AFRL-AFOSR-VA-TR-2024-0186

Hierarchical Assembly of Spider Silk Proteins: Exploring Structural Biology of Biomaterials from the Atomic to the Mesoscale

Holland, Gregory
SAN DIEGO STATE UNIVERSITY FOUNDATION
5250 CAMPANILE DR MC1947
SAN DIEGO, CA, 92182
USA

03/04/2024
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
20240304	Final	START DATE 20200801	END DATE 20230831

4. TITLE AND SUBTITLE
Hierarchical Assembly of Spider Silk Proteins: Exploring Structural Biology of Biomaterials from the Atomic to the Mesoscale

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
	FA9550-20-1-0103	61102F
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Gregory Holland

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
SAN DIEGO STATE UNIVERSITY FOUNDATION
5250 CAMPANILE DR MC1947
SAN DIEGO, CA 92182
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Air Force Office of Scientific Research 875 N. Randolph St. Room 3112 Arlington, VA 22203	AFRL/AFOSR RTB2	AFRL-AFOSR-VA-TR-2024-0186

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The ability to mimic the true hierarchical assembly of natural biomaterials in the lab remains a significant challenge. A major knowledge gap lies in our understanding of how the different length scales connect from the atomic to the macroscopic. Using an optical light microscope, one can readily observe length scales from the micron to the macroscale. However, connecting the atomic, molecular and mesoscales and the interplay between them is far more challenging in the context of biomaterials formation and performance. We aim to track the atomic, molecular and mesoscale assembly process of protein-based materials using spider silk as our target system. Spider silk is one of the Holy Grails of materials science, exhibiting mechanical properties that rival high tensile steel and toughness that surpasses Kevlar®, the material used to fabricate bullet-proof vests. The grand goal of this work is to probe and mimic the atomic, molecular and mesoscale structural and dynamical features and events responsible for natural materials assembly. We stress that this is envisioned as setting the stage for a new field of enquiry, namely, the Structural Biology of Biomaterials.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU	37

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
BENNETT IBEY	000-0000

Standard Form 298 (Rev.5/2020)
Prescribed by ANSI Std. Z39.18

1. Cover Sheet:

To: technicalreports@afosr.af.mil
Subject: Final Performance Report to Dr. Bennett Ibey

Contract/Grant Title: **"Hierarchical Assembly of Spider Silk Proteins: Exploring the Structural Biology of Natural Materials from the Atomic to the Macroscale"**

PI: Gregory P. Holland
Department of Chemistry and Biochemistry, San Diego State University, San Diego CA 92182-1030

Co-PI: Nathan C. Gianneschi
Departments of Chemistry, Biomedical Engineering, Materials Science & Engineering and Pharmacology, Northwestern University, Evanston, IL 60208

Contract/Grant #: FA9550-20-1-0103

Reporting Period: 1 June 2020 to 31 August 2023

2. Objectives (taken directly from the original proposal):

The ability to mimic the true hierarchical assembly of natural biomaterials in the lab remains a significant challenge. A major knowledge gap lies in our understanding of how the different length scales connect from the atomic to the macroscopic. Using an optical light microscope, one can readily observe length scales from the micron to the macroscale. However, connecting the atomic, molecular and mesoscales and the interplay between them is far more challenging in the context of biomaterials formation and performance. We aim to track the atomic, molecular and mesoscale assembly process of protein-based materials using spider silk as our target system. Spider silk is one of the Holy Grails of materials science, exhibiting mechanical properties that rival high tensile steel and toughness that surpasses Kevlar®, the material used to fabricate bullet-proof vests. The grand goal of this work is to probe and mimic the atomic, molecular and mesoscale structural and dynamical features and events responsible for natural materials assembly. We stress that this is envisioned as setting the stage for a new field of enquiry, namely, the Structural Biology of Biomaterials. We have the following objectives (**Figure 1**):

Objective 1: Atomic and Molecular Investigations of Silk Protein Dopes – Solution NMR and Modeling

Solution nuclear magnetic resonance (NMR) of native and recombinant spider and silkworm silk dopes will be conducted to determine the structural and dynamic interactions at the atomic and molecular levels that facilitate the silk protein mesoscale assemblies observed by dynamic light scattering (DLS) and cryogenic transmission electron microscopy (cryo-TEM) (Obj. 2). A focus will be placed on solution NMR methods typically applied to intrinsically disordered proteins (IDPs) as native silk dopes have been shown by our lab and others to be unfolded, highly dynamic random coils when stored in silk glands. Variation in biochemical conditions (pH, ions) in the spinning duct is critical for triggering protein folding and the mesoscale interactions that control hierarchical fiber assembly. Thus, NMR will be used to characterize silk protein structure and dynamics under various biochemical conditions. The experimental methods will be supported using a combination of molecular dynamics and basin hopping to address the challenge of simulating such large proteins. The modelling will make best use of the data from all the objectives and will test design rules to help interpret results across all the length scales.

Objective 2: Silk Dopes at the Mesoscale (100 nm – 1 μm) – Cryo-TEM and DLS

DLS will be applied to probe native spider and silkworm dopes in a high-throughput fashion to determine the presence of silk protein superstructures as a function of biochemical phase space. High resolution cryo-TEM with 3D tomography will be used to determine silk protein superstructures at high resolution for biochemical conditions deemed important by DLS and under controlled shearing conditions through MEMs devices. This will occur different spider silk types (major, minor, tubuliform), major spider silks from different species (orb weavers and cob weavers) and different silk producing organisms (silkworms) to characterize nanoscale silk protein pre-assemblies. Because distinct silks exhibit radically different mechanical and physical properties – and are used for entirely different applications – differences in silk processing and pre-assemblies are expected to fabricate the various fibers that are composed of very different proteins

Objective 3: Micron Scale Characterization of Silk Dope (μm – mm) – IR-microscopy Micro-rheology

IR microscopy will correlate changes in silk protein secondary structure through the native spinning system to micron level spatial resolution thus, correlating protein folding while moving through the duct where the silk proteins are in the nascent stages and the fiber begins to form. Through collaboration with the Robertson-Anderson Lab, we plan to continue using optical tweezers micro-rheology of silk protein dopes from spiders and silkworms. These measurements will inform on silk dope viscoelastic behavior on the micron length scale. Our preliminary data

exhibits the onset of an elastic response on micron scales (~20 μm) relative to the silk protein monomer (20 nm) and superstructures (200-300 nm) which suggests the presence of an entangled protein network on the micron length scale.

Objective 4: The Macroscale – Bulk Properties and Structure of Silk Fibers – Mechanical, EM and SSNMR

The ultimate aim is to utilize the information provided by the above experimental and modeling methods conducted on native and recombinant silk dopes to mimic the process of silk formation in the lab at all length scales and produce materials that have structures and bulk properties (mechanical) that match native fibers. As needed, we will conduct mechanical and structural characterization of native and synthetic silk fibers with MTS Nano mechanical testing, electron microscopy (SEM/TEM) and solid-state (SS) NMR.

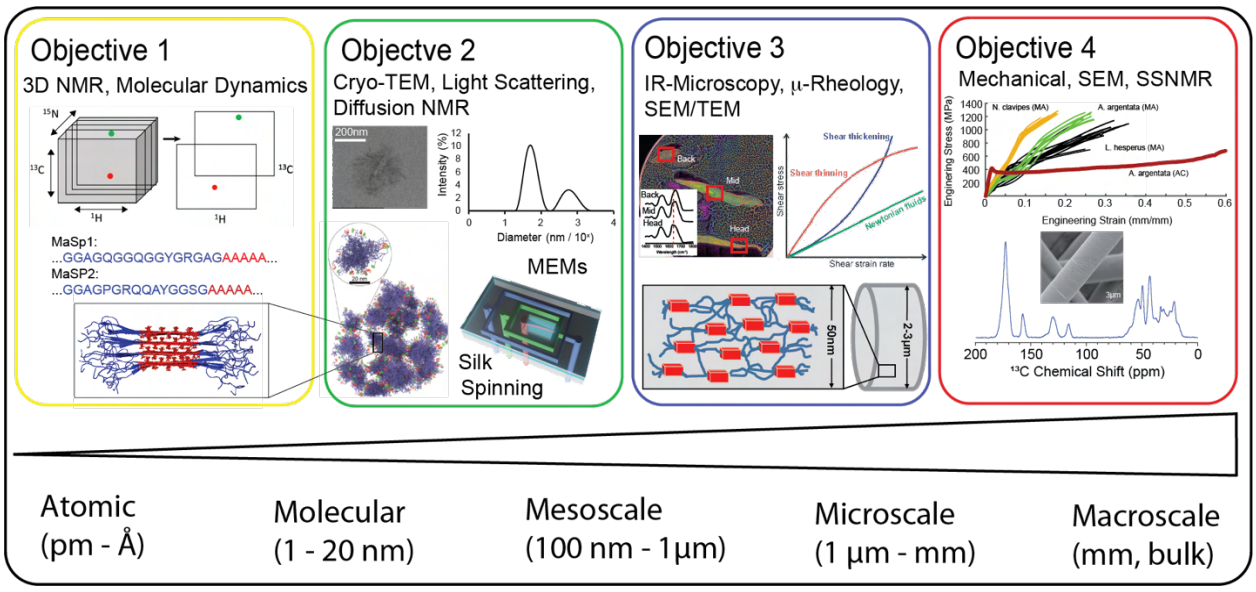

Figure 1. Multimodal approach to experimentally probe the progression of biomolecular assembly from the atomic to the macroscale. Obj. 1, Multi-dimensional, multi-nuclear solution NMR is applied to characterize silk protein structure and dynamics at the atomic and molecular length scales together with MD modeling. Obj. 2, Diffusion NMR, cryo-TEM and DLS are combined to track the molecular and mesoscale protein assembly processes together with MEMs silk processing and spinning under controlled physiochemical conditions. Obj. 3, IR-Microscopy and micro-rheology inform on the secondary structure and viscoelastic properties of the silk protein dopes and spatially map secondary structure through the native spinning system. Obj. 4, Mechanical properties are determined for bulk silk fibers together with electron microscopy and SSNMR to correlate nanoscale fiber features and secondary protein structure to observed mechanical properties. Observed biomaterial properties are correlated with atomic, molecular and mesoscale features of the protein pre-assemblies in the silk dope starting material.

3.1 Accomplishments/New Findings (Year 1):

Progress in MD and NMR of Spider Silk Dopes (obj. 1):

In-silico simulations are a critical component of this work to help guide spectroscopy experiments and interpret the results of complex NMR and CryoEM structural datasets. In addition to assisting protein structure determination, molecular dynamics (MD) is employed to determine intermolecular protein-protein interactions, the lynchpin of biopolymer assembly. We are beginning understand the structure of the two primary proteins that comprise spider dragline silk, MaSp1 and MaSp2, where we already have produced significant preliminary results this past year. The complete primary amino acid sequences of *L. hesperus* MaSp1 and MaSp2 spider silk proteins are known, and initial, unpublished MD structures have been completed (**Figure 2).** We have found that the structures are primarily random coil in agreement with 3D protein solution NMR however, they exhibit a loose coiling structure with poly(Ala) (both proteins) buried and the Tyr (in MaSp1) and Pro (in MaSp2) primarily solvent exposed. Another noteworthy observation from the MD structures is the anisotropic tubular shape of both proteins which could be a critical pre-structuring of the silk proteins to aid fibrillization and fiber formation. The next steps include: probing MaSp1-MaSp2 interactions using the open-source molecular dynamics software GROMACS (GROningen MAchine for Chemical Simulations) for protein simulations. Protein-Protein interaction energies as a function of protein orientation and inter-protein distance will be computed based on the Lennard-Jones (short-range) and Coulomb (electrostatic) interactions between the atoms comprising the protein structures. The computational setup can be readily extended to simulate interaction energies between a variety of silk proteins and are part of the work in year 2. We have continued our biophysical investigations of native *L. hesperus* MaSp proteins in year 1 where we are beginning to learn some of the intra- and intermolecular interactions responsible for their pre-assembly in protein micellar superstructures in the Ma gland. In the native gland environment, the MaSp1 and MaSp2 proteins self-associate to form hierarchical 200-300 nm superstructures despite being intrinsically disordered proteins (IDP's), a characteristic that is believed to be critical for the storage of extremely aggregation-prone proteins and prerequisite for effective spider silk spinning. In a very recent publication (3), we employed dynamic light scattering (DLS), three-dimensional (3D) triple resonance solution NMR and diffusion NMR to probe MaSp size, molecular structure and dynamics of these protein pre-assemblies diluted in 4M urea and identified specific regions of the MaSp proteins important for silk protein pre-assembly. 3D NMR indicates that the Gly-Ala-Ala and Ala-Ala-Gly motifs flanking the poly(Ala) runs, which comprise the β-sheet forming domains in fibers, are perturbed by urea, suggesting that these regions may be important for silk

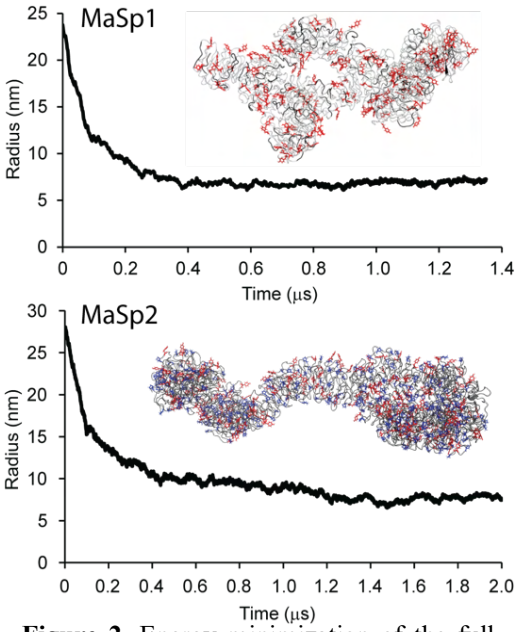

Figure 2. Energy minimization of the full repetitive core for *L. hesperus* MaSp1 (~200 kDa, top) and MaSp2 (~300kDa, bottom) spider silk proteins using MD simulations in GROMACS. Full atomistic structures of the two proteins are shown. Two thirds of the poly(Ala) runs are buried (black), while the Gly-Gly-X domains are solvent exposed with Tyr almost exclusively located on the periphery (red). For MaSp2, the blue amino acids are Pro residues that form kinking, elastin-like β-turn regions that contribute to extensibility.

protein pre-assembly stabilization and begin to explain how spiders store the protein without aggregation at such high concentrations.

The NMR chemical shifts of all identified residues correlate closely to those of random coils, but in a few cases, we observed significant perturbations between the intact gland and 4 M urea NMR spectra. The perturbations from the chemical shifts observed for native Ma silk glands are summarized in **Figure 3** that point to the Gly-containing regions that terminate the poly(Ala) runs as potential sites for intermolecular interactions in pre-assembled superstructures. This was accomplished by intentionally diluting Ma silk protein dope in 4M urea, destabilizing the silk protein superstructures as evidenced by DLS and diffusion NMR, and observing NMR spectroscopic perturbations that are suggestive of interactions that stabilize these structures at native concentration. Because the most common motifs in the repetitive core of the spidroins are three residues long, this permits characterization of the backbone structure and dynamics more deeply than previous studies. For example, the Gly-Gln-Gly motif is very common in the sequence of MaSp1 and MaSp2. By applying typical ^{1}H-detected 2D and 3D protein solution NMR strategies to silk protein solutions, assignments were made for up to three consecutive residues for all of the regions observed in the ^{1}H/^{15}N HSQC spectrum. In fact, three assignments covers ~82% of the full sequence for MaSp1 and ~50% for MaSp2. This is a tremendous step towards the complete NMR structural and dynamic characterization of the primary Ma spider silk proteins and opens the door to continuing investigations under various biochemical conditions (pH, salts, and denaturants) which is work that is presented in year 2 and 3 (below**, also see ref 3**).

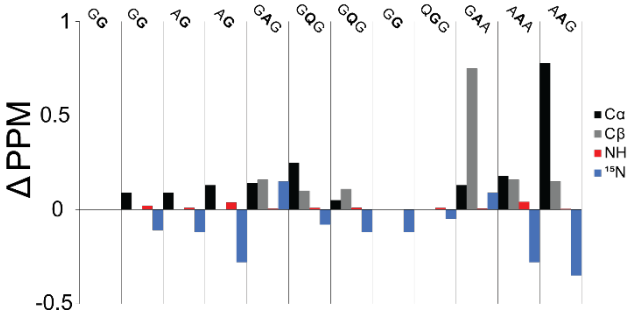

Figure 3. NMR chemical shift perturbations (Δ ppm) observed between the native intact Ma glands from *L. hesperus* spiders (~35 wt%) and diluted (~1 wt%) dope in 4M urea. The chart is arranged in the typical direction from the C- to N-terminus Gly-rich regions come before the poly(Ala) domains.

