From Soluble Protein to Anchoring Filament: Understanding the Structural and Mechanical Foundations of Pyriform Spider Silk

by

Jeffrey R. Simmons

Submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

August 2022

Dedication

For my family, friends, mentors and everyone who supported me along the way. To everyone who helped me grow so much professionally and personally over the last five years of my life. Those who gave me a hand when I was feeling down, and those who were there to raise a glass and celebrate the successes.

- Thank you

Table of Contents

List of Tables

List of Figures

Abstract

Spider silks are biomaterials used by spiders for many diverse adaptations, with toughness comparable to Kevlar and strength comparable to high strength steel. Orb-weaver spiders produce up to seven distinct classes of silk containing spider proteins (spidroins), with specific protein structural changes upon fibrillogenesis remaining unknown for most classes. The work detailed in this thesis focuses on the pyriform spidroin 1 (PySp1) from *Argiope argentata*. Based on the central pyriform silk repetitive domain of PySp1, I successfully engineered representative recombinant pyriform silk proteins. Silk fibres formed from a two-repeat-containing protein referred to as HPy_2 showed a relatively high combination of strength and extensibility, in contrast to extremes of one mechanical property vs. the other as seen in most silks. This was the first reporting of recombinant pyriform silk mechanical properties. Refinement of these expression methods in M9 media increased yield, qualitatively increased purity, and increased extensibility in spun fibers. To understand the structure of this protein in the solution state and how it changes during the fibre-forming process, I performed atomic-level solution and fibre structural studies. My results show that the repetitive domain units of PySp1 are modular and contain a highly ordered, 6-helical bundle, including an internal bridging helix, and long disordered linkers at each terminus. While the solution state structure is devoid of β-sheets, recombinant pyriform silk fibres show evidence of a partial α-to-β transition and supramolecular structure. As a whole, this work provided a benchmark for recombinant pyriform silk mechanical properties, the protein solution-state properties, and some understanding of the structural changes and assembly upon fibre formation.

List of Symbols and Abbreviations

σ: Stress

ε: Strain

δ: Chemical shift

$\Delta\delta$: Change in chemical shift

ω: Larmor frequency

ARIA: Ambiguous Restraints for Iterative Assignment

AS: As-Spun

CD: Circular Dichroism

CNS: Crystallography & NMR System

CP: Cross-Polarization

CS-Rosetta: Chemical shift ROSETTA

CSI: Chemical Shift Index

DANGLE: Dihedral Angles from Global Likelihood Estimates

DARR: Dipolar Assisted Rotational Resonance

dH_2O: Distilled water

DLS: Dynamic Light Scattering

DSS: 4,4-dimethyl-4-silapentane-1-sulfonic acid

DSSP: Define Secondary Structure of Proteins

ESI+: Electrospray ionization, positive ion mode

EtOH: Ethanol

HFIP: 1,1,1,3,3,3-hexafluoro-2-propanol

HSQC: Heteronuclear Single Quantum Coherence

IMAC: Immobilized Metal Affinity Chromatography

IPA: Isopropyl alcohol

IPTG: Isopropyl β-D-1-thiogalactopyranoside

LC: Liquid chromatography

LB medium: Luria-Bertani medium

Ma: Major ampullate (dragline) silk

MaSp: Major ampullate (dragline) silk spidroin

MeOH: Methanol

MS: Mass spectrometry

MS/MS: Tandem mass spectrometry

NOE: Nuclear Overhauser Effect

NOESY: Nuclear Overhauser Effect spectroscopy

NMR: Nuclear Magnetic Resonance

NMR-3: Nuclear Magnetic Resonance Research Resource

NRC-IMB: National Research Council Institute for Marine Biosciences

NTA: Nitriloacetic acid

PANAV: Probabilistic Approach to NMR Assignment and Validation

PCR: Polymerase chain reaction

PS: Post-spun

PSVS: Protein Structure Validation Software suite

PySp: Pyriform Spidroin (spider fibroin)

SDS PAGE: Sodium dodecyl-sulfate-polyacrylamide gel electrophoresis

SHG: Second harmonic generation

SUMO: Small ubiquitin-like modifier

TFA: Trifluoroacetic acid

TFE: 2,2,2-Trifluoroethanol

THG: Third harmonic generation

TuSp: Cylindriform/tubuliform silk spidroin

UV: Ultraviolet

Py_x: X number of block repeats from Pyriform central repetitive domain

W_3: Protein containing 3 block repeat units from the aciniform repetitive domain of *Argiope trifaciata*

Acknowledgements

I would like to start by thanking my supervisor, Dr. Jan K. Rainey. Throughout this journey, Jan has given me every opportunity to succeed. He gave me the freedom to explore different approaches and project directions, while not letting me go too far down the rabbit hole and readily gave advice when I was stuck on the project regardless of the day or time. He was also there for other aspects of the degree, from helping tinker with machines when something broke, to giving me the opportunity to travel for my work, which introduced me to a variety of different techniques, researchers in and out of my field, and drastically broadened my world view. I am truly grateful for every opportunity Jan has given me, and particularly for giving me my first chance in research.

During my time in Jan's lab, I had the privilege of working with many amazing colleagues and friends in the Rainey and Langelaan labs, who have been a key part of my success and enjoyment of my PhD. I would like to particularly thank a few people, starting with my first mentor in the Rainey lab, Nathan-Weatherbee Martin, who taught me many of the basic biochemistry techniques I needed in my day-to-day work and made my introduction to graduate level research enjoyable. I was also fortunate to work with Dr. Lingling Xu who patiently taught me molecular biology, and was a calm, constant advisor and friend throughout my degree. Thank you, Dr. Andy Song, for NMR advice and the occasional coffee run to take a break and recharge. Thank you, Bruce Stewart, who gave me the tools to quickly troubleshoot equipment issues and improvise new solutions. I would also like to thank the students I had the privilege to mentor along my journey so far. Specifically, Isaac Bridge, my first student who really helped me grow as a mentor and Skylah McLeod

van Wagoner whose keen attention to detail, scientific, and innovative rigor akin to that of a graduate student allowed her to continue pushing the pyriform project forward when I was out of the lab on an industrial co-op. Lastly, I'd like to thank Tam Pham and Anupauma Ghimire, who helped to make the little moments enjoyable.

I would like to thank those outside of the lab that have given me advice and instrumentation support throughout my degree, including my committee members Dr. Vanya Ewart, Dr. Stephen Bearne, and Dr. John Frampton. Your advice and support along the way have kept me moving in the right direction, and kept my studies focused on achievable goals within the length of my program.

Last but certainly not least, I would like to thank my family and those closest to me, without whom I would not have made it as far in my studies as I did. My late grandmother, Doris, whose constant love and support paved the way during my undergraduate degree to a successful future. I would also like to thank my best friend Will who was there for me through all the ups and downs during the last five years. I would like to thank my partner Kelly, who was an invaluable part of last few months of this degree, and who made the difficulties therein easier to handle. Most importantly, thank you to my parents Scott and Cathy, for constantly being there to guide, love, and support me through my degree, whether it be with a hot meal during busy moments, or phone calls to celebrate the successes and talk about the challenges.

