phase antiphase (IPAP) pulse sequence.[19] Nitrogen experiments are acquired as a single spectrum and are more straightforward to collect than ^{13}C experiments in a uniformly labeled sample. The chemical shift dispersion of ^{15}N is also slightly greater than the aliphatic or carbonyl carbon chemical shift regions. Presently, ^{15}N-detect versions of two of the most informative experiments, CON and CAN, have been demonstrated to have great value in structural analysis of IDPs [20] and thus, should provide similar advantages to the study of spidroins. These two experiments form the basis of a sequential assignment strategy similar to the HNCACB/CBCACONH used in more traditional protein NMR and used for the three residue MaSp1 assignment discussed above. In this modified version, the CAN experiment correlates backbone amide ^{15}N nuclei with

both the sequential *i* Cα as well as the preceding *i*-1 Cα. This experiment exists as either protonless or proton-start versions (hCAN)[21]. These proton-start versions have the benefit of reducing the acquisition time by using a traditional INEPT sequence to transfer polarization from the Hα to the Cα. Taken together, these direct-detect experiments form an elegant method for probing unstructured dynamic proteins that elude nearly all other methods of direct investigation. When coupled with different isotope labeling schemes, these approaches should provide an unprecedented level of detail for the inner workings of MaSp1 and MaSp2 silk proteins in a variety of different biochemical environments.

Generally, the highly efficient relaxation pathways of folded globular proteins make direct ^{13}C-detect experiments inefficient. However, IDPs that behave like spidroins suffer from significantly less signal loss due to much longer relaxation times because of the increased backbone dynamics (flexibility). In our previous work we have found backbone relaxation (T_1 and T_2) on par with those observed for other IDPs and are more in line with those typical of a small protein an order of magnitude smaller in size. However, even with beneficial relaxation properties, the significantly lower gyromagnetic ratio of ^{13}C and ^{15}N nuclei and lower labeling efficiency of spider silk proteins necessitates the use of a cryogenically cooled probe such as the one used for the experiments in **Figure 48**. Access to this specialized NMR hardware opens up a whole new class of multi-dimensional NMR experiments that has been shown to expand the resolving power of NMR on crowded, broadened, and overlapping spectra. This specialized equipment is made available to the Holland Lab in the UCSD Bimolecular NMR Center and provides some of the highest solution NMR sensitivity and resolution

available in the world for X-detect experiments which is a requirement for these cutting-edge X-detect IDP experiments. It should be noted that this is the first probe of its kind and it is at high field (800 MHz) with only a handful of similar instruments available in the world. This puts our research team in extremely unique position to take the NMR characterization of these silk proteins to an unprecedented level. Coupled with selective labeling strategies this increase in signal intensity, resolution, and site-specificity will likely allow for direct observation of the silk formation process.

To summarize, while NMR is one of the best techniques for non-destructive molecular-level investigation of both structure and dynamics, the unstructured, highly dynamic nature of spider silk proteins in solution exceeds the investigative boundaries of standard 2D and even 3D experiments. The source of the problem comes mainly from three sources: the large number of amino acid residues found in the silk proteins, and the highly degenerate and repetitive nature of the sequence, and the unfolded, unstructured dynamics. The success of NMR in application to the related field of IDPs gives us confidence that these problems are not insurmountable. To tackle these problems, we have been, and are continuing, to develop an NMR based investigative strategy involving using three main tools:

1. Selective labeling of amino acids, informed by the sequence of a given silk type to study changes in specific motifs of the spidroins.
2. Complimented ^{1}H detected and X-detect experiments to measure subtle differences in amino acid environments.
3. 4D and 5D NMR experiments to access greater resolution for crowded regions to overcome sequence degeneracy and slow relaxation.

With data from these molecular level experiments in hand, we will be able to inform future studies aimed at optimizing the silk spinning process.