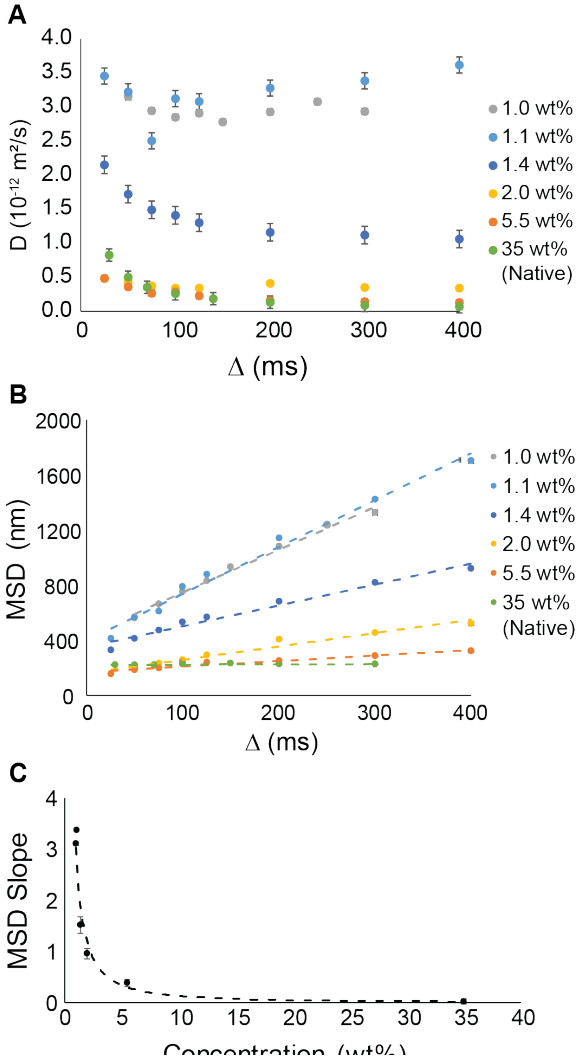

Figure 4. A, Diffusion coefficient measurements from PFG-STE NMR conducted at 800 MHz for native *L. hesperus* MaSp proteins as a function of concentration in 4M urea. B, MSD of MaSp proteins from A. C, Slopes of MSD from B plotted against concentration. Below ~5 wt%, MaSp proteins exist primarily as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

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

In order to properly attribute changes in structure and dynamics to either the dissolved MaSp protein monomers or the pre-assemblies, we first needed to determine the protein concentration at which these pre-assemblies dissolve. Due to the high protein concentration of spider silk dopes, DLS cannot be employed as the upper limit for this technique is ~1wt % protein. Instead by using diffusion NMR to monitor the disappearance of restricted diffusion as a function of concentration in 4 M urea, the critical concentration for pre-assembled protein structures can be determined. Previously, we have used diffusion NMR to determine that the MaSp proteins are restricted in the gland to a volume space of ~ 250 nm after observing a diffusion dependence on Δ and subsequently taking the MSD to approximate the size of the pre-assembled protein superstructures. Additionally, we calculated the size of a single MaSp monomer using dilutions between 0.1-1 wt%, where there is no dependence on Δ, and thus free diffusion occurs, and extrapolated to infinite dilution. Modifying this approach slightly, here we measure the diffusion of MaSp proteins at concentrations between native (35-50 wt%) and dilute (1 wt%) conditions, calculating the MSD and inspecting the MSDs for an abrupt change in slope corresponding to faster (unrestricted) diffusion. At native concentrations, MaSp proteins are restricted in the gland (**Figure 4A**) and the MSD is plateaued at ~300 nm (**Figure 4B**). This agrees well with our previously published data

showing that in the native gland environment the MaSp proteins are found in pre-assembled, micelle-like structures in the ~300 nm regime. As we begin to dilute the protein in 4 M urea, we see that the MSD increases slightly at 5.5 wt%, and abruptly between 1-2 wt% (**Figure 4B**). By plotting each slope from the calculated MSD at each concentration, we see a clear inflection between 4-5 wt% (**Figure 4C**), corresponding to the critical concentration of pre-assembled protein superstructures. For more on this work please see **ref 3**.

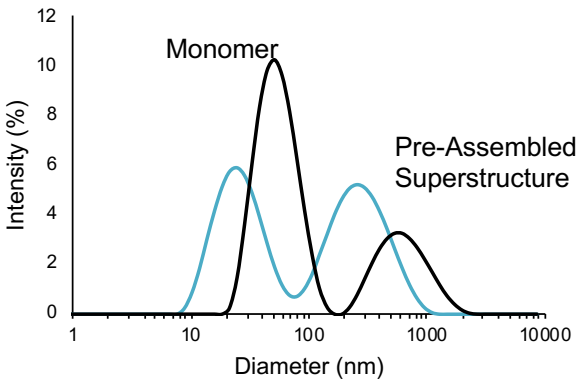

Figure 5. DLS of native *L. hesperus* spider silk dope diluted in 4 M urea (black) and recombinant MaSp spider silk protein (blue). Protein monomers and pre-assembled protein superstructures are observed. Protein concentration is ~1 wt% for both samples.

Progress in DLS and Solution NMR of Recombinant Dopes (obj. 1 & 2):

In year one we have also embarked on investigating a recombinant MaSp protein in collaboration with the Scheibel Lab at Bayreuth in Germany. This recombinant protein resembles MaSp2 and exhibits some similarities to the behavior of native spider silk dope with respect to the formation of pre-assembled silk protein superstructures which are believed to be critical for spinning fibers with mechanical properties approaching native material. DLS data is shown for *L. hesperus* native spider silk dope diluted in 4M urea and recombinant MaSp2-like spider silk protein. As can be seen from this data, protein monomers and pre-assembled superstructures in 200-300 nm length scale are observed. Not surprisingly, the recombinant protein exhibits a significantly smaller monomer and superstructure size that is presumably due to the significantly lower MW ~60 kDa compared to 300-350 kDa for the native MaSp proteins. Of significance, is that the addition of phosphate to the recombinant dope appears to promote the formation of MaSp superstructures and recombinant 'synthetic' spider silk fibers spun from dopes containing phosphate yield better mechanical properties than dopes without phosphate.

In order to better understand the molecular level interactions that govern the pre-assembled MaSp superstructures we have begun using solution NMR to investigate the recombinant MaSp protein under various biochemical conditions. We have successfully isotope enriched the recombinant MaSp protein and collected preliminary $^{1}H/^{13}C$ HSQC (**Figure 6**) and $^{1}H/^{15}N$ HSQC (**Figure 7**) NMR data for dopes with and without phosphate. The dope that contains phosphate is referred to as a biomimetic spinning dope (BSD) since it is known that phosphate is added to the spinning dope during the native spider silk spinning process while, the dope without phosphate The $^{1}H/^{13}C$ HSQC is nearly identical and superimposable for the two dopes with both samples exhibiting isotropic chemical shifts consistent with a random coil secondary

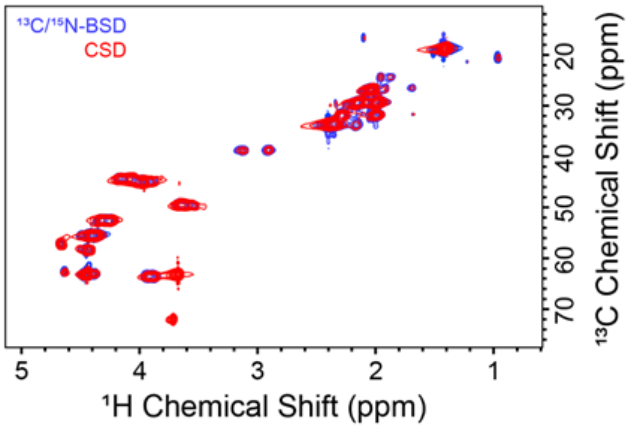

Figure 6. $^{1}H/^{13}C$ HSQC NMR spectrum of recombinant MaSp spider silk dopes. Biomimetic spinning dopes that contain phosphate (blue) and the classic spinning dope w/o phosphate (red) are shown at pH 7.

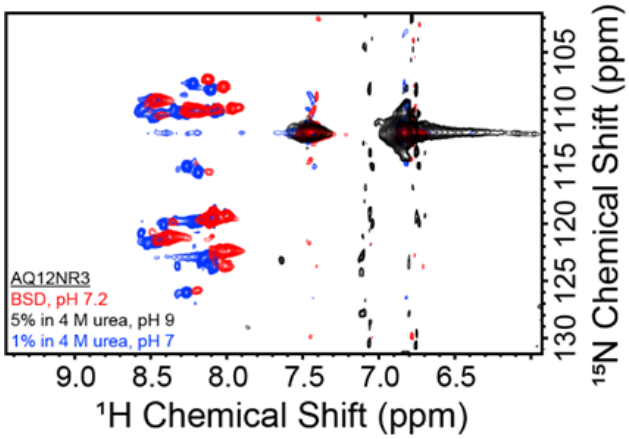

Figure 7. $^{1}H/^{13}C$ HSQC NMR spectrum of recombinant MaSp spider silk dopes. Biomimetic spinning dopes that contain phosphate (blue) and the classic spinning dope dissolved in 4M urea and two different pH 7 and 9.

structure. A random coil secondary structure is analogous to observations made for native MaSp spider silk dopes and illustrate that both dopes display little to no secondary structure and are highly flexible protein chains.

The $^{1}H/^{15}N$ HSQC NMR spectrum shown in Figure 7 shows the overlay of the BSD at pH 7.2 with recombinant spider silk dopes in 4M urea at two pH 9 and 7. The 4M urea dissolution process is used to prevent aggregation of native spider silk proteins to retain the native random coil structure and is used here to compare the BSD dope to the 4M urea dope state. There are some notable differences between the BSD and 4M urea preparations that we are in the process of analyzing that show that BSD prep differs from the random coil state. This notable but, subtle differences are being analyzed and further experiments/analysis are discussed in year 2/3, **ref 4**.

Progress in CryoEM of Native Spider Silk Dopes (obj. 2):

Cryogen Electron Microscopy (cryoEM) has become a powerful tool for resolving protein structures and assemblies of oligomers at near-atomic level. We are introducing this technique in combination with NMR to biomaterials-based problems, starting in the area of silkworm and spider silk proteins. To image the native, pre-spun silk protein complexes using cryoEM, a thin, aqueous layer of diluted silk sample must be vitrified prior to imaging. Ma glands of *L. hesperus* spiders and the silk glands of *B. mori* silkworms are being dissected and cleaned by carefully removing the outer membrane of the gland without shearing the dope. The dope is then solubilized in either 4M urea (*L. hesperus* glands) or water (*B. mori* glands) and subjected to a variety of conditions, including shear forces and changes in pH and salt concentration. Using a glass pipette to prevent additional shearing, approximately 5µl of sample are spotted onto graphene-oxide coated holey-carbon grids (Quantifoil R2/2, 400 mesh, Cu support). Developing grids for effective imaging of protein complexes is of great importance, as the substrate used to coat the grid, as well as the grid material itself, could interact with the sample and alter the conformation of the proteins. Recently, graphene supports on gold grids have been shown to improve image resolution and are being used in our lab. To explore the potential of different grids for high resolution imaging of silk IDPs, graphene-coated gold grids are being used for imaging as shown in **Figure 8**. Grids are blotted for 4s before vitrifying in liquid ethane using a Vitrobot, following previously established protocols described by our labs. This procedure will also be applied to recombinant silk proteins to identify discrepancies between the native and recombinant pre-spun Ma spidroins (see section above where we already have solution NMR data for recombinant spider silk samples). CryoEM images are being collected in the UCSD cryoEM facility utilizing the Titan Krios operated at 300keV which is equipped with Gatan Quantum energy filter, K2 Summit direct electron detector and Volta Phase Plates direct electron detection. Samples

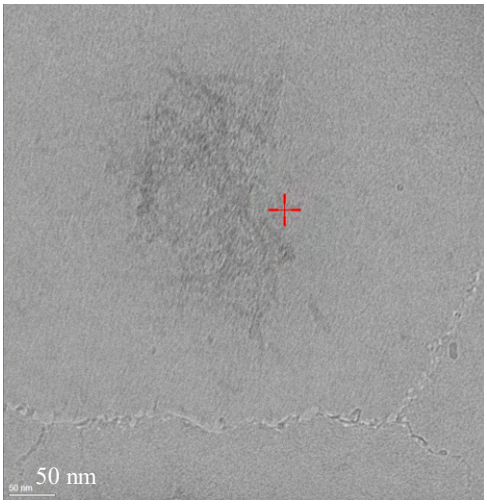

Figure 8. CryoEM image of native *L. hesperus* MaSp spider silk dope dissolved in 4M urea at a concentration of 0.5 wt%. The sample was quenched for 4s in liquid ethane with a Vitrobot.

will be imaged at a range of tilt angles for reconstruction of 3D tomograms. These tomograms will be exploited to gain insight into the mesoscale interactions occurring within protein pre-assembled structures. The data shown in Figure 7 is for L. hesperus spider silks dope diluted in 4M urea to concentration of 0.5 wt%. As can been seen in the image the silk proteins are present as larger 200-300 nm superstructures that contain 100's of protein monomers. We are in the process of testing the impact of various physiochemical conditions that are known to be important for spider silk spinning (pH, phosphate, shear, etc) to determine the impact on these protein pre-assembled structures.

Progress in IR-Microscopy of Native Spider Silk Spinning System (obj. 3):

IR-microscopy allows for cross-sections of the fixed spider or silkworm silk gland and duct to be imaged and secondary structure analyzed by deconvoluting peaks within the amide region of the spectra. Thus, the method allows for a spatial characterization of silk protein structural

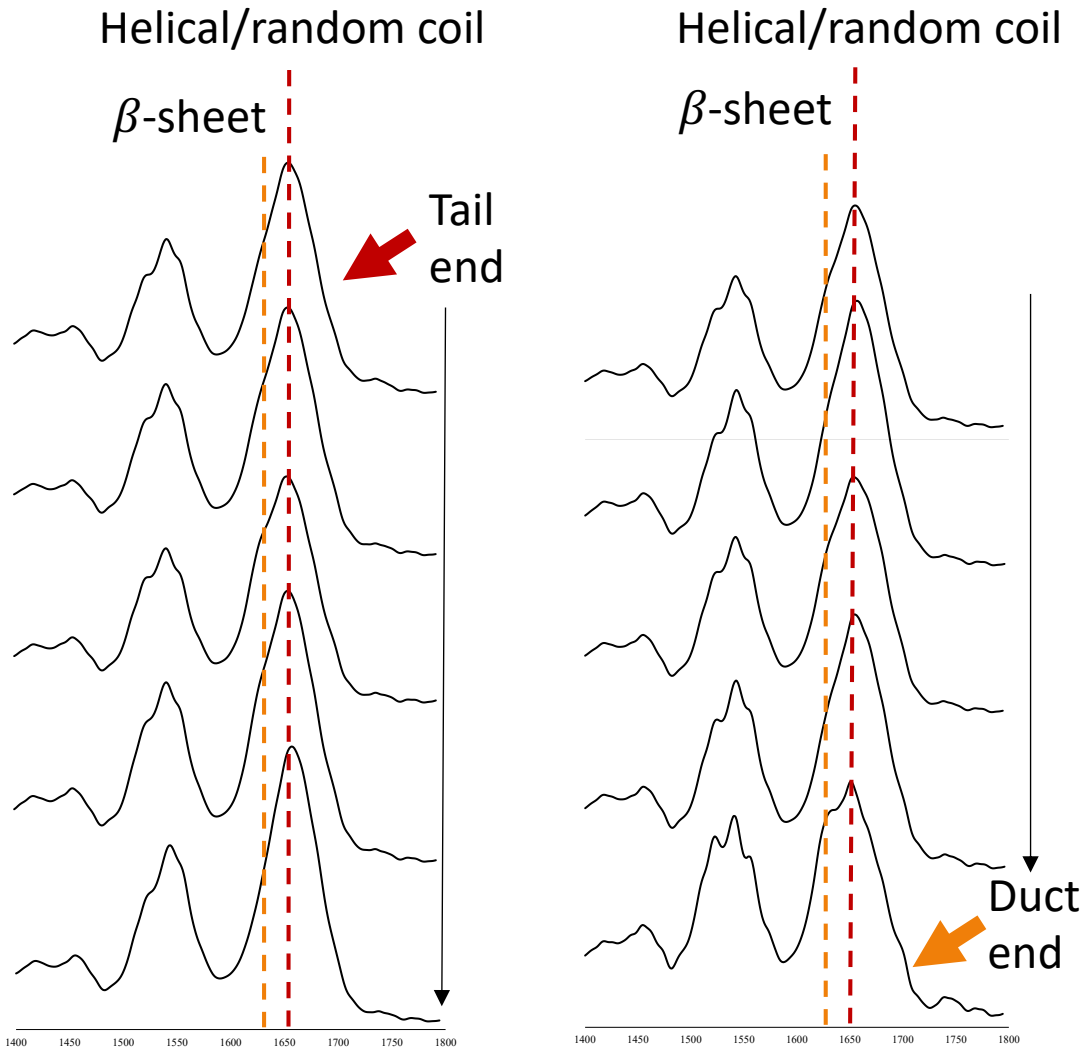

Figure 9. IR-Microscopy of *L. hesperus* spider silk spinning system. Images of the spider silk spinning system are sectioned and fixed for IR microscopy. The location of the section in the silk spinning system is shown to the left of the representative IR spectrum. The amide I band is sensitive to the silk protein secondary structure and exhibits a progressive increase in b-sheet content as the dope travels towards the end of the duct region of the silk spinning system.

evolution as the dope proceeds through the silk spinning system. Using IR-microscopy, we have mapped changes in secondary structure in the gland and duct regions of *L. hesperus* spiders where an increase in β-sheet content is observed in the final stages of the duct prior to fiber extrusion (**Figure 9**). Currently we are continuing IR-microscopy studies, focusing on tracking and quantifying the progression of silk protein secondary structure down the duct both in spiders and silkworms, eventually leading to fiber assembly. We hypothesize that the dope in silkworm glands will form micellular silk protein structures analogous to those found in spiders (**Figure 8**), based on the similarities of the glands, shared amphipathic nature of the proteins, and some initial literature investigations with light scattering and preliminary DLS data from our lab. As the protein complexes travel down the duct and the micelles elongate into fibers, we hypothesize that β-sheet character will increase as elongated silk protein micelles link up due to the acidic pH gradient along the duct and other changes in physiochemical variables that initiate silk protein folding. The hypothesized increase in β-sheet character as silk protein progresses towards the spider's spinneret is confirmed by the data presented in **Figure 9**.