Chapter 1

Introduction

1.1. Thesis Outline

The objective of this work was to develop an understanding of pyriform spider silk proteins in the solution and fibrous states to determine how this class of silk relates to the other six classes of silk spun by orb weaving spiders. Specifically, this work was carried out by recombinantly producing pyriform silk-based proteins used in structural and functional studies. In this chapter, I introduce fundamental knowledge of spider silk function and formation based on major ampullate (MA; dragline) silk, the only extensively characterized type of silk. I then place pyriform silk in context of the other relevant silk classes with what information is currently published, excluding the work summarized in this thesis. After this broad introduction of silk and silk classes, I then discuss the methods initially used and developed to express recombinant pyriform silk proteins, and the resulting mechanical properties that were the first reported for this class of silk (Chapter 2). These expression methods were further refined through studies discussed in Chapter 3, and silks spun from these proteins were characterized. Having optimized the methods of expressing and purifying pyriform-based recombinant proteins and physically characterizing these proteins, I proceeded to analyze the solution-state secondary structure at the atomic level (Chapter 4). In Chapter 5, this information was used to calculate the first tertiary structure of a pyriform silk recombinant protein, which was further analyzed through studies on temperature based chemical shift perturbations and wide-scale changes in secondary structure upon fibre formation. In Chapter 6, I conclude

with a broader examination of what insights can be obtained by combining of all of these findings, and how this combination of structural and mechanical properties clarifies how pyriform fibres compare to other classes of silk.

1.2. Spider Silk

Spiders depend on their silk to meet the most basic functions of their day-to-day lives, from protection from predators and adverse conditions to prey capture. This dependence for survival led to the development of specific, high-performance materials tailored for the various needs of the spider. Specifically, orb weaving spiders produce up to seven distinct classes of silks[1, 2] to meet these needs (Figure 1.1). It should be noted that some species also produce a unique class of silk (cribellate silk)[3] or perform specific modifications to silks to meet particular functional needs[4].

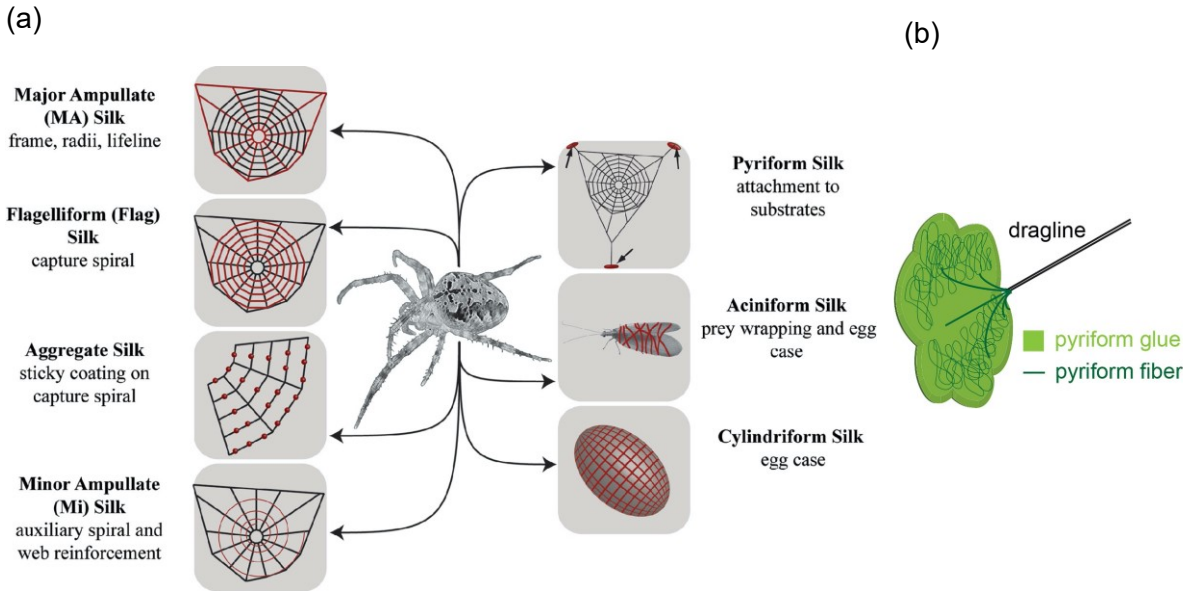


Figure 1.1. The seven main classes of spider silk and their uses by the spider[1] (a) with a schematic of the attachment disc[2] (b). Reprinted by permission from Insect Biochemistry and Molecular Biology[1] and Materials Today[2].

The distinct mechanical properties of spider silks which rival currently used materials in the textile and adhesion industries (Table 1.1) have incentivized the need to understand how silk fibres obtain these mechanical properties for the creation of new and sustainable materials. While class-dependent variations are important in the context of their respective uses and properties[5], this discussion will focus on silk classes with inherent extremes of mechanical properties. This is intended to provide a field-specific context of mechanical and structural properties relative to what was discovered in this work for pyriform silk.

Table 1.1. Mechanical properties of silk classes (grey shading) relative to commonly used materials. Unless otherwise stated, data are from *Argiope argentata*[6].

Material	**True Mechanical Strength (MPa)**	**Extensibility (%)**
Major Ampullate Silk*	1495 ± 65	20.5 ± 0.5
Minor Ampullate Silk*	923 ± 154	33 ± 3.3
Aciniform Silk*	1052 ± 120	40.4 ± 2.4
Flagelliform Silk*	534 ± 40	172 ± 5
Tubuliform Silk*	476 ± 90	28.57 ± 1.5
Drassodex cf. heeri Attachment Disc** (pyriform glue and silk)	511.0 ± 123.6	51 ± 26
Kevlar***	3600	2.7
Wool***	200	50

*****Data published by Blackledge and Hayashi[6]

******Data published by Wolff et al.[7]

***Tabulated by Gosline et al.[8]

1.3. Sequence, Structure, and Mechanical Character

Typically, spider silk proteins consist of a tripartite primary structure composed of a large core domain containing multiple block repeats that account for >80% of the overall protein[9] (Figure 1.2). This is flanked by the non-repetitive N- and C-terminal domains which play key roles in spider silk formation. Based on major ampullate silk, the strongest and most well characterized class of silk, it is reported that assembly of silk fibres into proteins occurs through a lock-and-trigger mechanism, in which an α→β transition occurs in the C-terminal domain, triggering the fibre forming process[10, 11]. This is followed by the formation of stable N-terminal domain homodimers of individual silk proteins[12], locking the structures into place (Figure 1.3).