References:

1. Dyksterhuis, L. B.; Carter, E. A.; Mithieux, S. M.; Weiss, A. S., Tropoelastin as a thermodynamically unfolded premolten globule protein: The effect of trimethylamine N-oxide on structure and coacervation. *Arch Biochem Biophys* **2009,** *487* (2), 79-84.

2. Roberts, S.; Dzuricky, M.; Chilkoti, A., Elastin-like polypeptides as models of intrinsically disordered proteins. *FEBS Lett* **2015,** *589* (19 Pt A), 2477-86.

3. Wong Po Foo, C.; Kaplan, D. L., Genetic engineering of fibrous proteins: spider dragline silk and collagen. *Advanced Drug Delivery Reviews* **2002,** *54*, 1131-1143.

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

5. Hronska, M.; vanBeek, J. D.; Williamson, P. T. F.; Vollrath, F.; Meier, B. H., NMR Characterization of Native Liquid Spider Dragline Silk from *Nephila edulis*. *Biomacromolecules* **2004,** *5* (3), 834-839.

6. 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*, 1947-1954.

7. Xu, D.; Yarger, J. L.; Holland, G. P., Are Spider Silk Proteins a New Class of Intrinsically Disordered Proteins? *Biophysical Journal* **2014,** *106*, 686a.

8. Felli, I. C.; Pierattelli, R., *Advances in Experimental Medicine and Biology*. Springer: 2014.

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

11. Andersson, M.; Jia, Q.; Abella, A.; Lee, X.-Y. 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. *Nature Chemical Biology* **2017,** *13*, 262-264.

12. Bastidas, M.; Gibbs, E. B.; Sahu, D.; Showalter, S. A., A primer for carbon-detected NMR applications to intrinsically disordered proteins in solution. *Concepts in Magnetic Resonance Part A: Bridging Education and Research* **2015,** *44*, 54-66.

13. Hinman, M.; Lewis, R. V., Isolation of a clone encoding a second dragline silk fibroin. *Nephila clavipes* dragline silk is a two-protein fiber. *J. Biol. Chem.* **1992,** *267* (27), 19320-19324.

14. Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V., Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *International Journal of Biological Macromolecules* **1999,** *24*, 271-275.

15. Liu, Y.; Sponner, A.; Porter, D.; Vollrath, F., Proline and processing of spider silks. *Biomacromolecules* **2008,** *9* (1), 116-21.

16. 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 β-turn structure in spider dragline silk. *Chem Commun* **2010,** *46*, 6714.

17. Shi, X. Y.; Yarger, J. L.; Holland, G. P., Elucidating proline dynamics in spider dragline silk fibre using H-2-C-13 HETCOR MAS NMR. *Chem Commun* **2014,** *50* (37), 4856-4859.

18. Piai, A.; Hošek, T.; Gonnelli, L.; Zawadzka-Kazimierczuk, A.; Koźmiński, W.; Brutscher, B.; Bermel, W.; Pierattelli, R.; Felli, I. C., CON-CON assignment strategy for highly flexible intrinsically disordered proteins. *Journal of Biomolecular NMR* **2014,** *60*, 209-218.

19. Takeuchi, K.; Heffron, G.; Sun, Z. Y.; Frueh, D. P.; Wagner, G., Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments. *J Biomol NMR* **2010,** *47* (4), 271-82.

20. Chhabra, S.; Fischer, P.; Takeuchi, K.; Dubey, A.; Ziarek, J. J.; Boeszoermenyi, A.; Mathieu, D.; Bermel, W.; Davey, N. E.; Wagner, G.; Arthanari, H., (15)N detection harnesses the slow relaxation property of nitrogen: Delivering enhanced resolution for intrinsically disordered proteins. *Proc Natl Acad Sci U S A* **2018,** *115* (8), E1710-E1719.

21. Gal, M.; Edmonds, K. A.; Milbradt, A. G.; Takeuchi, K.; Wagner, G., Speeding up direct (15)N detection: hCaN 2D NMR experiment. *J Biomol NMR* **2011,** *51* (4), 497-504.