Progress in SSNMR of Recombinant 'Synthetic' Spider Silk Fibers (obj. 4):

We have been characterizing the secondary structure of synthetic spider silk fibers spun from the CSD and BSD dopes. The ^{13}C cross polarization magic angle spinning (CP-MAS) SSNMR spectra is shown in **Figure 10** for the lyophilized protein and fibers spun from both the BSD and CSD dope. The ^{13}C isotropic chemical shift is dependent on the secondary structure (conformation) and these spectra can be used to determine changes in secondary structure between the lyophilized protein and fibers spun from different dope compositions or processed in various ways. Interestingly, fibers spun from both the BSD and CSD exhibit b-sheet conversion for poly(Ala) that is identical. Peak fitting routines have been used to determine that the b-sheet content in both fiber types is 77% which is very close to the 80% β-sheet fraction observed in native spider silk fibers. However, this does not explain why the BSD fibers perform significantly better in terms of mechanical properties compared to CSD fibers. This indicates that the differences in fiber mechanical properties may result from structural differences in the non-β-sheet forming domains. We are in the process of collecting additional two-dimensional (2D) homo- and heteronuclear SSNMR data to further characterize the β-sheet region and more importantly the loose helical and β-turn regions of the fiber to uncover structural differences that may help explain the differences in mechanical properties between

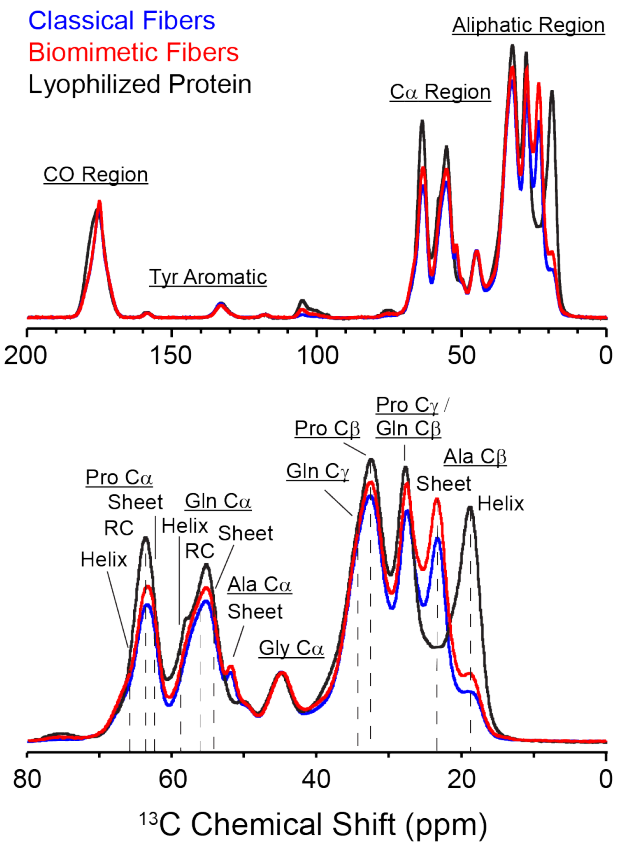

Figure 10. ^{13}C CP-MAS NMR of lyophilized recombinant MaSp protein (black) and fibers spun from the BSD (red) and CSD (blue) recombinant MaSp spider silk protein spinning dopes.

the two types of fibers. For more on this work (**see ref 4**).

4. Status of Effort (Year 2):

Our primary aim is to provide fundamental insight into the interactions and mechanisms connecting the atomic and molecular details (NMR, MD) with the mesoscale assembly process (cryo-TEM, DLS) that controls the conversion of solubilized silk proteins to fibers and establish the fundamental design rules for synthetic spider silk-based materials production. Most importantly, our research outlines an advanced multi-modal characterization blueprint for studying natural biomaterials assembly generally (**Figure 1**). We have made substantial headway towards our four main objectives in year 2. We have conducted numerous MD simulations of full length major ampullate spidroin (MaSp) spider silk proteins that are pointing to a tubular tertiary structure with certain repetitive amino acid motifs buried with others solvent exposed. We have continued 3D protein solution NMR of both native and recombinant isotopically enriched MaSp proteins expanding on our NMR assignment and structural and dynamic characterization of the proteins prior to fiber formation and have begun using computational tools to predict structural ensembles for the spider silk proteins in solution. Our NMR and computational studies have implicated the presence of a combination of β-turn and loose coil secondary structures that we believe lead to the tubular shape observed in our MD models. On our recombinant spider silk protein effort, we have completed a study where we tracked the formation of synthetic spider silk from Liquid-Liquid Phase Separated (LLPS) dopes that indicate a unique role for Tyr in mediating the intermediate LLPS step that results in very different Tyr side chain interactions when comparing the SSNMR spectra of fibers spun from classical spinning dopes (CSD, no LPPS) and biomimetic spinning dopes (BSD, w/LLPS) (**ref 4**). This helps to provide a molecular understanding of what is going on in LLPS dopes that result in improved mechanical performance when comparing CSD to BSD spun fibers.

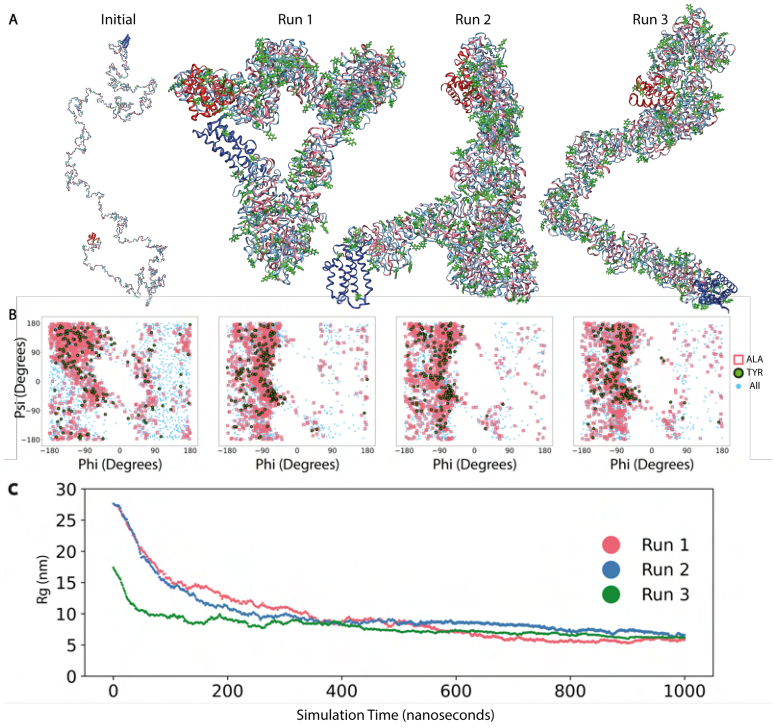

Figure 11. (A) Martini 2 MD simulations of the full repetitive core for *L. hesperus* MaSp1 spider silk protein using GROMACS. The simulations start with a random coil chain and three different completed simulations starting from three distinct random coil initial structures are shown. (B) Ramachandran plots showing the evolution of secondary structures with a preference for loose coils and β-turns in final structures. (C) The simulations were run out to 1 μs showing an equilibration in R_g that agrees with R_g from experimental SAXS data (8-

4.1 Accomplishments/New Findings (Year 2):

Progress in Atomic/Molecular Characterization of Native Spider Silk Dopes (obj. 1):

MD: In-silico simulations are a critical component of this work to help guide spectroscopy experiments and interpret the results of complex NMR and EM structural datasets. In addition to

assisting protein structure determination, molecular dynamics (MD) is employed to determine the 3D shape of the spider silk proteins. We are beginning to understand the tertiary structure of the two primary proteins that comprise spider dragline silk, MaSp1 and MaSp2, where we have already produced significant conclusive results in year 2. The complete primary amino acid sequences of *L. hesperus* MaSp1 and MaSp2 spider silk proteins are known, and MD structures have been completed (**Figure 11)** and the manuscript is now submitted for publication (**ref 6**). We have found that the structures are primarily disordered random coils in agreement with solution NMR however, they exhibit a loose coiling structure that contains some β-turns with poly(Ala) domain buried and the Tyr, Gln and Arg (in MaSp1) and Tyr, Gln and Pro (in MaSp2) residues primarily solvent exposed (data not shown). The implication that the poly(Ala) is primarily buried is supported by published work from year 1 where we showed with NMR relaxation measurements specifically, heteronuclear NOE that demonstrated the poly(Ala) had the most restricted backbone dynamics compared to any of the other repetitive motifs in the silk protein (**ref 3**). A more restricted dynamic environment is expected for regions that are less solvent exposed and buried in the protein core as is observed in the MD structures. Another noteworthy observation from the MD structures is the anisotropic tubular shape of both proteins which agrees with experimental SAXS data both in terms of the radius of gyration (R_g) (7-9 nm) of the simulated structures, and the overall anisotropic shape. The tubular shape of the silk protein molecules could be a critical nanoscopic pre-ordering that aids fibrillization and fiber formation.

NMR: We have been developing an X-detect backbone assignment strategy in combination with ^{1}H-detect data that are optimized for spider silk protein motifs following an approach similar to those applied to other IDP's. 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 solution NMR methods. Despite a lower gyromagnetic ratio, ^{15}N direct-detect methods compliment the ^{13}C direct-detect suite of experiments and have several unique advantages. 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 and thus, should provide similar advantages to the study of spidroin proteins and were further developed in year 2. These two experiments form the basis of a sequential assignment strategy similar to the HNCACB/CBCACONH used both in more traditional protein NMR and in the three and

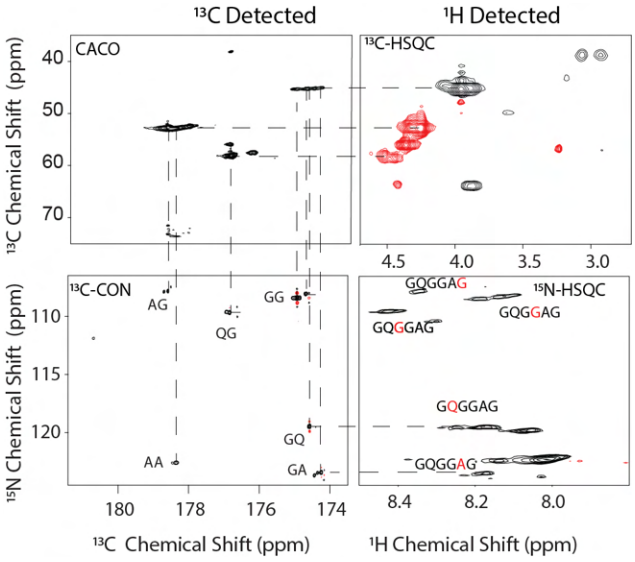

Figure 12. Sequential backbone assignment of residue pairs from the correlated X-detect experiments CACO, CON, and traditional ^{1}H-detect ^{13}C and ^{15}N HSQC. The sample was ^{13}C/^{15}N-Ala labeled Ma spider silk protein from *L. hesperus* spiders. The data was collected in the UCSD Biomolecular NMR facility at 800 MHz with TXO indirect NMR cryoprobe.

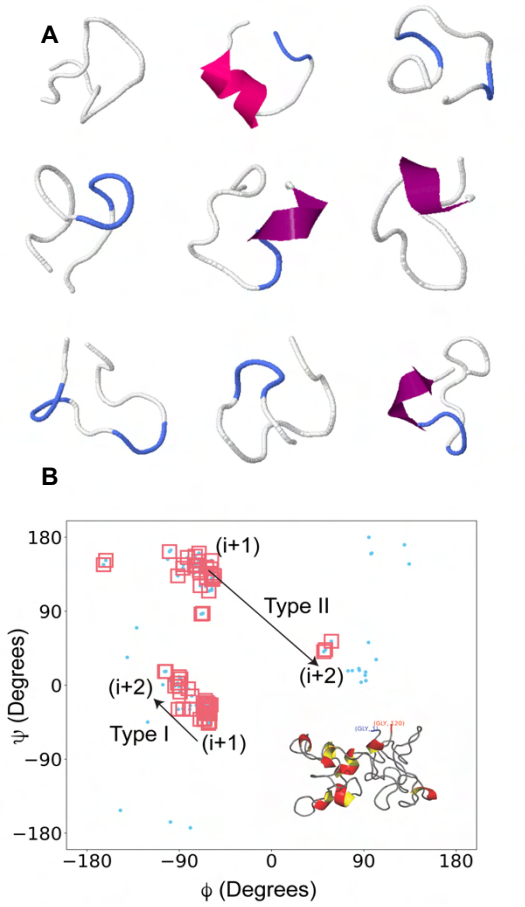

Figure 13. (A) Nine lowest energy structures generated by CS-Rosetta using NMR chemical shift data from the fully assigned GQGGAGAAAAAAAAG repeating unit of MaSp1. Generally, the structures lack defined secondary structure but, rather exhibit a loose β-turn structure. (B) Ramachandran plot of a 60-residue structure generated in CS-Rosetta with the 3D structure shown inset. The Ramachandran plot is consistent with a combination of Type I and Type II β-turn conformational ensembles. Ala residues are boxed in red.

six residue MaSp1 that we recently published (**ref 3**). 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. Taken together, these direct-detect experiments form an elegant method for probing unstructured, dynamic protein structures that elude nearly all other methods of direct investigation. When coupled with different isotope labeling schemes, these approaches provide an unprecedented level of detail for the inner workings of MaSp1 and MaSp2 silk proteins in a variety of different biochemical environments and the complete chemical shift data obtained can be used together with NMR computational tools TALOS-N and CS-Rosetta to determine silk protein structural ensembles in solution.

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 this backbone relaxation (T_1, T_2, NOE) to be on par with those observed for other IDPs (**ref 3**). 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 optimized for X-detection. 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. Coupled with selective labeling strategies, this increase in signal intensity, resolution, and site-specificity is allowing us to greatly expand our ability to structurally characterize silk proteins with solution NMR. This specialized equipment is made available to our lab by the UCSD Bimolecular NMR Center and provides some of the highest solution NMR sensitivity and resolution available in the world, which is a requirement for these cutting-edge X-detect IDP experiments.

Through a combination of ^{1}H-detect conducted at SDSU and X-detect NMR approaches conducted at UCSD we have been able to fully assign and extract the ^{1}H/^{13}C/^{15}N chemical shift data for all sites in a 15-residue repeating unit of the MaSp1 protein (**Figure 12**). The 15-residue repeating unit is GQGGAGAAAAAAAAG which, appears 39 times in the MaSp1 sequence making it highly representative of the silk proteins including the GGX and poly(Ala) motifs.

Assignments for two triple residue repeats in the GGX region were linked together and then placed alongside assignments from the poly(A) region of the protein. Notably, the X-detect experiments provide outstanding resolution in the ^{13}C carbonyl region of the spectrum as evidenced by a broad stretch of Gly resonances that informs on conformational variability for the Gly-rich motifs. These X-detect experiments are extremely promising for expanding our ability to structurally characterize both MaSp proteins and we have begun using this data to determine the conformational ensembles of this repeating unit with NMR computational tools. We have generated structural ensembles in year 2 and are actively working on a manuscript that combines the MD models (**Figure 2, 11**) with these experimental NMR results. From the complete chemical shift data set obtained for GQGGAGAAAAAAAAG, a combination of computational techniques is used to determine structural ensembles of this heavily represented residue repeat in MaSp1. Two NMR computational tools are being used TALOS-N and chemical shift (CS)-Rosetta. TALOS-N uses chemical shifts and a combination of neural network analysis and database comparison to find restraints on phi/psi angles of individual residues. The output angles are also scored based on how well the prediction is likely to be and how large the expected deviations are. These parameters can then be fed as restraint files into CS-ROSETTA, a set of tools coupled to the ROSETTA *ab initio* structure prediction software. Here the predicted angles and secondary structure predictions are used to select protein fragments from the BMRB which most closely resemble the input sequence. From these fragments, Monte Carlo structure generation and relaxation using ROSETTA results in a landscape of predicted possible structures with various energies. If the resulting ten lowest energy structures are extremely similar, the results are said to converge, and final structures are probably representative of the true structural ensemble. The converged results of the MaSp1 fragment above gave rise to the top lowest energy structures shown in **Figure 13.** These results are extremely promising and show that this motif in MaSp1 lacks most traditional secondary structure however, a consistent loose β-turn structure is observed. We have further extended this to larger structures generated in CS-Rosetta where we linked four of the 15-residue repeats to generate a 60-residue structure. As can be seen from the Ramachandran plot of this structure, it is best described as a combination of loose type I and II β-turns (**Figure 13B**). The results are extremely promising and represent the first 3D solution structures of native spider silk protein motifs ever conducted and continuation of this work will enable building real structural models for silk proteins and is ongoing in year 3.