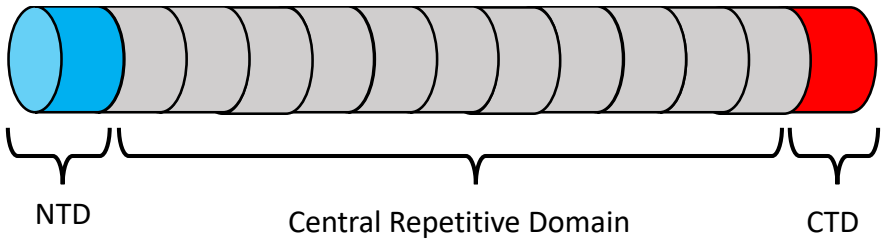


Figure 1.2. Representative primary structure of spider silk proteins. Typical silk proteins consist of a conserved core domain repeated multiple times in tandem, with flanking, non-conserved terminal domains.

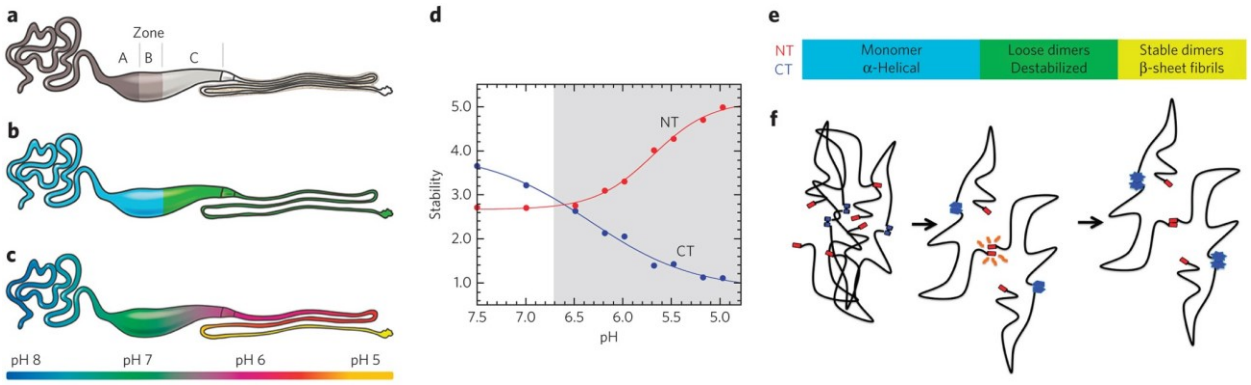


Figure 1.3. Representation of the major ampullate gland and fibrillogenesis process. (a) Major ampullate silk gland with its winding tail (Zone A), central sac (Zones A, B, and C) and spinning duct (post Zone C). (b) Spidroin synthesis takes place in the tail and sac (blue) while carbonic anhydrase responsible for decreasing the gland pH is found in the sac and spinning duct. (c) Schematic of decreasing pH along the silk gland, that lead to changes in thermodynamic stability of N- and C-terminal domains (d), changing N- and C-terminal structural features (e), and the lock-and trigger mechanism of native fibre formation (f). Reprinted by permission from Springer Nature, Nature Chemical Biology, Toward spinning artificial spider silk, Rising A. et al., 2015[9].

While the terminal domains are, in general, highly conserved between species and to some extent between silk classes to properly function during fibrillogenesis[13], central core-domains vary greatly based on the class of silk[5], leading to drastic differences based on the functional role of the silk protein. The class-specificity of these central regions has led to the discovery of alanine- and glycine-rich motifs in the primary sequence that have been linked to mechanical extremes observed in ampullate silks[5], while pyriform silks have been reported to contain proline-rich and serine/glutamine-rich domains[1, 14, 15].

Understanding how these differences lead to different functional capabilities is therefore critical in understanding these classes of silks and applying silk-based materials in other applications.

1.4. The Journey from Gland to Spigot

Distinct differences in silk functions are highly reflected in silk gland morphology[16]. Owing to the high mechanical strength of dragline silk, and the higher body of work surrounding it due to easier natural collection relative to other silk classes[17], current understanding of spider silk formation in the gland is based primarily on this silk. Information can be inferred relative to other silk glands, but differences in morphology and size indicated different mechanisms in different types of glands (Figure 1.4). While this fails to answer many specifics for why drastically different gland size and shapes are required for different silks[16], this provides a basic framework for understanding how silks are formed, with almost all silks losing α-helical character and gaining β-sheet character[18] as the proteins transition from a soluble mixture to a fibrous network (Figure 1.5).

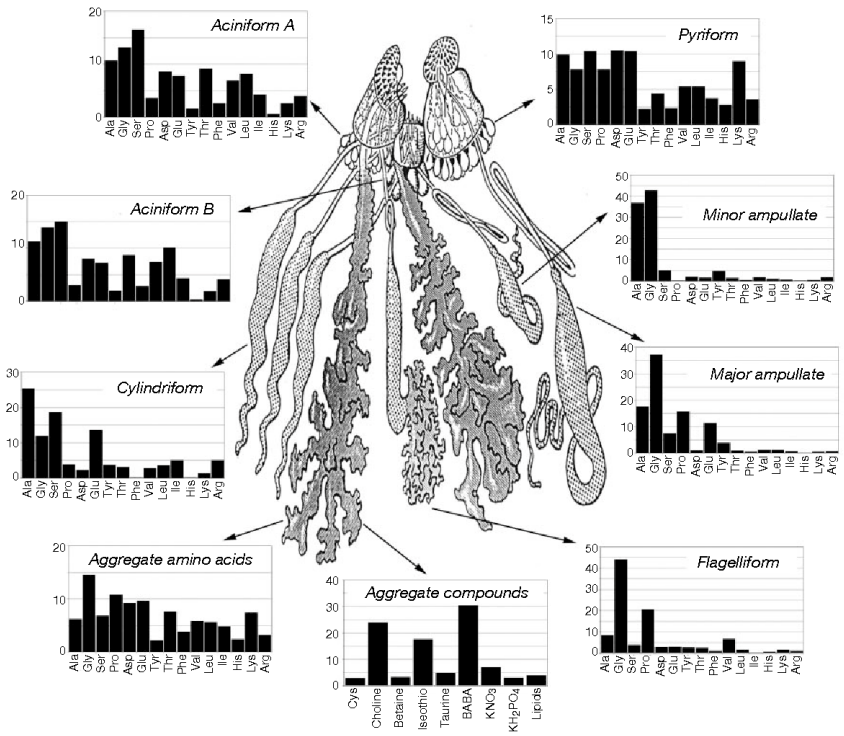


Figure 1.4. Morphology of the glands of all different spider silks and their major constituent composition. Reprinted by permission from Springer Nature, Nature, Liquid crystalline spinning of spider silk, Vollrath F et al, 2001[16].