<u>Progress in Atomic/Molecular Characterization of Recombinant Spider Silk Dopes (obj. 1):</u>

In addition to isotopically labeling native silk proteins for NMR studies, we have pursued alternative strategies for generating silk proteins to investigate. The Scheibel Group expresses recombinant spider silk proteins on large scales for fiber spinning and have produced synthetic spider silks with high toughness and some of the best mechanical properties for a synthetic spider silk to date. The Scheibel Group has a firm grasp on the interplay between amino acid sequence, secondary and higher-order structure on the resulting silk fiber mechanical properties. Using native sequences as their template, they have assembled MaSp2 synthetic protein analogs with this sequence: (AQ_{12}: GPYGPGASAAAAAAGGYGPGSGQQGPGQQGPGQQGPGQQGPGQQ) with different repeat lengths, with and without N-termini, and with and without C-termini to study their effet on overall fiber mechanical properties. Various ion exchange processes and an acidic pH gradient in the duct system are believed to facilitate silk protein organization and assembly into fibers. Recently, the Scheibel group published strong evidence that their biomimetically spun, $N1L(AQ)_{12}NR3$ fibers created from a dope with the correct balance of salt ions and pH have a similar toughness to native fibers, strengthening the idea that fine tuning the biochemical

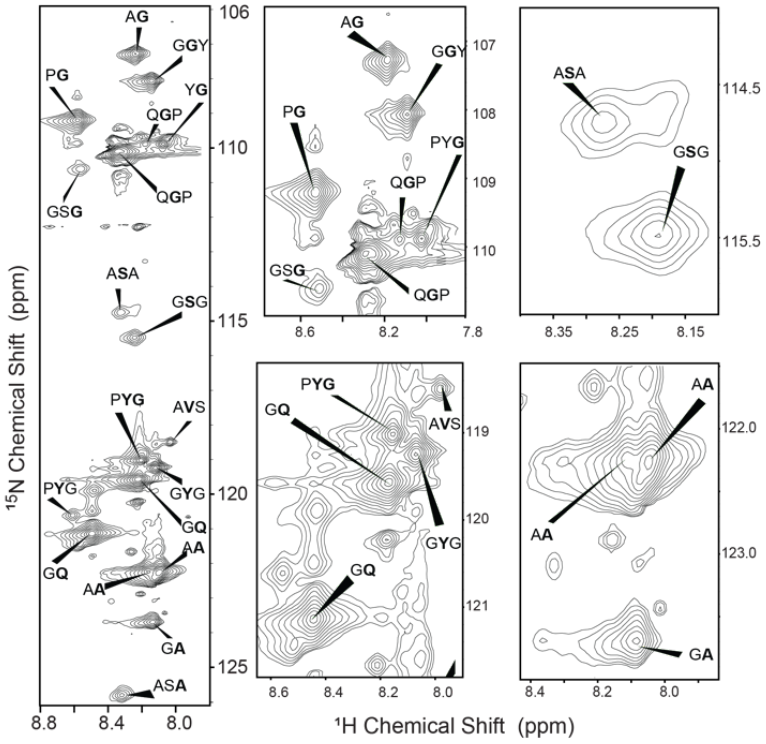

Figure 14. ^{1}H-^{15}N HSQC NMR spectrum of fully labeled $^{13}C/^{15}N$-$(AQ)_{12}$NR3 BSD collected at 850 MHz. Bolded amino acid indicates the resonance identity with di- and tri peptide repeats based on 3D experiments as discussed in our recently published manuscript (ref. 4) .

conditions of the spinning dope is critical to producing fibers that more closely resemble native ones in terms of mechanical properties. However, it should be noted that although the fibers were tough because of their extensibility, they exhibited considerably lower strength compared to native dragline silk, illustrating that there is still work to be done in order to produce synthetic spider silks with properties that match native ones.

Building on this work, the goal is to understand the role of salt and pH in the assembly of these recombinant spidroin proteins on a molecular level with solution NMR. We understand the native conditions during assembly, including the nanostructure of the pre-assembled silk proteins that are important for the ultimate mechanical properties of the synthetic fibers, with the hypothesis that proper salt and pH balance helps to organize the proteins into 200-300 nm hierarchical pre-assemblies. In year 2, we have further tuned these recombinant synthetic silk protein dopes and characterize their structure and dynamics with some of the solution NMR methods described above to compare to the native system. *Isotopic enrichment of recombinant spider silk proteins is ongoing through the Schieble lab collaboration and we have already published our first manuscript* (**ref 4**). In the manuscript we discuss our initial backbone assignment and interpretation of the silk protein structure (**Figure 14**). We are continuing this work in year 3 using the above-described X-detect NMR approaches combined with computational NMR modeling tools TALOS-N and CS-Rosetta to determine the structural ensembles of the recombinant system and compare to the native ones we have already begun determining (**Figure 13**). This collaboration has been ongoing for over three years, with already secured travel funding between labs through the Bavaria California Technology Center (BaCaTeC) Internationalization of the High-Tech-Initiative. The Scheibel and Holland Lab have already hosted multiple student and PI trips from each research group. The Scheibel group will continue to provide material to Holland Lab in order to investigate the structure and organization of the silk dope at the molecular and mesoscales including the role of phosphate and LLPS (below).

We are interested in comparing the behavior of recombinant spider silk proteins in the presence of phosphate that initiates LLPS leading to fibers with improved mechanical performance. We embarked on a study of LLPS in recombinant protein dopes from the Scheibel group with NMR spectroscopy and recently published these results (**ref 4**). In order to examine more closely the structural differences of the protein in two distinct dope preps, ^{13}C direct-detect spectra were collected and compared for dopes that have undergone LLPS in the presence of

phosphate and dopes prepared in tris buffer. The former is coined BSD and the latter CSD. ^{13}C direct detect experiments have been shown to provide valuable structural information with improved resolution for IDP's, as the ^{13}C dimension has a considerably larger chemical shift range compared to either ^{1}H or ^{15}N providing information regarding both the backbone and sidechains in a single experiment. Assignment in the ^{13}C NMR spectrum was based on 3D and ^{1}H-^{13}C HSQC NMR experiments (data not shown). A number of subtle but, interesting differences were observed between the two dopes in the ^{13}C solution NMR spectra (**Figure 15**). The first obvious difference was resonance line sharpening observed for the BSD when compared to the CSD. Line sharpening was most notable for the Cβ Ala resonances G**A**, S**A** and A**A**, Pro Cα resonances G**P**Y and G**P**G and carbonyl resonances. IDP's like spider silk proteins display structural heterogeneity due to an ensemble of conformational environments. The observed line sharpening indicates that addition of phosphate in the BSD dope induces a degree of structural ordering. This pre-ordering of the silk protein prior to fiber spinning in BSD suggests an important step that could lead to spun fibers with improved mechanical properties and we aim to further expound upon this with NMR relaxation measurements in year 3 to further elucidate silk protein dynamics in the pre-ordered state.

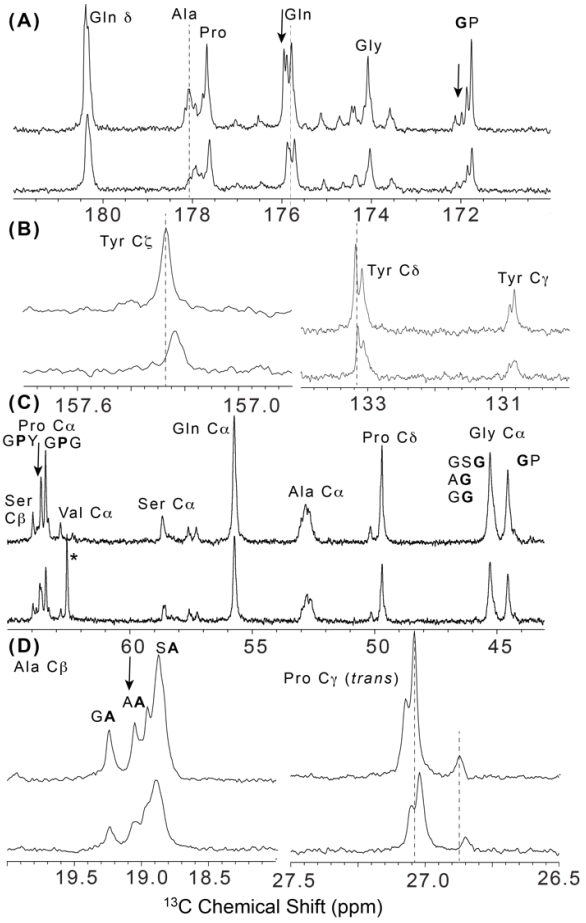

Figure 15. ^{13}C solution NMR spectrum of recombinant $(AQ)_{12}NR3$ CSD (bottom) and BSD (top) collected at 600 MHz. The (A) carbonyl, (B) aromatic, (C) backbone and (D) aliphatic region of the spectrum are displayed. Resonances that exhibit line sharpening are indicated with arrows. Tris buffer signal is assigned an asterisk (*).

Other notable perturbations observed in the ^{13}C direct solution NMR spectrum of the two dopes included slight chemical shifts to higher ppm for the Pro Cγ resonances, the Tyr aromatic ring environment and the Tyr C-OH moiety with a pronounced shift for the Ala carbonyl resonance. This indicates that these environments are the most impacted by the BSD environment. One thing that stands out is the change of the Tyr aromatic C-OH (Cζ) which also exhibits a line width that is ~2x sharper in the BSD dope. The importance of the Tyr environment in the final spun fibers (discussed below) and differences observed in Tyr ring interactions could be traced back to some unique differences in the two dopes where the Tyr environment was significantly more ordered for the BSD dope likely due to hydrogen-bonding differences between the two dopes resulting in varying Tyr ring packing. The importance of Tyr is very interesting in light of our MD results (**Figure 2, 11**) on native MaSp that points to Tyr located almost exclusively on the protein surface. The pronounced shift for Ala carbonyl resonance also indicated an increased ordering effect for this environment showing that there is a notable effect on the poly(Ala) in the BSD dope.

Progress in Mesoscale Characterization of Native Spider Silk Dopes (obj. 2)

Cryogen Electron Microscopy (cryoEM) has become a powerful tool for resolving protein structures and assemblies of oligomers at near-atomic level. We are introducing this technique in combination with NMR to biomaterials-based problems, starting in the area of silkworm and spider silk proteins. We have already obtained high quality cryoEM of native spider silk protein superstructures and have been working on a high through put negative staining (NS)-TEM method for imaging silk protein assemblies that match cryoEM images. Briefly, Ma glands of *L. hesperus* spiders and the silk glands of *B. mori* silkworms are being dissected and cleaned by carefully removing the outer membrane of the gland without shearing the dope. The dope is then solubilized in either 4M urea (*L. hesperus* glands) or water (*B. mori* glands) and subjected to a variety of conditions, including shear forces and changes in pH and salt concentration. Using a glass pipette to prevent additional shearing, approximately 5µl of sample is spotted onto TEM grids and stained with a staining agent. We have screened numerous staining agents and protocols for this and have found a stain and protocol that yields high quality TEM images of the silk protein assemblies that match our cryoEM images previously published that were obtained in a more native, hydrated environment. The stain of choice is ammonium molybdate at pH 7 (see **Figure 16**). The images are of excellent quality and very close to the silk protein nanoparticle (micelle-like) assemblies previously published. One of the main goals is to look at some of the biochemical conditions that trigger spider silk fibrillization and image starting, intermediate and final structures to image the progression of the fiber formation process. A crticial trigger for fibrillization in both spiders and silkworms is an acidic pH gradiet in the gland duct system en route to extrusion. Through careful selection of an acidic staining agent (uranyl acetate, pH 3) and wicking of the fluid to mimic shear forces and fluid flow in the gland we have triggered the early stages of fibrillization directly on the TEM grid. These results are shown in Figure X. These results are particualrly exciting as one can directly observe some of the smaller micelle-like structures assembling into fibers with very good contrast. These results are very exciting and we are playing with different staining agents and shearing protocols to mimic fibrillizaiton in vitro. We are continuing these efforts in year 3 and working on a protocol manuscript on NS-TEM.

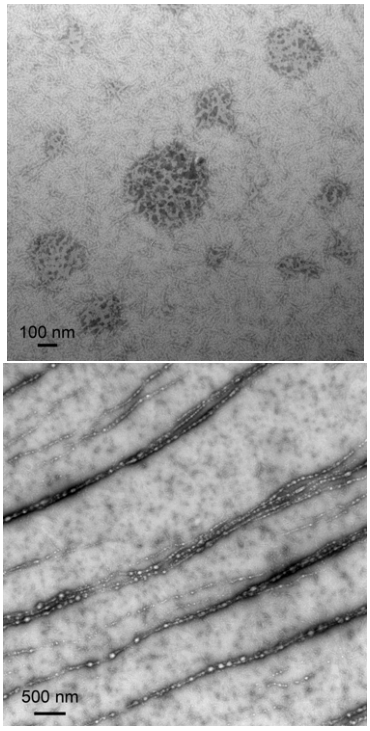

Figure 16. (Top) NS-TEM of native silk protein assemblies (protein micelles). The staining agent is ammonium molybdate pH 7. The morphology and dimensionality of the silk protein assemblies is very close to those observed previously by our lab in a hydrated environment with cryoEM. (Bottom) NS-TEM with uranyl acetate pH 3 staining agent together with wicking to mimic shear flow. Fibrillization is observed on the TEM grid with spherical silk nanoparticles assembling into nanofibrils.

<u>Progress in SSNMR Structural Characterization of Native and Recombinant 'Synthetic' Spider Silk Fibers (obj. 4)</u>

The ultimate aim of the grant is to utilize the information provided by the above experimental and modeling methods conducted for spider silk dopes to mimic the process of silk formation in the lab at all length scales and produce biomaterials that have hierarchical structures and bulk properties (mechanical) that match native fibers. In order to develop a better molecular understanding of native and recombinant spider silk fiber structure, we have been using SSNMR

to investigate sidechain interactions in isotopically enriched spider silk fibers, focusing on residues with functional sidechains including Tyr, Gln, Arg, Ser and Pro to detect the presence of sidechain interactions not only in solution and during intermediate dope transitions (above), but also in the final spun fiber. We made substantial progress in this direction in year 2 where we have $^{13}C/^{15}N$-isotopically enriched black widow Ma silk at Ala, Tyr and Arg (which labels Pro) and utilized the chemical shift selection pulse sequence developed by our lab to simplify ^{13}C MAS SSNMR spectra and conduct chemical shift analysis that provides information on secondary structure and sidechain interactions through selective excitation dipolar assisted rotational resonance (DARR) build-up curves (**Figure 17**). In this example we were able to determine Pro-Tyr inter-residue contacts and determine distances (**see Table insets**). The results clearly illustrate that the Tyr and Pro residues interact at a distance ~ 4 Å. Through atomistic simulation on the GPGXX motif where X is Tyr, we were able to determine potential interactions that included Pro-Tyr (**Figure 17D**) as well as Tyr-Tyr through π-π stacking (**Figure 17E**). We are combining these SSNMR results in combination with MD simulation to propose structural models for sidechain interactions in spider silk fibers. We are actively working on the manuscript this summer.

We have already begun characterizing and comparing the secondary structure of synthetic spider silk fibers spun from CSD and BSD dopes (*w/Scheibel*) and recently published the results (**ref 4**). We found that although the mechanical properties are far superior for BSD fibers in terms of toughness, the nanocrystalline β-sheet conversion is identical with both fiber types exhibiting a 77% β-sheet content. Because of the differences observed for Tyr in BSD dopes (**Figure 15)**, we began to explore the Tyr environment in the spun fibers to look for differences. This prompted us to implement SSNMR 2D heteronuclear correlation experiments (^{1}H-^{13}C HETCOR) to correlate ^{1}H and ^{13}C nuclei close in space. Because of the strong ^{1}H dipole-coupling network in proteins, high MAS speeds are required to average out these effects in combination with LG ^{1}H homonuclear decoupling. From the 2D HETCOR

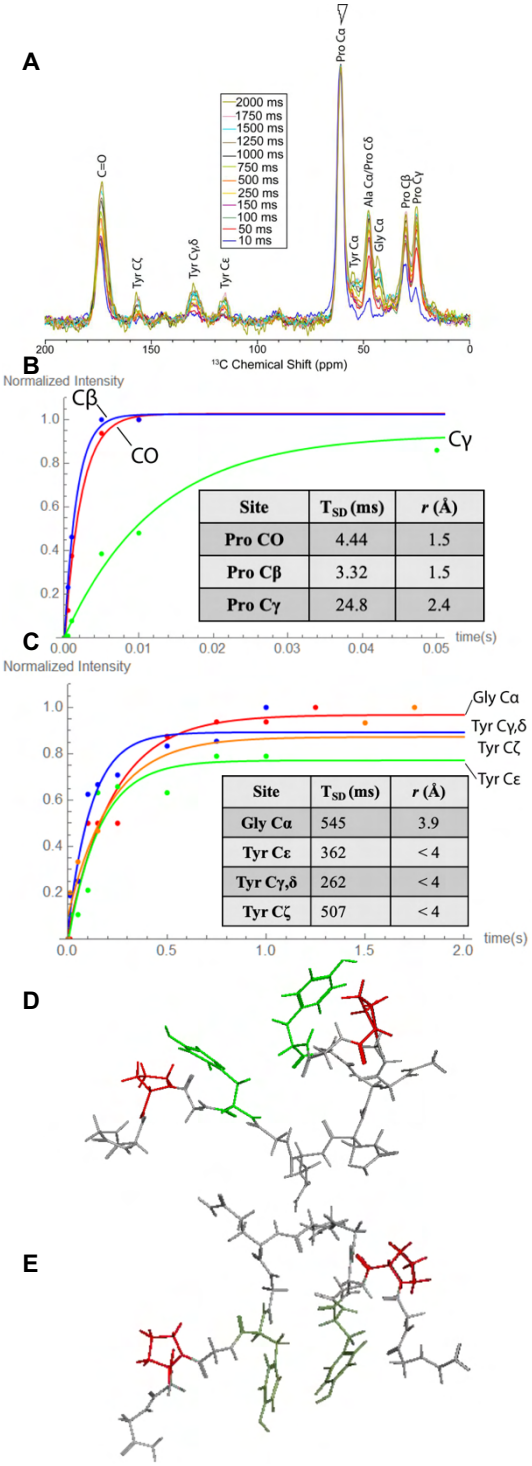

Figure 17. (A) ^{13}C selective DARR as a function of mixing time. The Pro C_α was selected and build-up curves were monitored between residue sites to determine distances. (B) Short-range intra-residue and (C) long range inter-residue build-up curves are shown. Atomistic MD models of the GPGXX dragline spider silk motif showing (D) Pro-Tyr and (E) Tyr-Tyr sidechain interactions. The sample was $^{13}C/^{15}N$-Tyr and $^{13}C/^{15}N$-Pro enriched black widow (*L. hesperus*) Ma silk.

spectrum, we saw two major differences between the BSD and CSD fibers (**Figure 18**). First, the region between approximately 50-65 ppm corresponding to the Cα backbone of the protein appears shifted downfield (to high ppm) in the ^{1}H dimension for BSD (red) compared to CSD fibers (blue). The same was true for the Tyr ring region at 131 ppm ^{13}C. In fact, the shift downfield for the protons on the Tyr ring were more dramatic than those seen in the Cα backbone region. This suggested that the Tyr ring as well as the Cα resonances in BSD are located in a more polar environment, shifting their ^{1}H resonances downfield. The Tyr ring is shifted nearly 1 ppm higher in the BSD compared to CSD fibers suggesting differences in Tyr ring packing between the two fibers. One possibility is that there are differences in Tyr π - π stacking or other Tyr π related interactions (excluding the Pro-Tyr interaction which are similar in both systems) between the two fibers as aromatic ring interactions are known to greatly perturb ^{1}H chemical shifts resulting in aromatic chemical shift shielding effects for systems exhibiting strong π stacking interactions. This Tyr packing difference could be traced back to the dope starting material where differences in the Tyr environment were observed (**Figure 15**) that might help drive LLPS resulting in Tyr structural perturbations observed in the final spun fibers when comparing CSD and BSD systems.

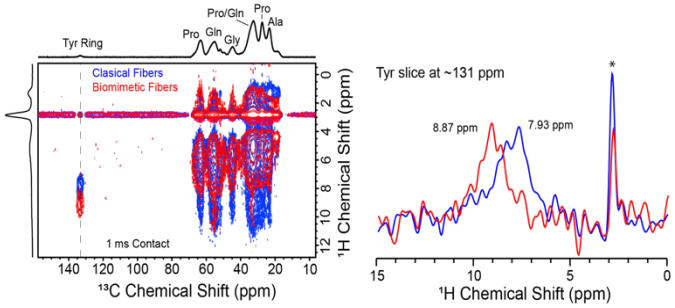

Figure 18. 2D ^{1}H-^{13}C HETCOR SSNMR spectrum collected with FSLG homonuclear proton decoupling for recombinant CSD (blue) and BSD (red) fibers. Slices were extracted at the Tyr sidechain resonance (right) to illustrate the differences between the ^{1}H chemical shift of the Tyr aromatic ring for the two fibers. This downfield (high ppm) ^{1}H chemical shift in the BSD fibers is also observed in the Cα region, but not in the aliphatic region. The DC offset glitch due to FSLG homonuclear decoupling (*).

We are actively continuing characterization of the secondary and sidechain structure of the various native and synthetic spider silk fibers with SSNMR in year 3. We have developed and applied a number of SSNMR techniques for characterizing silk structure for many years and this broad range of SSNMR experiments are applied routinely in the Holland Lab and will continue to characterize the conformational structure of various native and synthetic spider silk fibers. Information gained from these studies are being correlated back to the molecular, meso and micron scale structure of the dope starting material which is the primary focus of the proposed work. We are particularly interested in the role of sidechain interactions and will combine SSNMR data with MD simulation to elucidate sidechain contacts in the fibers. While most of our previous studies have focused on secondary structure, we aim to focus on sidechain interactions where essentially no structural information is available. Our initial studies on native (**Figure 17**) and synthetic (**Figure 18**) spider silk are beginning to point to sidechain interactions as critical for silk fiber performance and improved mechanical properties.

5. Status of Effort (Year 3):

Our primary aim is to provide fundamental insight into the interactions and mechanisms connecting the atomic and molecular details (NMR, MD) with the mesoscale assembly process (cryo-TEM, DLS) that controls the conversion of solubilized silk proteins to fibers and establish the fundamental design rules for synthetic spider silk-based materials production. Most importantly, our research outlines an advanced multi-modal characterization blueprint for studying natural biomaterials assembly generally (**Figure 1**). We have made substantial headway towards our four main objectives in year 3. We have completed numerous MD simulations of full length

MaSp spider silk proteins that are pointing to an anisotropic tubular tertiary structure with certain repetitive amino acid motifs buried with others solvent exposed. We have continued 3D protein solution NMR of both native and recombinant isotopically enriched MaSp proteins expanding on our NMR assignment and structural and dynamic characterization of the proteins prior to fiber formation. Further, we have completed one study where we use computational tools including CS-ROSETTA to predict structural ensembles for the spider silk proteins in solution. Our NMR and computational studies have implicated the presence of a combination of β-turns, bends and loose coil secondary structures that we believe lead to the tubular shape observed in our MD models. This work is under revision in *PNAS* (**ref 6**) and we are hoping for a positive decision in the coming months. The latter work is particularly impactful as it is the first structural model with atomistic detail for spider silk proteins in solution (prior to fiber formation) that explains the NMR structural and dynamic data and published small angle X-ray scattering (SAXS) data. We have also completed a study on sidechain interactions in spider silk fibers with SSNMR and MD simulations that was just accepted for publication in the ACS journal *Biomacromolecules* (**ref 5).** In this work we were able to illustrate sidechain packing interactions for Tyr and Pro rings and our MD simulations, while confirming these results, also point to other sidechain packing interactions to pursue in future work. This work is particularly important because although the secondary structure (β-sheet, β-turn, 3_1-helix) has been well established, sidechain packing interactions have not been considered. It is also imperative to note that solution studies of the silk protein are beginning to point to the importance of sidechain interactions in the dope phase and they also play a big role in liquid-liquid phase separation (LLPS) intermediate steps in silk protein assembly into fibers. Thus, continuing to understand how sidechains interact in the solution phase, under LLPS conditions and in the finalized fiber will be critical information for a full understanding of silk self-assembly process.

5.1 Accomplishments/New Findings (Year 3):

Progress in Atomic/Molecular Characterization of Native Spider Silk Dopes (obj. 1):

In year 3 we completed a study where we combined coarse grain (CG)- and atomistic MD simulations together with 3D ^{1}H and X-detect solution NMR to determine a structural model for MaSp proteins in the dope phase. The CG-MD simulations conducted with Martini2 were reported in **Figure 11** and showed that the MaSp proteins folded in a disordered tubular structure with selective partitioning of certain amino acid motifs. Regardless of the starting configuration, a consistent tubular tertiary structure emerges in all simulations, with selective partitioning of specific amino acid motifs; the poly(Ala) is primarily buried and specific polar amino acids (Gln, Tyr, Arg) are located almost exclusively on the protein surface. We conducted an extensive analysis of these structures that we report here. Having observed a consistent tubular shape between different simulations, deeper analysis (**Table 1**) reveals a hierarchical structure characterized by three different lengths scales for both MaSp1 and MaSp2. The longest scale is *averaged contour length* (ACL) which relies on the determination of a central tubule axis that matches our intuitive notion of the tubule shape. Such a curvilinear axial trajectory can be determined by taking a local center of mass of 150 residue sequences every 10 residues along the primary protein sequence. Averages over fewer than 150 residues did not remove sufficient fine detail about the protein backbone, whereas averages over 400 residues lost information about the tubule shape. These considerations suggest an appropriate size for an intermediate coarse grain unit between 150 and 400 residues, corresponding to 5 to 13 repeats of the amphiphilic sequence patterns.

The MaSp1 ACL is around 50 nm yielding a size parameter for the tubule structure which matches our intuition of the length of the tubule structures and corresponds to the length scales for monomeric units suggested elsewhere for closely related silkworm silk fibroin. The full length of the protein backbone (~1160 nm long for MaSp1, ~1400 nm for MaSp2) is considerable and is packed into this 50 nm tubule. The second length scale is the *Max Euclidean* (MaxE) length between any two atoms within the structures. For our MaSp1 structures, Max E is around 16 nm. When ACL and MaxE are compared, we perceive the amount of higher order folding of the 50 nm tubule as it begins to supercoil into a space 16 nm across. The overall gradient of the R_g plateau suggests longer time scale simulations would find more compact folds of the tubule. Such higher order folding is sure to be impacted by neighboring proteins in the silk dope that are not present in the simulations here. The third and final characteristic length scale is the tube radius, R_t, which is between 1.6 and 2 nm for the different conformations analyzed. We define R_t as the 90^{th} percentile of the distribution of radial distances between each residue and the nearest point on the axis at the center of the tubule. The measured diameter for the MaSp1 and MaSp2 tubules is therefore between 3.2 and 4 nm. Intriguingly, this value is very close to the diameters of nanofibrils found in silkworm and Ma spider silk fibers suggesting these nanofibrils could be representative of individual proteins.

In view of the order of magnitude span of hierarchical lengths that characterize the tubule (ACL, MaxE and R_t), it is clear that conventional experimental measurements of protein size such as R_g (radius of gyration) and R_h (hydrodynamic radius) mask significant details about the shape and nature of the spidroin molecule. However, R_g and R_h do enable meaningful comparison of our models with the literature. Taking the ratio of R_g and MaxE yields a value of 0.3 which is generally taken as a good figure of merit for the structural anisotropy. The anisotropy of our structures matches the anisotropy of silk dope as measured using SAXS where a value of 0.3 was also found. That said, the anisotropy defined as R_g/MaxE is relatively consistent across all structures we simulated (~0.3) and is therefore a poor discriminator for understanding differences in our structures. We found that R_g/R_t and R_g/R_h (the shape factor) were better order parameters for capturing anisotropy differences between our simulated structures. For a globular (spherical) protein, R_g/R_h is predicted to be ~0.7 while a coil shape has a value of ~1.0, and ~1.7 is an extended rod. The values determined for the simulated structures indicate that the coil is a good description of MaSp1 and MaSp2 which have shape factors in the region 0.9-1.1.

As a result of this simple analysis, we expect that models with 400 residues or more should exhibit tube-like properties, since they are sufficiently long to be comprised of more than one 150-200 residue sub-units. On their own, such 150-200 residue structures tend to be compact globular structures. Such an observation is borne out in the mutational simulations and size assay discussed below, where shape factors range from spherical to coil depending on the simulation. The R_g/R_t ratio also reveals insightful trends.

Table 1. Geometric parameters of the tubule MaSp structures. Averaged Contour Length (ACL) is the contour length of the tubular space curve found by averaging the center of mass over successive 150 residue blocks. Max Euclid (MaxE) is the longest Euclidean distance between any two atoms in the structure. R_g is the radius of gyration. R_h is translational hydrodynamic radius computed using HullRad. R_t is the tube radius that contains 90% of the residues averaged over seven 400 residue sets per MaSp1 molecule and nine 400 residue sets for MaSp2. R_g/R_h is the shape factor (sphere: 0.77, coil: 0.9 and rod 1.73). R_g/R_t and R_g/MaxE are anisotropy ratios between R_g and the tube radius and MaxE.

Name	ACL (nm)	MaxE (nm)	Rg (nm)	Rh (nm)	Rt (nm)	Rg/Rh	Rg/Rt	Rg/ MaxE
MaSp1								
aa_2	53.0	15.8	5.4	5.9	1.7	0.91	3.13	0.34
aa_5	50.8	20.7	6.1	5.9	1.8	1.03	3.41	0.30
aa_10	52.4	29.3	8.5	7.1	1.9	1.19	4.59	0.29
aa_20	52.3	20.3	6.0	6.3	1.8	0.97	3.39	0.30
aa_30	52.9	23.6	7.3	7.1	1.6	1.03	4.52	0.31
b_1	51.8	19.2	5.9	6.3	1.7	0.95	3.5	0.31
b_2	51.1	21.4	6.6	6.5	1.7	1.02	3.86	0.31
b_3	50.7	18.5	5.5	5.7	1.8	0.97	3.11	0.30
MaSp2								
aa_1	67.1	23.8	7.6	6.7	1.9	1.14	4.05	0.32
aa_2	62.6	25.7	7.1	7.0	1.9	1.01	3.75	0.28
aa_3	59.9	27.2	8.5	7.0	1.8	1.21	4.6	0.31
aa_4	64.2	20.4	6.9	6.7	2.0	1.02	3.45	0.34
AlphaFold – MaSp1								
200	1.1	11.1	3.3	3.5	-	0.95	-	0.30
400	6.0	12.2	3.2	3.6	2.1	0.90	1.51	0.26
800	14.0	17.8	4.9	5.1	2.2	0.96	2.27	0.28
1600_1	57.0	23.7	7.1	7.4	2.6	0.96	2.75	0.30
1600_2	52.2	22.6	6.6	7.2	2.4	0.92	2.70	0.29
1600_3	46.8	19.2	5.6	6.5	2.2	0.87	2.61	0.29
1600_4	36.7	19.4	5.8	6.3	2.3	0.92	2.46	0.30
1600_0	89.2	26.6	7.2	6.9	3.2	1.03	2.25	0.27
Mutational Study								
4A Poly(G)	4.9	6.8	2.0	2.5	1.4	0.83	1.5	0.30
4B Poly(A)	4.3	6.1	1.9	2.4	1.7	0.81	1.13	0.31
4C Poly (AG)	4.4	6.0	1.8	2.3	1.6	0.77	1.16	0.30
4D (A->G)	5.8	5.8	1.9	2.5	1.6	0.75	1.22	0.33
4E (-TYR)	5.6	7.7	2.2	2.6	1.4	0.86	1.62	0.29
4F (WT)	5.6	7.8	2.3	2.6	1.4	0.89	1.62	0.29
4G1 Rand	5.5	7.1	2.1	2.5	1.5	0.87	1.42	0.30
4G2 Rand	6.4	8.4	2.5	2.7	1.3	0.93	1.89	0.30
4G3 Rand	6.9	9.0	2.7	2.8	1.3	0.97	2.13	0.30
4H	5.8	6.4	2.0	2.5	1.5	0.81	1.37	0.32
4I	8.1	9.3	2.7	2.6	1.4	0.83	1.50	0.29
4J	5.5	7.6	2.3	2.6	1.3	0.88	1.76	0.30
4K	6.1	8.3	2.6	2.7	1.4	0.94	1.90	0.31

For all the tubule structures under consideration the identity of solvent exposed residues was determined by the rolling ball method using the GetArea algorithm and the results are presented in **Table 2** for the MaSp1 and MaSp2 proteins (see Methods). A set of eight MaSp1 conformations, drawn from six simulations, and four MaSp2 conformations, drawn from four simulations, were analyzed. The surface areas for each residue type were summed for each conformation, as well as the relative populations of inside and outside residues of that type.

There are relatively few residue types found in the spidroin proteins. 95% of the sequence is drawn from Gly, Ala, Gln, Tyr, Ser, and Arg in MaSp1 with the addition of Pro in MaSp2. The bulk of the residues are Gly and Ala (75.0% MaSp1 and 64.6% of MaSp2). A simple model for packing spheres into cylindrical shells suggests a 70-30 split between bulk and surface residues for geometries matching our tubule parameters (50 nm long cylinder with a diameter of 4 nm) and total number of amino acids found in the MaSp proteins (3100 for MaSp1 and 3779 for MaSp2). The values arising from this model are in close agreement to results from solvent exposure analysis when *all* the residues in the protein are considered (blue text in **Table 2**). On average, 66% of the residues are buried in MaSp1 and 15% solvent exposed. The remaining 19% possess similar solvent exposure ratios as their standard random coil value, and thus lie somewhere between buried and exposed, i.e., near the surface. For MaSp2, these figures are 68%, 14% surface exposed and 17% somewhere between exposed and buried. Thus, both MaSp1 and MaSp2 solvent exposure analysis agrees well with the 70/30 split from our geometric model. Other dimension cylinders have considerably different surface to volume ratios.