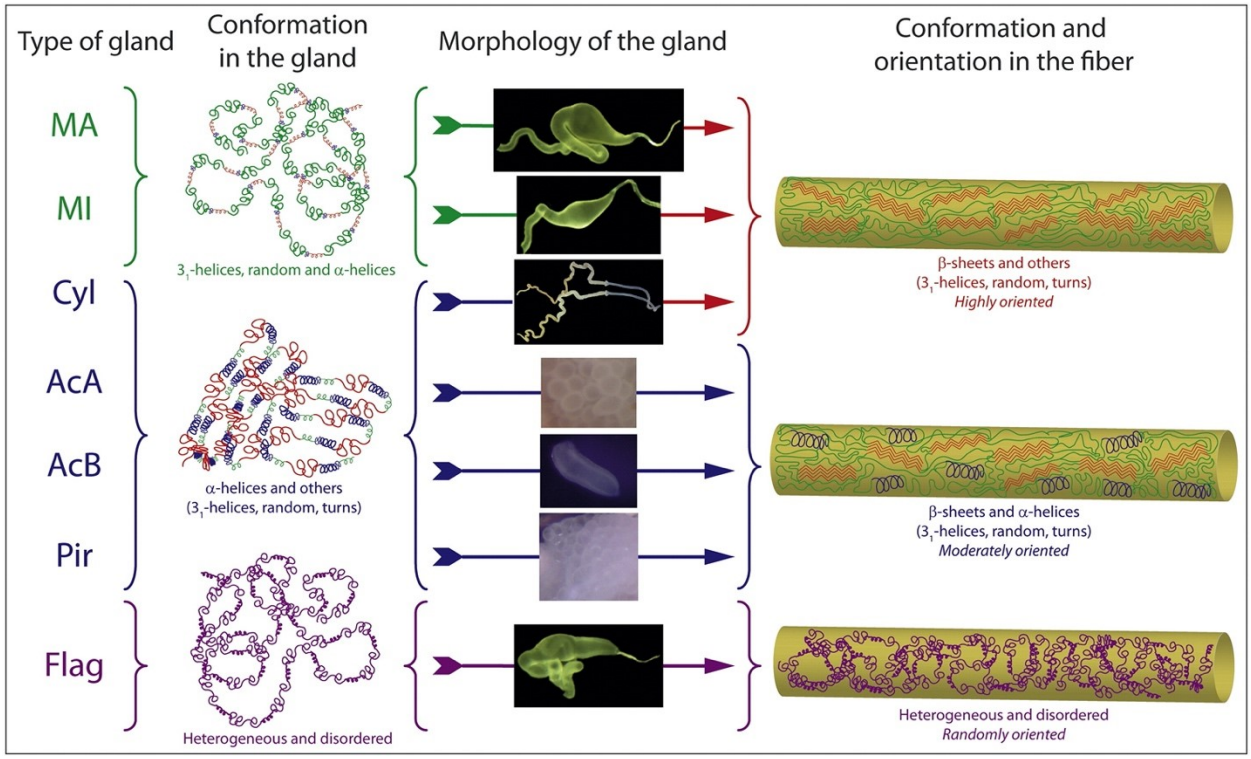


Figure 1.5. Schematic for the soluble, and fibre state protein structures of each class of silk based on *Nephila clavipes*. Reprinted by permission from the Journal of Molecular Biology[18].

As detailed in Figure 1.3, the major ampullate gland can be divided into three distinct regions. Andersson et al.[19] found that most MaSp silk proteins are excreted into zone A, and some proteins are excreted into zone B, leading to a zone A-derived core and a zone B-derived skin. Each of these regions are maintained at a neutral pH, which would aid in preventing unwanted aggregation of silk proteins[11]. Higher levels of carbonic anhydrase in zone C leading to gradually decreasing pH and increasing pCO_2 from the tail to the duct, with contrasting N- and C-terminal effects[11]. Expanding on what is shown in Figure 1.3 and starting with the N-terminal domain, decreasing pH values lead to the formation N-terminal dimers by pH 6, which are stabilized between pH 5-6, locking the proteins into

larger, stable structures[12]. At the same time, the C-terminal domain structure converts from α-helix to β-sheet, triggering amyloid-like fibre formation[10, 11]. As the dope travels through the spinning gland, shear forces are applied to the proteins through contact with the walls of the gland, and narrowing of the gland closer to the spinneret, forming a longitudinal orientation of protein units.

Although the full mechanisms underlying the process of silk fibrillogenesis in nature remain enigmatic, there are two theories of how this solution-to-fibre transition takes place based on distinct proposed packing arrangements of silk proteins: the liquid crystalline theory, and the micellar theory. The liquid crystalline theory[16] of Vollrath and Knight proposes that silk proteins stack together in hexagonal columnar liquid crystalline-like structures, then are oriented longitudinally to the forming fibre by shear forces along the gland[20]. In contrast, Jin and Kaplan's micellar theory[21] proposes that silk proteins form micelles that are elongated due to shear forces. Figure 1.6 shows a summary of the two theories.

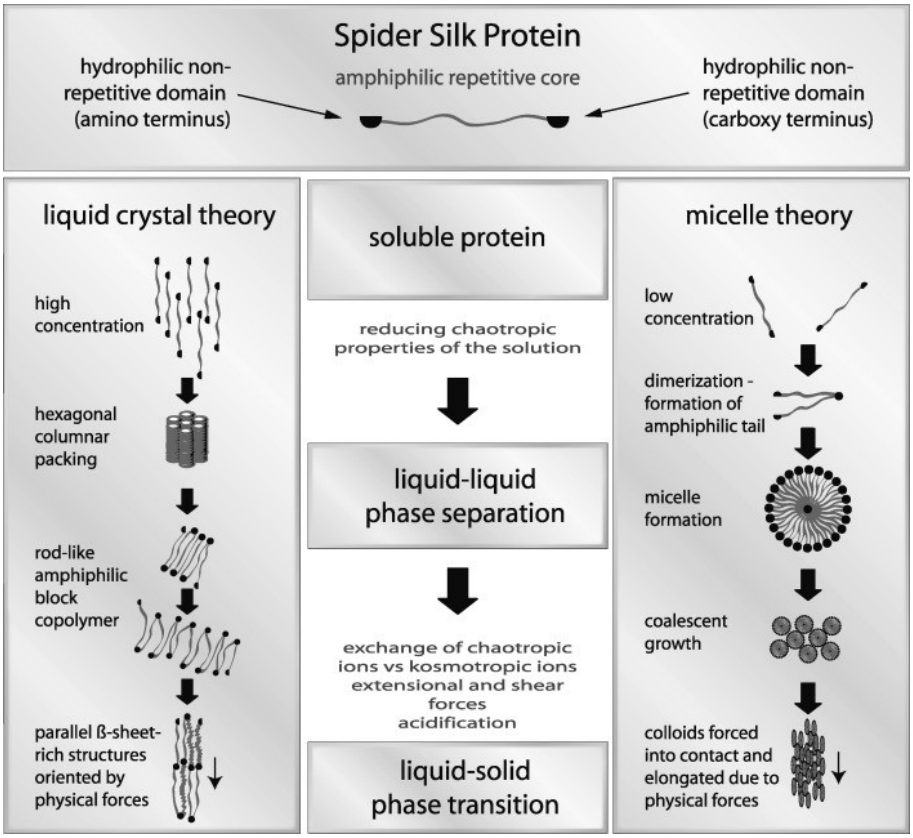


Figure 1.6. Summary of the liquid crystal[16] and micelle[21] theories of spider silk formation. Reprinted by permission from Angewandte Chemie International Edition[20].