Comparing sub-populations of different residue types against this standard 70/30 split is illuminating. For MaSp1 we see that Gly and Ala are significantly over-represented in the core (~76%) and under-represented on the surface (~22%). Gln, Tyr and Arg are considerably under-represented in the core with only around 20-50% of those residues within the interior, but 30-31% are fully solvent exposed, and 20% partially exposed (i.e., somewhere between a 20-80 and 50-50 split for interior to exterior). For MaSp2 the situation is similar. Gly, Ala and Ser all have ~80% in the core and 5-11% in the surface, compared with Pro, Gln, Tyr and Arg showing between 20-40% in the core and 17-36% located in the surface region. The concept that the poly(Ala) is buried in the protein core agrees with NMR relaxation data collected on the same system where the poly(Ala) was shown to have the most restricted dynamics by heteronuclear NOE measurements. It is also noteworthy to mention that SSNMR has shown that Ala, Gly, and Ser comprise the nanocrystalline β-sheet domains in *L. hesperus* Ma fibers and the simulated models show that these domains are primed for β-sheet assembly by being buried in the solvent-poor, protein core. Pro adopts mostly a surface role as expected. Another way of viewing the same results is that Gln and Tyr account for only 16.2% and 12.1% of the overall number of residues in MaSp1 and MaSp2 respectively but provide 42.3% of the surface area in MaSp1 and 30.8% of the total surface area for MaSp2, considerably punching above their number weight in terms of solvent accessibility.

Finally, it is worth noting that although the bulk of the Ala, and Gly residues are indeed within the core of both proteins, there are still substantial numbers of Ala and Gly with solvent accessibility. In fact, they form some 40% of the outer surface. This may be critical for inter-protein interactions in building fibril formations between protein chains, suggesting a role for the outer residues such as Gln, Tyr and Arg in driving and maintaining the formation of a tubular structure as a means of managing inter-protein contacts. Although speculative, this is an interesting prospect that can be tested with future modeling of multi-chain interactions.

Table 2. The relative number of each type of residue (N) and the solvent accessible surface area occupied by those residues expressed in absolute terms and as a percentage from the MD structures. MaSp1 is averaged over 8 structures and MaSp2 over 4, with the standard deviation in brackets. Also indicated are the relative proportion of the residues that are buried (inside) and on the surface (outside) as determined by the GetArea algorithm. The ratio between the occupied surface area and the number of that type of residue (A/N) or the ratio of outside to inside (O/I) are indicators of residue distribution throughout the structure. Red denotes residues with indicators less than the whole protein, and green denotes residues with indicators greater than the whole protein. Residues colored red are considered predominantly buried residues, whereas those colored green are considered as surface residues.

Residue	N	N (%)	Surface Area	Area (%)	A/N	Inside (%)	Outside (%)	O/I
			MaSp1 (8)					
ALL	3132	100.0	73276 (4627)	100.0	1.00	66 (2)	15 (1)	0.23
GLY	1330	42.5	15468 (783)	21.1	0.50	76 (1)	8 (1)	0.11
ALA	1019	32.5	14230 (1345)	19.4	0.60	76 (2)	13 (2)	0.17
SER	76	2.4	2355 (232)	3.2	1.33	51 (3)	26 (6)	0.51
GLN	357	11.4	20570 (1466)	28.1	2.46	34 (4)	31 (4)	0.91
TYR	151	4.8	10408 (993)	14.2	2.95	31 (7)	22 (4)	0.71
ARG	51	1.6	3890 (206)	5.3	3.26	24 (7)	30 (4)	1.25
Others	148	4.8	6355	8.7				
			MaSp2 (4)					
ALL	3779	100.0	88954 (1291)	100.0	1.00	68 (0.4)	14 (0.4)	0.21
GLY	1267	33.5	12044 (454)	13.5	0.40	81 (1)	5 (0)	0.06
ALA	1176	31.1	14750 (401)	16.6	0.53	79 (1)	11 (1)	0.14
SER	267	7.1	4136 (205)	4.6	0.66	76 (2)	7 (1)	0.09
PRO	324	8.6	14131 (319)	15.9	1.85	37 (2)	36 (3)	0.97
GLN	261	6.9	15738 (543)	17.7	2.56	33 (2)	31 (1)	0.94
TYR	197	5.2	11632 (218)	13.1	2.51	42 (1)	17 (2)	0.40
ARG	68	1.8	5738 (234)	6.5	3.59	20 (5)	35 (4)	1.75
Others	217	5.7	10785	12.1				

A set of silk mutants were simulated (**Figure 19**) to reveal the role that each class of residue plays in the emergence and detailed character of the tubule structure. A sequence of 474 residues taken from the wild-type sequence does indeed form tubular structures as expected. The sequence chosen is taken from the main body of the native sequence, and consists of 16 Gly-rich and poly(Ala) repeating motifs. The details of the modelled mutations reveal hypotheses for the nature of the amphiphilic pattern in the primary sequence. Why should there be 5 to 10 residue repeats of Ala and 17 to 25 residues for the Gly-rich repeat unit, and what is the role of the disrupting X-residues such as Gln, Tyr, and Arg which our model suggests are predominantly surface moieties? The starting structure for each of the mutational simulations was a random coil similar to the one shown in **Figure 11**.

When the sequence is 474 Gly (**Figure 19A**) we see the radius of the final tubule is a little narrower than the wild type with a smaller R_g. However, it is significantly more ordered. The residues bundle into β-sheet-like stacks forming a cuboid shaped tertiary structure with a shape factor of 0.83 (**Table 1**, R_g/R_h column) which is highly compact. The details of the highly organized packing are reminiscent of amyloid structure. The length of these Gly strands are around 10 residues long (~3 nm), which is as long as the poly(Ala) sub-blocks in MaSp proteins. As the structure folds throughout the trajectory, we see it coalesce into beads of 245, 55 and 174 residues whose internal structures end up as highly organized sub-domains, that we have colored distinctly in the interim trajectory structures in the figure. Such sizes of intermediate structures are again consistent with the idea that 150-200 residues long are the minimal coarse grain unit.

A similar result happens when the structure is purely Ala as shown in **Figure 19B**. In this case the resulting organized strands are 15 to 20 residues long (3.5 - 4 nm), which is about the length of the amphiphilic blocks of Gly. By the end of the trajectory the pure Ala structure forms a single organized domain of aligned strands with a shape factor of 0.81 (**Table 1**) , categorized as

spherical. We note that in our model the Gly strands naturally form structures of the size of the poly(Ala) domains in the spider silk, and that the Ala strands naturally form structures about the size of the Gly domains. We do not know if this is significant or merely a coincidence; the chemical character of Ala and Gly is similar.

When we create a pattern consisting purely of alternating blocks of pure Gly or pure Ala in precisely the same block structure seen in the native silk (we achieve this by mutating all non-Gly or non-Ala residues into Gly) we see similar single domain structure emerge which, again, is almost a cuboid volume and consisting of entirely aligned strands (**Figure 19C**). This structure has an R_g/R_h value of 0.77 indicating it is spherical in nature (**Table 1**). It is likely that the identity of the Gly and Ala is irrelevant in this structure, as it forms very similar structures to pure Ala, and pure Gly alone. However, detailed visual inspection of the structure in VMD suggests that the Ala strands are quite well isolated from each other by Gly regions with fewer Ala-Ala region contacts. This observation suggests that a principle of frustration may be at work, in which the block pattern is selected to disrupt the intrinsic structures that each type of residue adopts singly. Perhaps the irregularity of the repeating pattern is an emergent moiré pattern between the competing natural sizes for each kind of residue and the block-like nature of the pattern in which sometimes the Ala regions occupy a loop, sometimes strands. This pattern would ensure that some Ala are on the surface and not buried in the core. We note that the overall diameter of the mixed Gly-Ala structure is midway between the diameters of the pure Gly and pure Ala structures.

The picture changes radically when we create a sequence in which there are no Ala at all, yet we leave all the disrupting polar X-residues in the structure (**Figure 19D**). In this case the structure first forms a tubule which then readily collapses and forms a compact globule in which there is no long-range anisotropic structure at all. The shape factor is 0.75 which is the most spherical shape in any of the simulations conducted. The inset highlights the backbone cartoon which contains only hairpin turns throughout, with no β-strand forming regions or alignment between the strands of the type we saw in the first two mutational models. Therefore, our model suggests that spacing of the non-Ala and non-Gly residues has the effect of disrupting the formation of a regular strand-like structure. This is an obvious test target for future studies to see how the amount, number and distribution of X-residues affect the formation of a regular amyloid-like alignment between the poly(Ala) blocks of the spider silk proteins.

It is only when we add all the components—the Ala and Gly in a block pattern and the X-residues—that suddenly a tubule structure re-emerges. Both the Tyr knock-out structure (**Figure 19E**) and the Wild type (**Figure 19F**) have shape factors that jump back towards coil (**Table 1**). The Tyr knock out shape factor is 0.86 and the Wild type 0.89. It is intriguing to note that the spacing of the disrupting residues, such as Gln and Tyr, (within the Gly blocks) is such that a portion of the Gln and Tyr are exposed to the external surface of the protein. It is likely that Gln-Gln, Tyr-Tyr, and/or Gln-Tyr contacts form a kind of energetic barrier that helps to stabilize the tubule, like surface tension, and helps to mediate contact between the molecules in the Ma silk dope, while the poly(Ala) are trying to form regular crystalline β-sheets. In the Tyr knock out (**Figure 19E**), we see that the tubule structure is still present suggesting that Gln contributes significantly in stabilizing the tubule structure. The Gln residues are predominantly on the surface and there is a much less organized character to the turn and strand structure of the protein. Thus, we conclude the Tyr and Gln disrupt the natural propensity of the Ala and Gly to form long 3-4 nm strands, resulting in a much more tubular shape with similar radius but much less pre-

organization between the Ala-rich regions. The latter could be responsible for preventing premature β-sheet aggregation of the proteins in the spider silk gland environment.

The three scrambles (**Figure 19G**) show that a randomized sequence also forms tubules (Shape factor 0.87, 0.93 and 0.97), but there is quite a wide variance in shape factor that brackets the wild type (**Table 1**). We note that in scramble 1 all the Tyr randomly collected in one half of the structure. This asymmetry we believe resulted in a kink emerging in the tubule, as one half did not have the stabilizing effect of the Tyr at the surface. Although such kinks were observed elsewhere, they were somewhat less pronounced. This experiment of opportunity suggests a hypothesis that the long-range persistence length of the higher order tubule may be affected by the distribution of the disrupting residues.

The results from the second panel of **Figure 19 (H-K)** support the emerging picture. In this sequence of simulations, we selectively rearrange the sequence and move entire blocks of Ala rich regions around, while keeping the overall proportions of residues the same. In this way we change the block pattern of the protein. We see that when the poly(Ala) is in a single continuous stretch, as in **Figure 19H**, they cause the tubule to have a much larger radius at one end than the other, resulting in a shape factor of 0.81 which is moving back towards spherical (Table 1). This result is intriguing. If there is too much Ala to dominate the structure (as in **Figure 19B, H and I**) the shape factor becomes more spherical. As we systematically redistribute the Ala back into the 16-mer pattern of the wild type we see a systematic change in shape factor back to coil. Thus, **19J** and **19K** in which there are two and four regions of Ala, there is a more coiled character. This trend is direct evidence that the distribution of the Ala-rich regions directly controls the overall geometry of the tubule and that there is a delicate balancing act between the distributions of Ala, Gly, Tyr and Gln, which yields a tubular structure that is at least meta-stable.

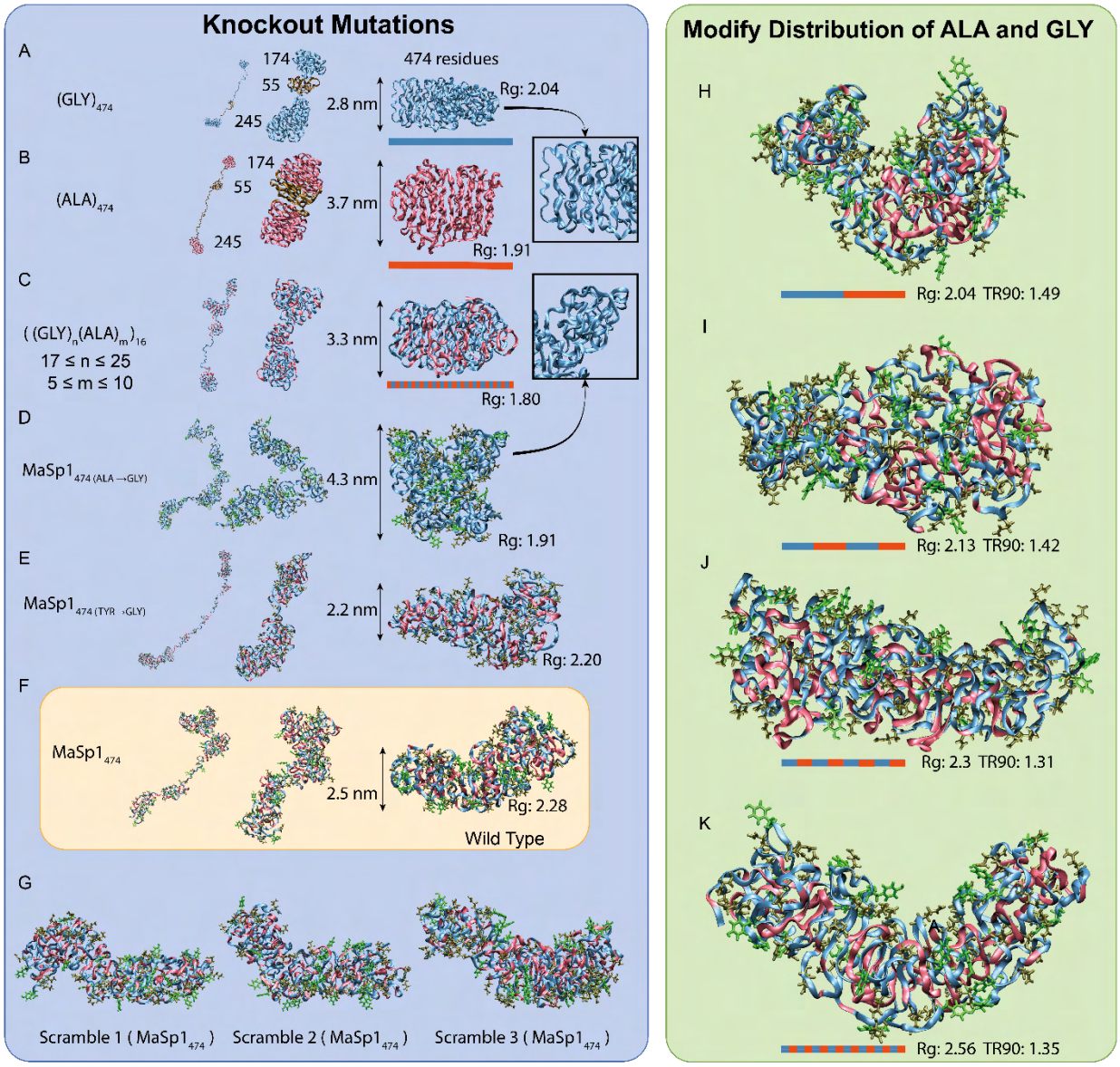

Figure 19. CG MARTINI simulations of a sequence of 474 residues taken from MaSp1 show the effect on the tubular structure of various mutations in the sequence. (A) A chain of 474 Gly, (B) A chain of 474 Ala, (C) All residues in the 474 residue MaSp1 sequence, except Ala, are replaced with GLY, leaving the block co-polymer character of the sequence intact. (D) MaSp1 sequence with all Ala mutated to Gly. (E) All Tyr in the wild-type sequence is mutated to Gly. (F) The 474 residue sub-sequence taken from wild type MaSp1 that is the starting point for all the sequences. (G) Three random scrambles of the wild type sequence. (H) The 16 blocks of poly(Ala) sequences are gathered into a single block and placed at the C-terminus of the sequence. (I) The 16 blocks of poly(Ala) sequence are gathered into two groups and placed at separate parts of the sequence. (J) The poly(Ala) blocks are gathered and placed at four places in the sequence. (K) The poly(Ala) blocks are gathered and placed at 8 places in the sequence.

To probe the relevance of the simulated protein architectures in native Ma silk dope, we used solution NMR to obtain experimental structural data. Solution NMR chemical shifts are a long established and widely used source of secondary structure information. When combined with sequence information and database-based search analysis of published structures, these chemical shifts can provide full three-dimensional solution structures in many cases. We therefore employed solution NMR to characterize secondary structure and determine low-energy structural ensembles for native MaSp proteins from NMR chemical shifts to further compliment the MD simulated structures with experimental data. The determined solution NMR chemical shifts were used as inputs to NMR computational tools TALOS-N and CS-Rosetta for secondary structure characterization and structural ensemble determination. Ma glands were enriched in NMR-active nuclei, excised from black widow (*L. hesperus*) spiders, and used for a broad range of multidimensional structural NMR experiments discussed below.