Further investigation into the formation of spider silks has led to a modification of the micellar theory[22]. Namely, it was proposed that silk proteins, when stored in the gland, assemble into flake-like subdomains that further assemble into micelle-like structures. Under this theory, upon exposure to shear forces, individual subdomains form elongated fibrils remain in the overall micellar-like assembly. Developing an understanding of the fibre forming process is still an active area of research and will give key insight into how silk proteins assemble into fibres on the microstructure level.

1.5. Properties of Pyriform and Related Spider Silks

While there are seven classes of silk in nature, I will be focusing this discussion on three relatively poorly characterized silks with potential links to PySp: flagelliform silk, aggregate silk, and aciniform silk. This is followed by an extensive review on the target of my work PySp.

1.5.1. Flagelliform Silk: The Most Extensible Silk

Forming the netting of the spider web, flagelliform silk functions as a highly extensible framework to prevent web damage from insect collisions and outdoor weather conditions. With its intrinsic net-like function, it is not surprising that this is the most extensible class of silk (Table 1.1). In lieu of the poly-alanine motifs of major ampullate silks, flagelliform silk has a much higher abundance of GPGGX motifs, which have been proposed to yield high extensibility through the formation of β-spirals[23]. To function in its required role, flagelliform silk is aided by glue-like droplets secreted along the threads called aggregate silk. Aggregate silk is a highly complex combination of lipids[24], small polar aliphatic compounds[25], and other chemical components[26], with the exact combination and degree of variability between species' remaining unknown. In studies of flagelliform silk with or without aggregate secretions, it was shown that the presence of aggregate secretions leads to a plasticization of the flagelliform fibres[27]. This type of interaction could, in theory, provide some insight into different features of the other composite pyriform silk-glue system used in spider webs.

1.5.2. Aciniform Silk: The Toughest Silk

Aciniform, or wrapping, silk is the toughest of all silk classes, and the aciniform protein AcSp1 contains large central repetitive domain units relative to other classes of silk (200 amino acids for each block repeat in *Argiope trifasciata*, for example). Unlike previously mentioned silk proteins, AcSp1 lacks the general alanine-rich and glycine-rich motifs found to give other spider silks high mechanical strength and extensibility, respectively[28]. Structural studies of the central repetitive domain have shown variations in the central block repeat solution-state structure, with AcSp1 from *Argiope trifasciata* organized into five helical regions with short linker regions. In this silk, the 5th helix is the first to unfold[29, 30] and most dynamic on the ps-ns timescale[31], implying a key role in the fibre forming process[30]. Contrarily, the AcSp1 central block repeat from *Nephila antipodiana* was found to be a tightly clustered 6-helix bundle[32], indicating potential differences in structure between silks of different species. Interestingly, the aciniform gland is similar in size and shape to the pyriform gland, suggesting some similarities in silk formation despite having distinct differences in silk application and composition[16].

1.5.3. Pyriform Glue Silk: The Enigmatic Silk

Pyriform (piriform) silk is the least investigated of the seven silk classes. This silk is utilized in composite materials called attachment discs[33], in which a pyriform microfibril mesh is embedded in pyriform glue proteins. Prior to the 1970s, it was thought that attachment discs were only composed of a glue material as opposed to the matrix of glue and silk that is now known. The glue and silk components function to attach different silks together, to attach spider webs to objects, and in prey capture (Figure 1.7)[34]. Early studies of the

gland and silk suggested that pyriform glue is coated onto pyriform fibres prior to extrusion from the gland and, potentially, prior to full assembly of the pyriform proteins into silk[35, 36].

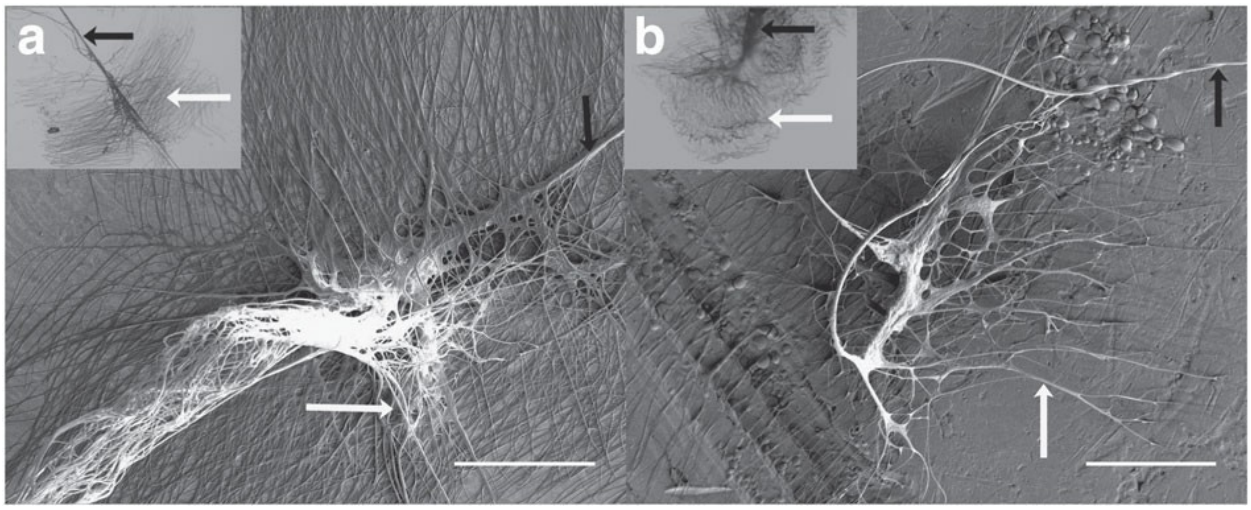


Figure 1.7. SEM images of *A. tepidariorum* scaffolding discs (a) and gumfoot discs (b). White arrows indicate pyriform fibres, while black arrows indicate major ampullate fibres. Reprinted by permission from Springer Nature, Nature Communications, Cobweb-weaving spiders produce different attachment discs for locomotion and prey capture, Sahni V. et al., 2012[34].

Like the aggregate glue, there is little known about the makeup of the glue component, but a hierarchical ordering with silk proteins has been stipulated in the attachment discs[37]. These networks of glue and fibres are postulated to interact through polar side chains to give cohesion in the attachment discs and adhesion to substrates[37]. Likely helping aid these interactions, it is proposed that pyriform glue is composed of one or more proteins rich in acidic amino acids[35], a carbohydrate component and, potentially, lipids[37].

While pyriform silk as a class is still highly enigmatic, an understanding of this silk class has started to be developed in recent years. So far, there have been two sub-classes of pyriform silk proteins identified, PySp1[1, 15] and PySp2[14], each being classified currently on the basis of a small number of partial or full-length sequences (Figure 1.7). PySp1 of *Argiope argentata*, the silk protein this work is based on, is a protein containing many long (>200 amino acid) block repeats, non-repetitive N- and C-terminal linkers, a long N-terminal domain, and a shorter C-terminal domain (Figure 1.8). Mechanically, what is known for full length silks was only tested in the context of the attachment disc[7, 38], and hence is representative of the complex attachment disc context, but not of the pyriform silk component itself. Furthermore, there is a lack of sequence information for the full-length pyriform silk. At the time this work was initiated, only one full-length pyriform gene had been sequenced (Figure 1.9)[1], while a second full-length sequence[15] and the first full-length PySp2 sequence[14] were recently reported (Figure 1.10). Although full sequence comparisons are limited to these three protein sequences, this sequence information revealed a distinct amino acid composition of the central core-domain relative to all other silk classes, and a lack of the alanine- and glycine-rich motifs. Instead, pyriform silk proteins have a PX-rich motif at the C-terminal end of each block repeat, which is reported to be a common finding in pyriform silks of orb weaving spiders[1], and a repetitive glutamine and serine rich motif near the N-terminal end (Figure 1.11). The function of these distinct motifs remain unknown, but understanding their structure could provide key insight into how these fibres are functioning in their spider web applications.

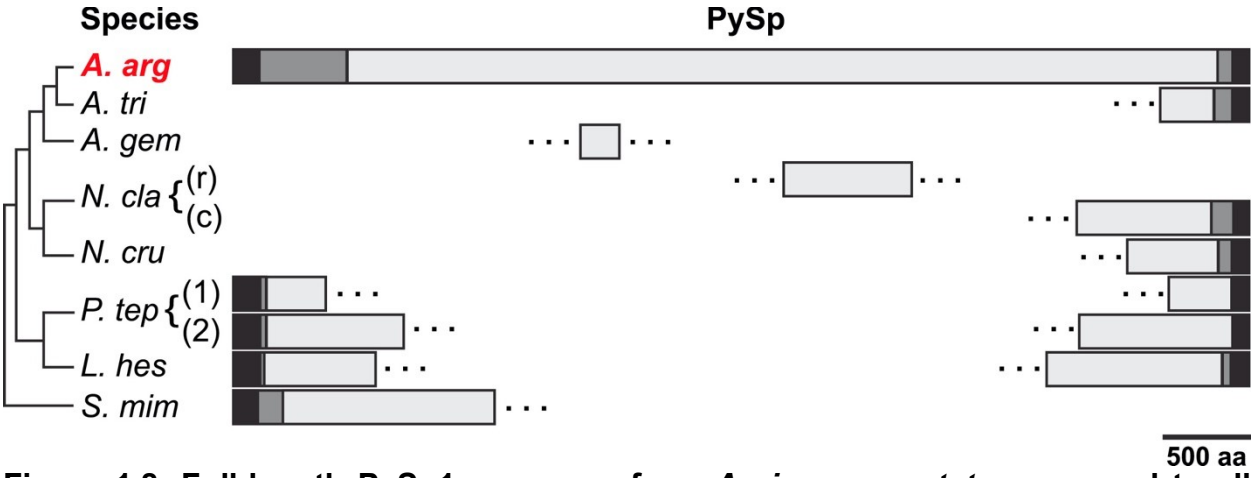
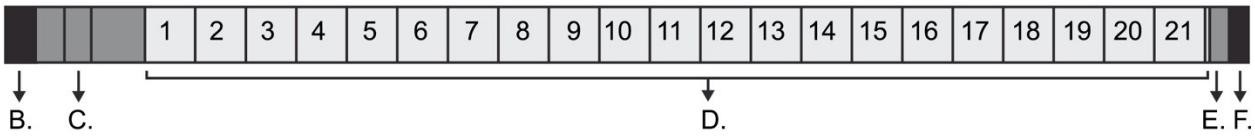


Figure 1.8. Full-length PySp1 sequence from *Argiope argentata* compared to all other PySp partial sequences. Reprinted by permission from Insect Biochemistry and Molecular Biology[1].

A. Full length *Argiope argentata* pyriform spidroin (PySp1; 5,759 aa)



B. N-terminal region

MSWLSVLSLCVVLFASNADALSLAKSWVQDEGTAIYFLNRAIEYLRECGVLKIDQERDAIETM
FETSSLYQKNAKPSKIKHSVASKLAEIIIEGLEGASDTSYKMDCVSQAIASASEATTGTVDDN
FIESVQELVVVMYNNDIEDK...

C. N-linker region, repetitive section

...TSIAQQQYEASQQQSSVSQQQYEASQQQSSGSQQQYEASQQQSSVSQQQYEASQLQASAAQQ
QYEASQQQASIAQQQYEATQQQASIAQQQYEASQQQASIAQQQYEATQQQ...

D. Repetitive region, consensus of 21 complete repeats

QSSVAQQSAVAQQSSVSQQSSAAQQSSVAQSQQTSYSAATNAGSSVSQSQAIVSSAPVYFNSQ
TLTNNLASSLQSLNALNYVSNGQLSSSDVASTVARAVAQSLGLSQGSVQNIMSQQLSSIGSGA
STSSLSQAIANAVSSAVQGSQAAAPGQEQSIAQRVNSAISSAFAQLISQRTAPAPAPRPRPAP
LPAPAPRPRPAPAPRPAPVYAPAPVASQFQASASSQSSAQENSFT

E. C-linker region

...EGLYSSQSQNLQSANVLSSSQVQQVSSGSSSDFGSSSTVFAAPSSSAFYSPAAISSSASVGA
SSISMASSSGFSTSALPTSGSPLAASAA...

F. C-terminal region

...QQLLSPAANQRITALSNSLKSTFSGGQVNYGALSNSLASAASQIQSSSGLSKNEVLVEVLLE
TLAALLDSLSVSGSSSAQFAQAVVQAFA

Figure 1.9. Diagram of the different domains found in the full-length *A. argentata* PySp1 sequence. Reprinted by permission from Insect Biochemistry and Molecular Biology[1].