The biggest challenges with NMR analysis of spidroin proteins stem from many factors including their massive size (250-350 kDa), highly repetitive block co-polymer nature, and intrinsically disordered conformational state. This results in extremely low chemical shift dispersion and extensive broadening of the observable signals that cannot be overcome simply by going to higher field NMR (although we do see some increase in resolution at high field). In order to tackle these issues, we applied an arsenal of multidimensional experiments including traditional ^{1}H-detected 2D and 3D NMR together with 2D X-detect correlation experiments CACO and CON. **Figure 12** shows the correlation between the two X-detect experiments, and their 2D ^{1}H-based complements. The broadening and overlap in the ^{1}H-detect experiments can be successfully separated out using the X-detect experiments, which allows for sequential assignment for residue pairs. Two residue carbonyl assignments were made using this combination of X-detected CACO and CON experiments for residue pair sequential assignments. These pairs of assignments were then added to the sequences based on their order of appearance in the full *L. hesperus* MaSp1 sequence. Traditional 3D NMR experiments were used to make sequential assignments for non-carbonyl nuclei in a 15-residue repeat sequence of MaSp1. A full 6-residue assignment from the Gly-Gly-X region was assigned and then placed alongside assignments from the poly(Ala) region (see Methods). This resulted in a nearly complete backbone nuclei assignment for the following sequence: GQGGAGAAAAAAAAG**.** This sequence repeat is highly represented throughout the repetitive core of the *L. hesperus* MaSp1 sequence and appears 39 identical times in the sequence comprising ~20% of the repetitive core. This 15-residue domain contains the Gly-Gly-X (X = Gln) repeat and a long poly(Ala) block containing eight Ala. We have previously reported the chemical shift assignments for five of the backbone nuclei in this short sequence (**ref 3**). Now we complete the assignment with the addition of the carbonyls. This primes the data for high quality structure determination using a combination of NMR computational tools.

Building on the results presented in **Figure 13**, we used these complete NMR chemical shift assignments and a combination of computational techniques (TALOS-N and CS-ROSETTA) to make determinations about probable structural ensembles from the NMR data and conduct secondary structure analysis. The top-ten lowest energy CS-ROSETTA structures determined for the 15-residue domain and longer 120-residue (comprised of eight 15-residue repeats linked together) are shown in **Figure 20** along with their Ramachandran plots. Close inspection of the Ramachandran plots reveals a high population of Type I and Type II β-turns with very low β-sheet and helical content, which is consistent with the MD models presented above. Complete secondary

structure quantification for the CS-ROSETTA structures was conducted with the DSSP algorithm (**Table 3**).

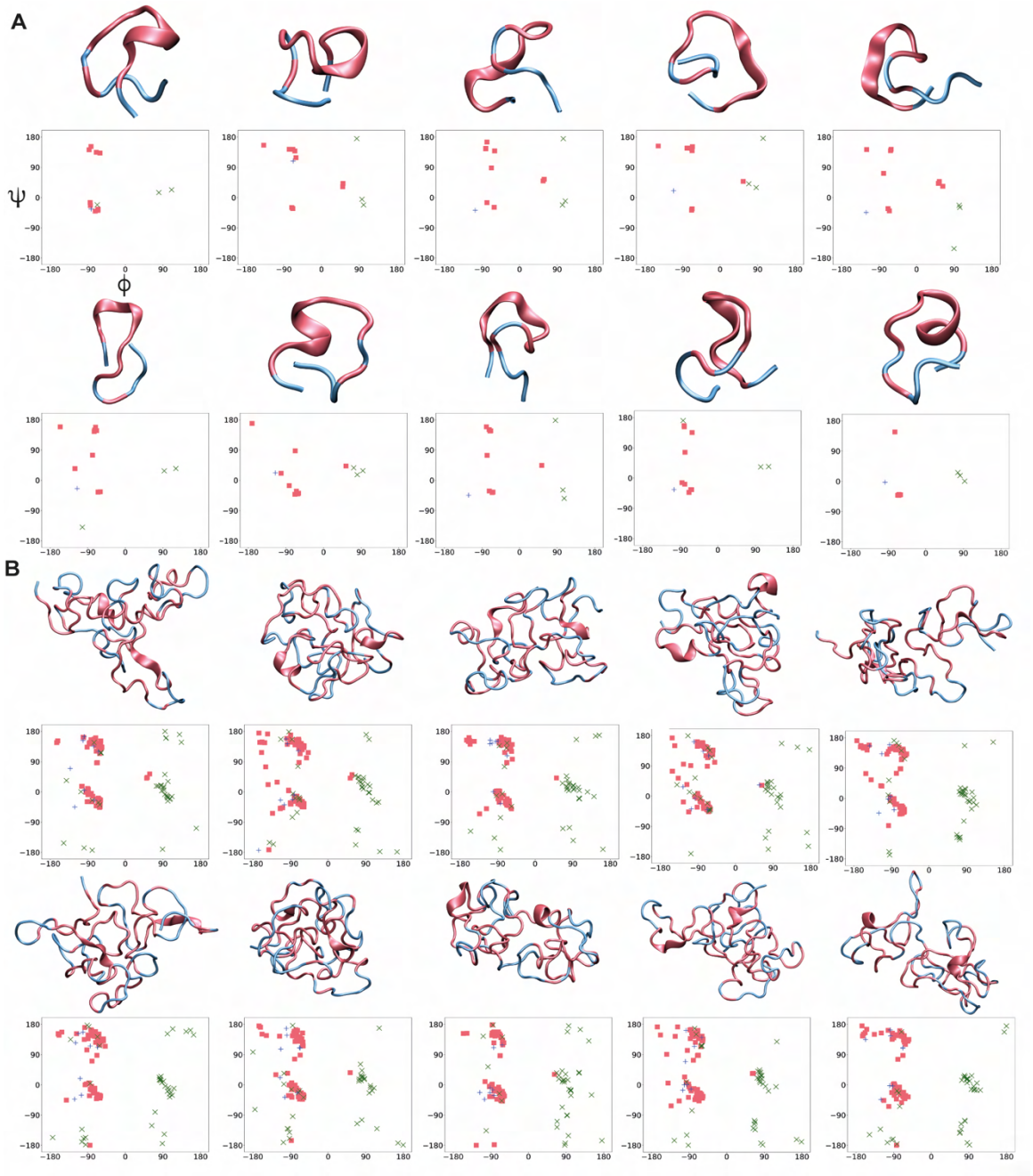

Figure 20. Top ten lowest energy structures determined with CS-Rosetta from solution NMR determined chemical shifts for (A) 15-residue and (B) 120-residue GQGGAGAAAAAAAAG repeating unit. The Ramachandran plot for each structure is shown below with Ala indicated by red squares, Gly by green x and Gln by blue +. The Ramachandran plot shows strong clustering for residues exhibiting Type I and Type II β-turns with nearly no evidence for β-strand and low population of helices. Pink represents Ala in structures.

Table 3. Secondary structure quantification for the NMR-determined CS-ROSETTA ensemble structures. The top ten lowest energy 15- and 120-residue structures were analyzed with DSSP and secondary structure quantification is expressed as a percent.

Secondary Structure	15-residue	120-residue
α-Helix	4.0	3.8
β-Bridge	0.0	3.2
β-Strand	0.0	0.6
3_{10}-Helix	2.0	5.1
π-Helix	0.0	0.0
β-Turn	22.0	28.0
Bend	30.7	25.3
Unstructured	41.3	34.1

The results from DSSP also show that the structures on average contain high populations of β-turns (defined H-bond) and bends (no defined H-bond) along with the more dominant unstructured (random coil) conformation. DSSP analysis shows that β-strand structure is essentially non-existent with a low occurrence of helices (α-helix and 3_{10}-helix). β-turns and bends represent greater than 50% of the secondary structure with unstructured regions accounting for 35-40% in both 15-residue and 120-residues structures. Essentially no β-sheet structure is detected in any of the low energy CS-ROSETTA structures with a slightly higher helical content observed in the 120-residue structures but, still accounting for <10% of the total conformation. Although this is not a complete secondary structure assessment of the MaSp proteins, the high occurrence of this 15-residue repeat in MaSp1 and the substantial β-turn/bend fraction observed in the NMR-determined structural ensembles, one can easily envision that this could lead to the tubular shaped tertiary structure from the MARTINI CG-MD simulations (**Figure 11**).

Progress in EM of Native Spider Silk Dopes (obj. 2):

The Northwestern (Gianneschi Lab) electron microscopy portion of the spider silk project is targeted at characterizing the native morphology of silk proteins prior to and during the silk spinning process. Our team's prior work yielded the first direct verification of the silk micelle

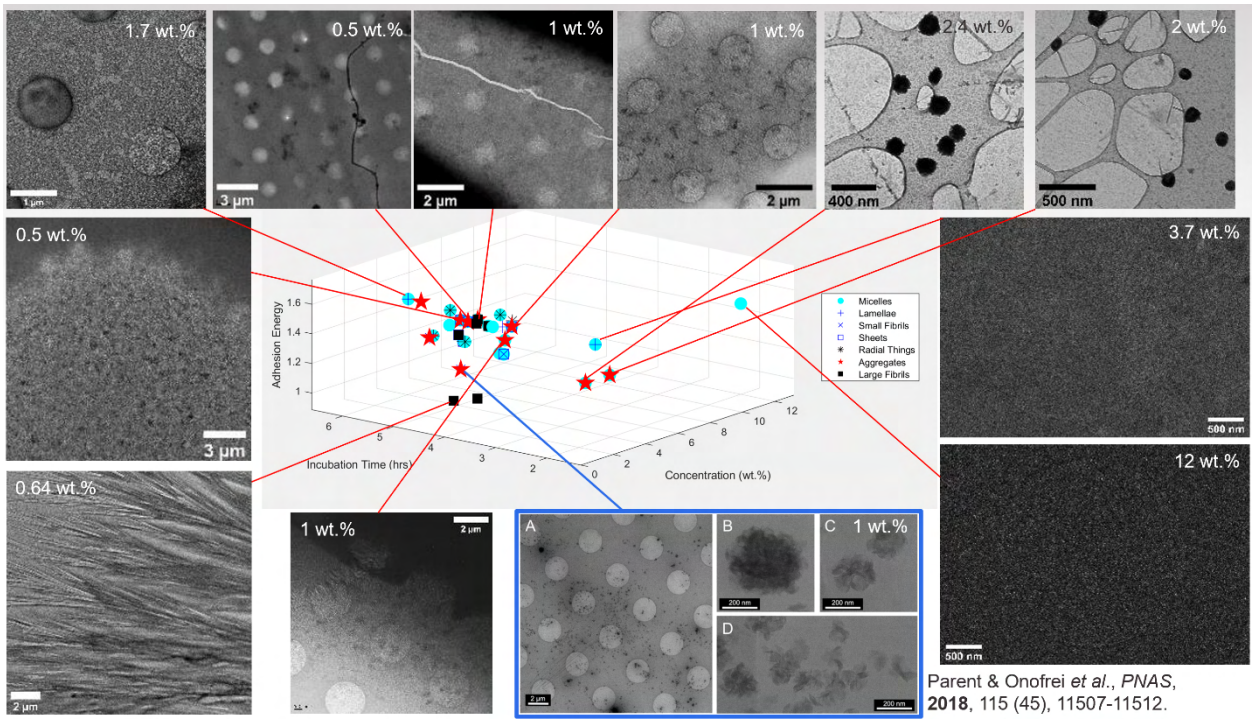

Figure 21. Cryo-TEM of dissected, 4M urea-diluted black widow MA silk. Attempts to replicate and extend the previous work by Parent and Onofrei (blue) at different concentrations and under different salt and pH conditions ran into consistent sample preparation issues with appropriate silk loading and clean vitreous ice production. A significant variety of silk structures were observed – dense packing of small micelles occurred at high concentration, while collapsed structures occasionally occurred at low concentrations.

hypothesis using diffusion NMR and tomographical tilt-series cryo-TEM reconstructions of *ex situ*, highly dilute silk protein complexes in urea, as well as the response of the complex to vigorous pipette shearing. One initial objective of the Northwestern portion of this collaboration was to reproduce that work and extend it, probing the responses of the silk assemblies to typical spinning conditions, including pH, salt concentration, and phosphate addition, and the second microscopy objective was probing the spinning process itself. Significant exploration of the spidroin phase space (**Figure 21**) showed that the silk micellar structures in the *ex situ* dope often adopted a condensed, collapsed phase. The secondary objective of exploring the spinning process using microfluidic devices output onto cryo-TEM grids was abandoned after realizing the variability of the silk proteins just from the dope; to pursue that objective, we pivoted to ultrastructure characterization of the whole gland using structural biology techniques. Typically, that kind of structural biology uses chemical fixation methods. These methods use a combination of chemical fixatives, stains, and organic solvents to crosslink the proteins and membranes in the tissue of interest, and the features in question are then imaged by TEM after the resin-infiltrated block is cured and sectioned. However, as spider silk is well known for its extreme sensitivity to perturbation – even the gentle prod of a tweezer can produce enough shearing to pull fibers, and pH shifts from 7 to 5.5 are attributed drivers of natural fibrillarization – we were very concerned that chemical fixation would destroy the silk protein structure while preserving the cellular structure in the silk gland membrane. This worry was vindicated by a paper from Casem *et al.* that chemically fixed the silk glands of black widow spiders to study the cellular machinery in the gland – and yielded very collapsed, dehydrated silk in the gland lumen. To avoid these issues with chemical fixation, we used the other available structural biology fixing technique: high pressure freezing followed by freeze substitution (HPF-FS).

In HPF-FS, small tissue samples are flash-frozen into a thin puck by high-pressure jets of liquid nitrogen. This method allows for similar vitreous ice production as the liquid ethane plunge-freezing that is standard in cryo-TEM sample preparation, but for samples hundreds of microns thick instead of the tens of nanometers possible in plunge freezing. Once frozen, HPF samples can be stored safely in liquid nitrogen for weeks to months with minimal degradation until freeze substitution.

The freeze substitution processes are long and very detailed, but at their core entail doing the chemical fixation steps at cryogenic temperatures to preserve structure, then doing any further labeling and resin embedding at room temperature after the structures are fixed.

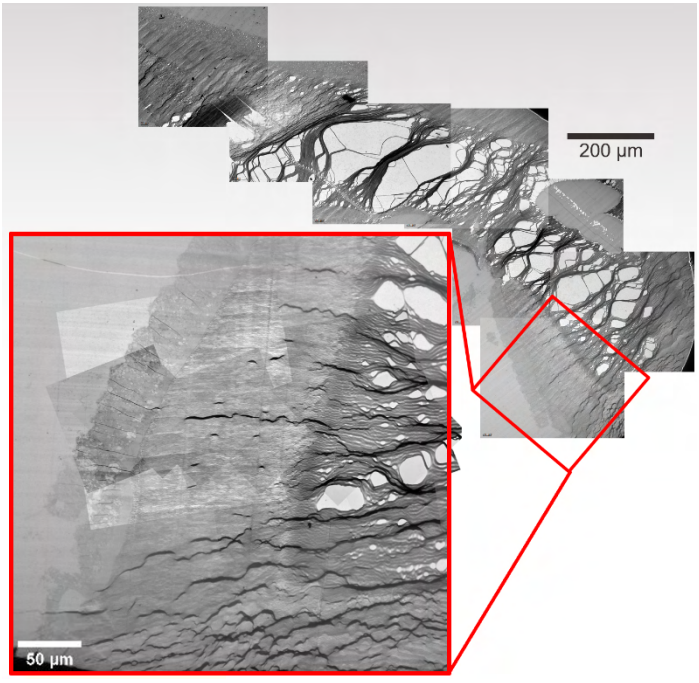

Figure 22. HPF-FS round 1. The stitched low-mag series extends 1.5 mm across a 2x1 mm slot grid, and shows unusual fiber behavior perpendicular to the gland's flow direction. However, this phenomenon is due to poor resin infiltration and fixing of the silk lumen, which leads to tearing of the sections parallel to the sectioning knife edge. In the inset (red), the maximum extend of resin infiltration is easily seen by the dark line where the thicker fibrils start appearing.

First, the vitrified ice in the frozen tissue is very gradually etched away at cryogenic temperatures with an acetone solution that contains chemical crosslinkers and stains over the course of several days. Thanks to the low temperature and slow rate, the etching of structure-preserving vitreous ice and the introduction of the chemical crosslinkers occurs almost simultaneously, severely limiting if not completely removing the possibility for the crosslinker addition to drive morphological changes in the silk protein. After chemical fixation, further *en bloc* staining using dilute heavy metals (e.g. Os, U, and Fe) produces the differential contrast required for TEM imaging. Finally, the fixed blocks are gradually infiltrated with an epoxy-based resin, embedded in a resin block, and cured. After FS is completed, the resulting sample chit can be sectioned normally using ultramicrotomy, and then imaged using standard dry-state TEM methods.

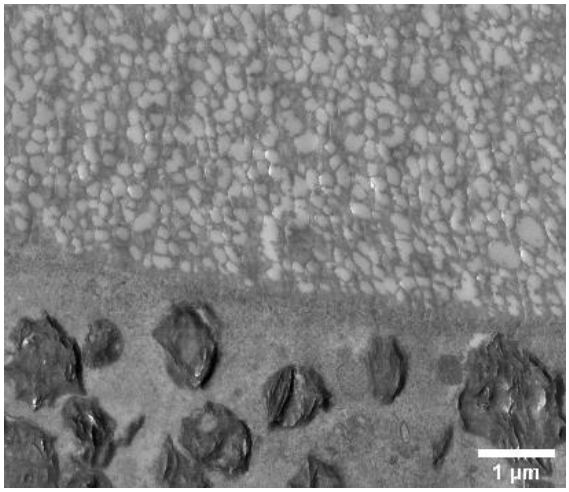

Figure 23: HPF-FS round 2. Using glutaraldehyde and longer resin infiltration times yielded better fixation, but the LLPS-esque behavior in the gland lumen raised questions about crystalline ice inclusions during HPF.