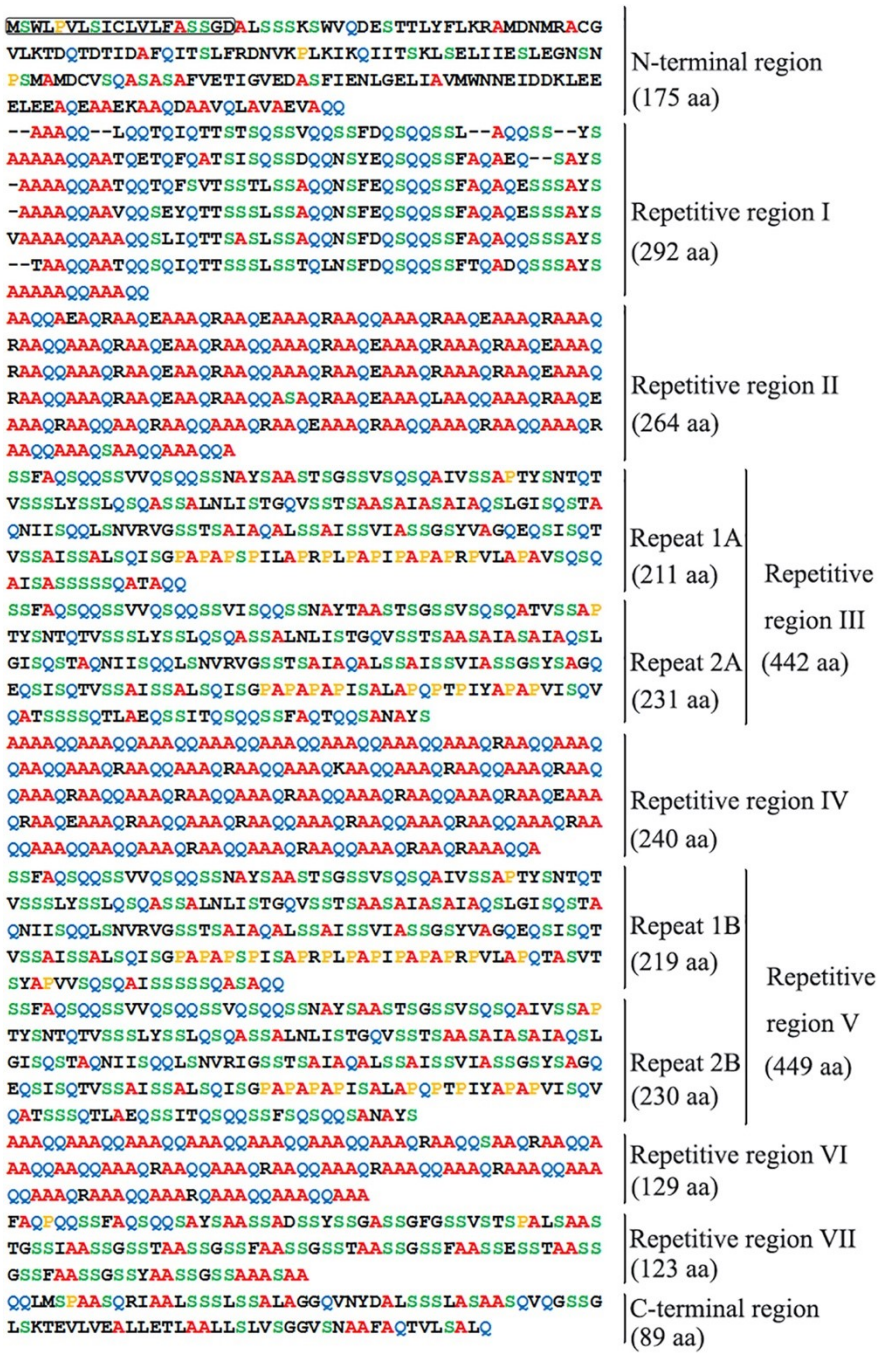


Figure 1.10. Diagram of the different domains found in the full-length *Araneus ventricosus* PySp2 sequence. Reprinted by permission from Insect Biochemistry and Molecular Biology[14].

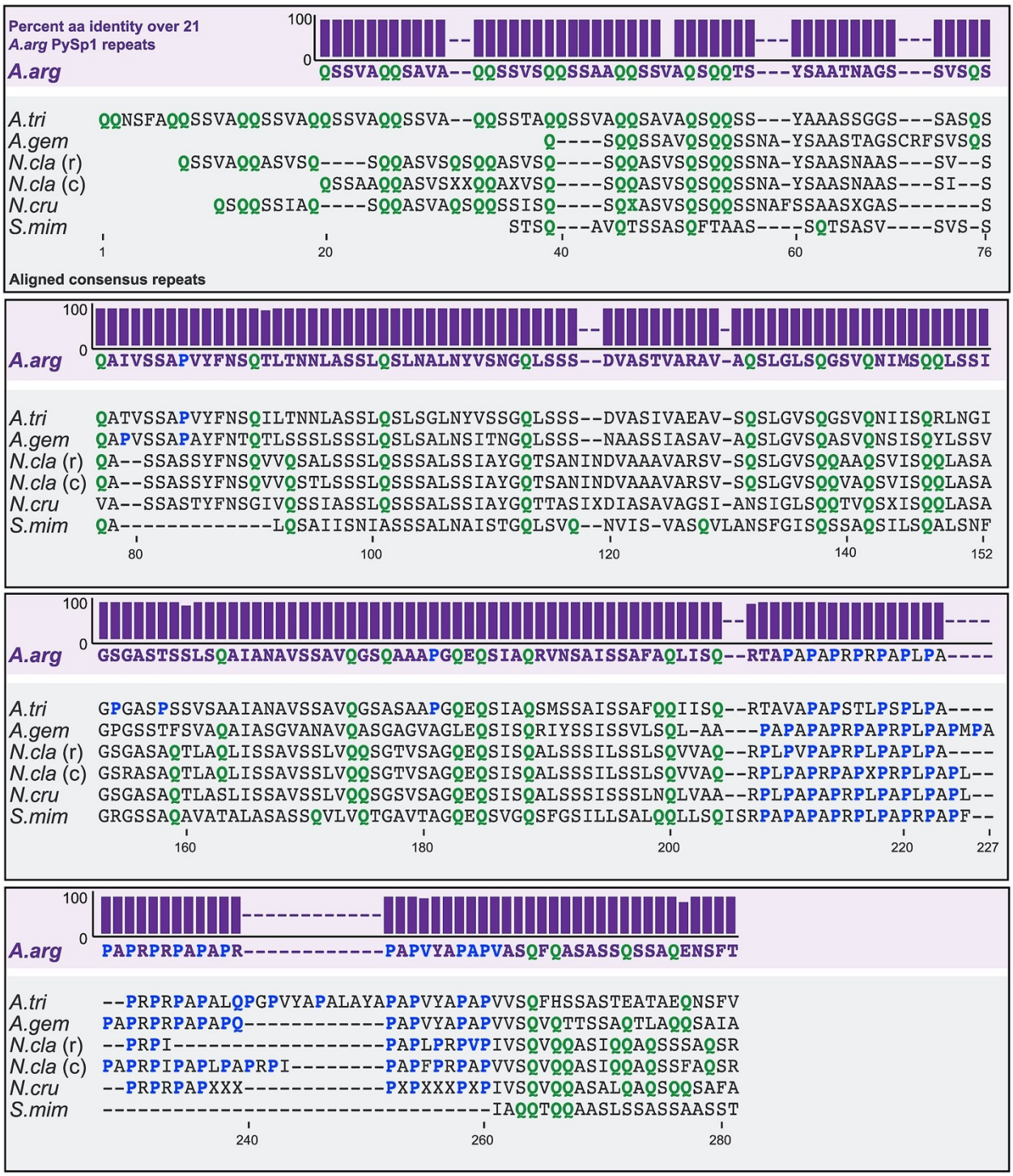


Figure 1.11. Alignment of PySp conserved repeats from different species. Reprinted by permission from Insect Biochemistry and Molecular Biology[1].