The first attempt at HPF-FS of silk glands yielded highly detailed, well-fixed membrane cells and unexpectedly fibrillar silk structures in the gland lumen (**Figure 22**). Assembling high-magnification micrographs into stitched collages provided a picture of a much more structurally diverse silk gland than expected for the pre-spinning region. After discussion with structural biology experts at Northwestern, we determined the shift in fibril size at a regular depth into the gland was an artifact of incomplete resin infiltration followed by shearing of unfixed proteins during sectioning. To remedy this, the next HPF-FS round extended the resin infiltration time by several hours. Additionally, considering the composition of the silk gland relative to the cellular structures that HPF-FS usually targets, glutaraldehyde was added to the FS protocol as a protein crosslinker to improve fixing of the gland lumen.

The second round of HPF-FS fixed the resin infiltration and ultramicrotomy shearing issues from the first round, but raised new concerns over HPF-induced artifacts (**Figure 23**). With uniform composition and features across the lumen of the gland, the observed features cannot be blamed on a partial infiltration or incomplete fixing. Our initial hypothesis was that we had managed to capture early stages of liquid-liquid phase separation in the silk gland, as the FS process would have been prevented fixative-induced phase separation. However, as LLPS in spider silk has been demonstrated to be an effect of phosphate addition, and recognizing that phosphate concentrations in the gland ampulla is far below the concentrations that have been observed to trigger LLPS, we were very uncertain of that hypothesis. Discussions with relevant faculty raised the possibility that these features could be the effect of ice crystal inclusions formed during HPF which were then etched away during FS. That effect has been seen in HPF of clay slurries previously – clay slurries with similar salinities but lower clay mineral content with thicknesses of 200 µm or more produced micron-scale hexagonal ice crystallites that massively altered the microstructure of the sample. To characterize this extent of ice contamination, synchrotron cryogenic wide angle x-ray scattering (cryo-WAXS) of post-HPF pre-FS samples was undertaken at Sector 5 of the Advanced Photon Source.

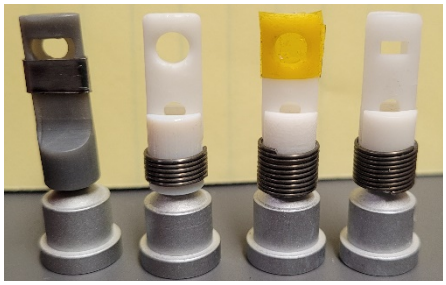

Figure 24. Evolution of cryo-WAXS sample holders. From left, initial cryo-fragile version with metal clip ring; sandwiched clamshell, sandwiched clamshell with Kapton windows, and final slotted clamshell.

Synchrotron cryo-WAXS is a particularly attractive technique to probe ice crystallite sizes in an otherwise disordered HPF silk gland sample due to its crystal habit sensitivity, penetration depth, and cryogenic sample preservation. In the WAXS region, the only species scattering x-rays from the silk gland is the crystalline ice present inside or on the surface of the sample, and in the event of perfectly vitrified samples, a characteristic amorphous ice hump can be observed. Additionally, analyzing the WAXS pattern using a Scherrer peak width analysis allows us to calculate the average ice crystallite size for that sample, which can then be used as a "feature floor" for observed structures: any LLPS-like features significantly larger than the ice crystallite size are likely real, and any that are smaller are likely etched-out ice inclusions. Adding cryo-WAXS to the workflow from a sample perspective does not alter or damage the samples prior to FS as long as the silk glands are kept at cryogenic temperatures during cryo-WAXS; installing a cryojet at the beamline allows for those temperatures, though installing the cryojet at the beamline we used did require remounting the sample goniometer at 90° to its normal position to avoid freezing it, as well as designing custom sample holders for HPF pucks (**Figure 24**) and developing sample preparation protocols to extract the HPF samples from the HPF carriers intact. The holder redesign was key to success – initial holder designs based on recollections of previous cryo-WAXS experiments proved ineffective, but the final clamshell-and-clamp design worked almost perfectly (with only occasional cryo-induced shattering). The sample release problem proved the most intractable part of the cryo-WAXS efforts: over the course of four cryo-WAXS beamtime slots, a total of 1, 3, 7, and 13 samples were characterized on respective beamtimes. The increasing sample number was almost entirely due to improvements in sample release from the HPF: by moving from the HPF-standard coatings (1% lecithin in CHCl3, 2% lecithin in CHCl3, and n-hexadecane) to more aggressive lubricants (3-in-1 Lock Dry Lube), the HPF pucks popped out of the planchets much more completely, and dissection-to-beamline success rates went from 17% to 72%. Cryo-WAXS demonstrated that there are easily distinguished differences in ice crystallite inclusion sized across different samples and using different freezing methods (**Figure 25**), and these discrepancies can be readily seen by TEM as well. Using the same FS protocol for both plunged and HPF samples, radical difference in morphology can be observed (**Figure 26**). The plunged sample exhibits very large, uniform contrast voids with erratic shapes and sizes, indicative of etched out crystalline ice inclusions. Meanwhile, the HPF sample retains extremely well-preserved cellular features and well as well-fixed silk in the gland lumen, with no LLPS type features present. Additionally, the lumen silk exhibits a globular morphology very similar to the high concentration

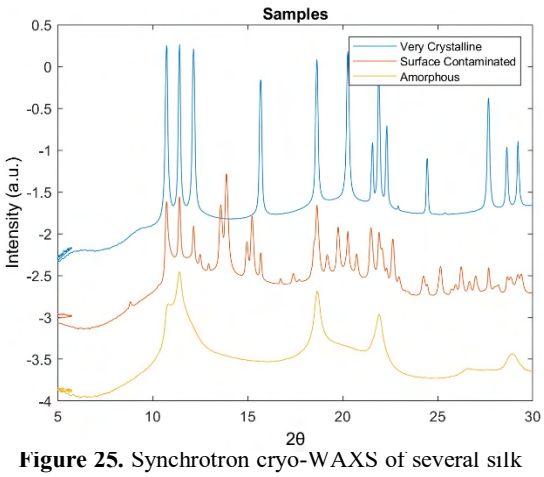

Figure 25. Synchrotron cryo-WAXS of several silk gland samples. The very crystalline sample (blue) is a plunge-frozen gland with large ice crystallites; the amorphous sample (yellow) has only small crystallites in an amorphous matrix after HPF; the orange sample has mid-sized crystallites from HPF as well as contamination doublet peaks from the planchet release coating.

samples probed by traditional cryo-TEM (Fig. 1, 12 wt.%) and corresponds well with AFM images of intermediate stages of silkworm silk coating deposition in the literature, where silk proteins initially form globules before merging into a film or fiber. Most characteristically, the tears in the plunge-frozen sample are randomly located and shaped throughout the section; the discontinuities and tears in the HPF sample all trace membrane-lumen boundaries and correspond to the mechanical property differences between the two phases.

With a high-fidelity method in hand to characterize the natural spider silk spinning process and robust error-identifying methods available to identify artifactual features, all that remains is fully mapping the silk evolution through the gland and duct. Serial acquisition and automated image stitching algorithms are presently being adapted for this purpose; there are still some issues to fix, but full section maps should be finished in the near future as this collaborative effort continues.

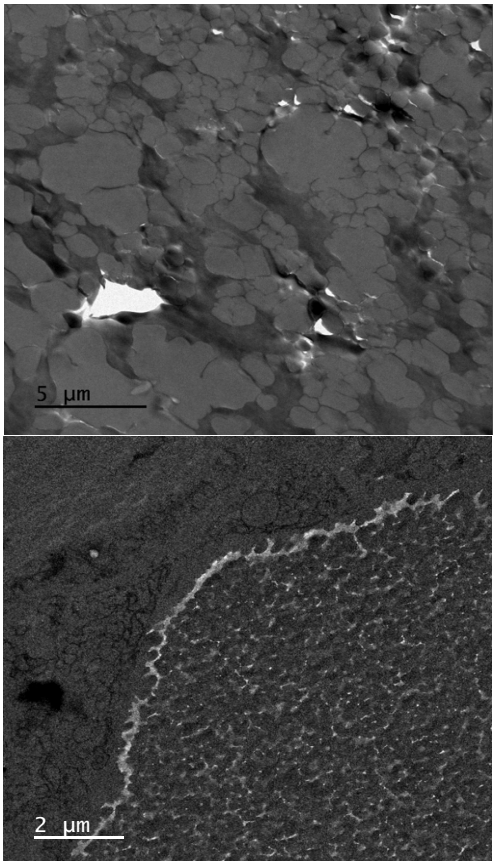

Figure 26. Comparing plunge-frozen (PF, top) and HPF (bottom) silk gland samples. The PF sample is the highly crystalline blue trace in Fig. 4; the HPF sample is amorphous yellow trace. Identical FS protocols were used on both samples. The PF sample exhibits severe ice inclusions, while the HPF sample maintains exquisite sample preservation of both cellular membrane structures and in the silk lumen, where ~300 nm globular structures can be observed. These structures correspond nicely to the high-concentration prior work (Figure 21, 12 wt.%) as well as literature silk assembly work.

6. Personnel Supported:

1) Gregory P. Holland (PI, SDSU), Supported one month each year.

2) Dr. David Onofrei (Research Scientist, SDSU), Supported one month each year.

3) Hannah Johnson (Ph.D. Student, Research Assistant, SDSU), Supported 2 years.

4) Dr. Kevin Chalek (Postdoc, SDSU), Supported 2 years.

5) Chris Sharpe (Ph.D. Student, Research Assistant, U. Northwestern), Supported 2 years.

6) Dr. Chris Forman (Research Scientist, Northwestern), Supported one month per year.

6. Publications:

1) Addison, D., Stengal, D., Bharadwaj, V.S., Happs, R.M., Doeppke, C., Wang, T. Bomble, Y.J., Holland, G.P., Harman-Ware, A.E. "Selective One-dimensional ^{13}C-^{13}C Spin Diffusion Solid-state Nuclear Magnetic Resonance Methods to Probe Spatial Arrangements in Biopolymers Including Plant Cell Walls, Peptides and Spider Silk", *J. Phys. Chem. B* **2020**, *124* 9870-9883. ***[Made the Cover]***
2) Stengal, D., Addison, J.B., Onofrei, D., Huynh, N.U., Youssef, G., Holland, G.P. "Hydration-induced β-sheet Crosslinking of α-helical-rich Spider Prey-wrapping Silk**"** *Adv. Funct. Mater.* **2021,** *31* 2007161. ***[Made the Cover]***
3) Onofrei, D., Stengal, D., Jia, D., Johnson, H.R., Trescott, S., Soni, A., Addison, B., Muthukumar, M., Holland, G.P. "Investigating the Atomic and Mesoscale Interactions that Facilitate Spider Silk Protein Pre-Assembly" *Biomacromolecules* **2021**, *22*, 3337-3385.
4) Stengel, D., Saric, M., Johnson, H.R., Schiller, T., Diehl, J., Chalek, K., Onofrei, D., Scheibel, T., Holland, G.P. "Tyrosine's Unique Role in the Hierarchical Assembly of Recombinant Spider Silk Proteins: From Spinning Dope to Fibers" *Biomacromolecules* **2023**, *24*, 1463-1474.
5) Chalek, K., Soni, A., Lorenz, C.D., Holland, G.P. "Proline-Tyrosine Ring Interactions in Black Widow Dragline Silk Revealed by SSNMR and Molecular Dynamics Simulations" *Biomacromolecules* **2024**, *Accepted*.
6) Forman, C., Onofrei, D., Stengel, D., Aldana, J.A., Gianneschi, N.C., Holland, G.P. "Molecular Dynamics and NMR Reveal Tubular Tertiary Structure Rich in b-turns and Bends for Dragline Spider Silk Proteins in Solution" *PNAS* **2024**, *In Revision*
7) Stengel, D. (2021). Understanding the Native Spider Silk Spinning Process Using Recombinant Silk Proteins. *UC San Diego*. ProQuest ID: Stengel_ucsd_0033D_20959. Merritt ID: ark:/13030/m5w73fxg. Retrieved from https://escholarship.org/uc/item/9wr500pm

7. Interactions/Transitions:

a. Participation/presentations at Meetings, Conferences, Seminars:

1) **Holland, G.P.**, "Hierarchical Assembly of Spider Silk Proteins: Exploring the Structural Biology of Natural Materials from the Atomic to the Macroscale" DOD-AFOSR (Natural Materials and Systems) Program Review, Fort Walton Beach, FL (2023). [***Invited***]
2) **Holland, G.P.**, "Elucidating Spider Silk Structure and Assembly with NMR" Gordon Research Conference: Silk Proteins and the Transitions to Biotechnologies, Bryant University, Smithfield, RI (2023) [***Invited***]
3) **Holland, G.P.**, "Elucidating Spider Silk Structure and Assembly with NMR" North Jersey ACS NMR Topical Group (2023) [***Invited***, held virtually]
4) **Holland, G.P.**, "Hierarchical Assembly of Spider Silk Proteins: Exploring the Structural Biology of Natural Materials from the Atomic to the Macroscale" DOD-AFOSR (Natural Materials and Systems) Program Review, Fort Walton Beach, FL (2022). [***Invited***]
5) **Holland, G.P.**, "Hierarchical Assembly of Spider Silk Proteins" Frontiers in Soft Matter and Macromolecular Networks, University of San Diego, San Diego, CA (2022). **[Invited]**
6) **Holland, G.P.**, "A Multimodal Biophysical Approach to Understand Biomaterials Formation" Biomaterials Seminar - University of Bayreuth, Germany (2022). [***Invited***]
7) **Holland, G.P.**, "Hierarchical Assembly of Spider Silk Proteins: Exploring the Structural Biology of Natural Materials from the Atomic to the Macroscale" DOD-AFOSR (Natural Materials and Systems) Program Review, Fort Walton Beach, FL (2021). [***Invited***]
8) **Holland, G.P.** "Developing Aciniform Spider Silk Biomaterials with Unique Structural Transitions and Properties" Seminar to the Army Research Labs (2021). [***Invited***, held virtually due to the pandemic]
9) Onofrei, D.**,** Stengel, D., Johnson, H.R., Puzio, B., **Holland, G.P.** "Spider Silks as Model Systems for the Design of Functional Protein-based Materials" AAAFM-UCLA International Conference on Advances in Functional Materials, UCLA, LA, CA (2021). [Held virtually due to the pandemic]
10) Johnson, H.R., Adams, K., Layana, C. Vallejo, S., **Holland, G.P.** "Informing Tunable Bio-composite Design with Fiber Formation in Spiders and Silkworms" AAAFM-UCLA International Conference on Advances in Functional Materials, UCLA, LA, CA (2021). [Held virtually due to the pandemic]
11) Stengel, D., Addison, B., Onofrei, D., Holland, G.P., "Water-induced β-sheet Cross-linking of α-helical-Rich Spider Prey-Wrapping Silk" American Chemical Society 261st National Meeting, (2021). *Held Virtually Due to the Pandemic*
12) Johnson, H.R. and Holland, G.P., "Spider Silk Fiber Formation as a Model System for Characterizing Biomaterials Using Infrared Microscopy" American Chemical Society 261st National Meeting (2021). *Held Virtually Due to the Pandemic*
13) Holland, G.P. and Onofrei, D., "X-detect solution NMR for Investigating the Structure and Dynamics of Spider Silk Pre-assemblies" American Chemical Society 260th National Meeting, San Francisco, CA (2020). ***[Invited]*** *Held Virtually Due to the Pandemic*

14) Holland, G.P., “Hierarchical Assembly of Spider Silk Proteins: Exploring the Structural Biology of Natural Materials from the Atomic to the Macroscale” DOD-AFOSR (Natural Materials and Systems) Program Review (2020). *Held Virtually Due to the Pandemic*
15) Addison, B., Stengel, D., Holland, G.P., Ware, A., “Selective 1D [^{13}C-^{13}C] Spin Diffusion Solid-state NMR to Probe Spatial Arrangements in Biomolecules” Experimental NMR Conference (ENC) (2021). *Held Virtually Due to the Pandemic*
16) Gianesschi, N. “Protein-like Polymers” North Carolina State University, NC (2020) *Invited Virtual Seminar*
17) Gianesschi, N. “Protein-like Polymers” Biotech by the Lake, Northwestern University, IL (2020) *Invited Virtual Seminar*

b. Consultive and Advisory Functions: None
c. Technology Assists, Transitions, and Transfers: None

8. New discoveries, inventions, or patent disclosures:

U.S. Patent WO2022150163A2·2022-07-14 (US2021063587W) titled "BIOMATERIALS AND BIOTEXTILES AND METHODS FOR MAKING SAME" originally filed December 15, 2020.

9. Honors/Awards:

1) Holland, G.P. SDSU, College of Sciences Exceptional Service Award (2021)
2) Johnson, H.R. SDSU, University Graduate Fellowship (2021)
3) Stengel, D. PMSE Best Poster Award, Spring 2021 Virtual ACS Meeting (2021)
4) Sharpe, C. Northwestern U., PPG Student Travel Fellowship (2020)
5) Li, Y. SDSU, Rees-Stealey Research Foundation Fellowship (2020)