1.6. Recombinant Spider Silk Production

While native silk studies are important for understanding the mechanical and structural properties of silks spun from full-length proteins, work on native silks is limited by difficulties in collection and isolation of silks, efficiency of protein isotopic enrichment for atomic-level structural studies by NMR spectroscopy, and inability to probe specific regions in the whole protein. To solve these challenges, recombinant silk expression can be utilized, and is the focus of this thesis work. For this type of work, the most commonly used expression systems are those based on *Escherichia coli*, due to the potential for high protein yield and the relative simplicity of expression and purification systems. Precedents on recombinant pyriform prior to this work[39] and confirmations after this work[40] showed this system is amenable to recombinant pyriform silk production. This provides a way to express targeted versions of silk proteins, and potential to express truncations[41] and variations[42] of silk proteins for interrogation or enhancement of mechanical properties. These protein segments, whether it be the central repetitive domain which is the focus of this thesis, or a non-repetitive terminal domain, can be used to create the simplest models to evaluate the structural and mechanical contributions of different regions of a given protein, and hence its utilization in this work.

1.7. Atomic-Level Solution State Structures of Spider Silk Repetitive Domain Proteins

Currently, only a small number of solution state structures have been reported for spider silk repetitive domains. These include the repetitive domains from AcSp1 of *Argiope trifasciata*[29] (10.2210/pdb2MU3/pdb), and, from *Trichonephila antipodiana*, AcSp1[32], both

TuSp1[43] repeats one and two, and a deposited but non-published MiSp1 structure from *Trichonephila antipodiana* (10.2210/pdb2MQA/pdb). These proteins have all been shown to comprise ordered cores of 5-6 parallel α-helices, either tightly clustered [AcSp1 from *T. antipodiana*[32] (10.2210/pdb2LYI/pdb) and TuSp1 units from *T. antipodiana*[43] (10.2210/pdb2K3N/pdb;10.2210/pdb2K3O/pdb), and MaSp1 from *T. antipodiana*)], or somewhat loosely packed (AcSp1 from *A. trifasciata*[29]). Each of the tightly packed proteins share a similar structural arrangement, with most having a short helix bridging over the bundle (Figure 1.12). It should be noted that the underlying study reporting the solving of the MiSp1 repeat unit protein has yet to be published and the reported PDB validation metrics are low relative to other structures, making its robustness difficult to assess. This information indicates a potential structural similarity in the repetitive domains of soluble silk proteins; however, the two aciniform silk proteins show very different repetitive domain structure (noting that they only share 22% pairwise sequence identity[29]), which means more information is needed before this hypothesis can be fully tested. As such, solving the structures of additional central repetitive domains is essential to provide insight into whether there is a general architecture for packing of soluble spider silk proteins, or whether there is greater structural diversity than is currently represented by the handful of determined structures, especially keeping in mind the already shown existence of exceptions.

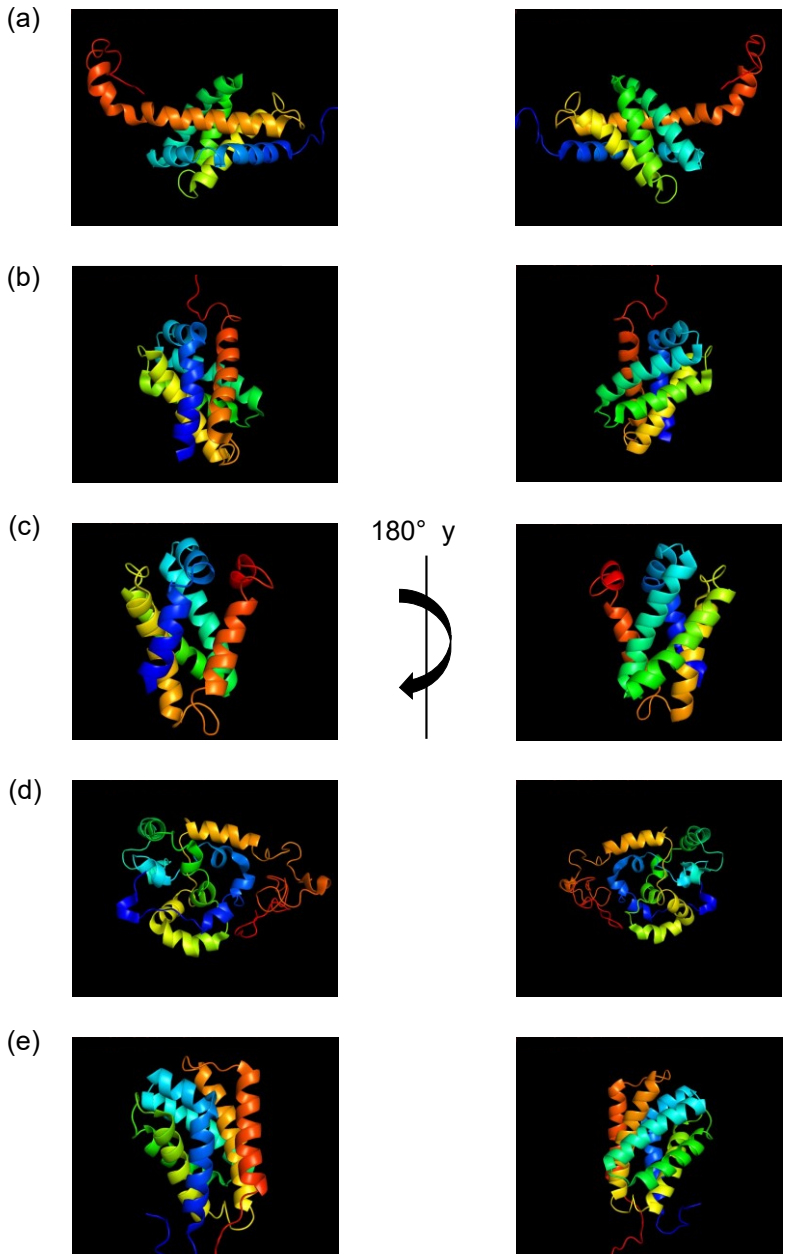


Figure 1.12. Lowest energy structures from repetitive domain ensembles. Shown are the solved structures for TuSp1-RP1[43] (a; PDB DOI: 10.2210/pdb2K3N/pdb), TuSp1-RP2 (b; PDB DOI:10.2210/pdb2K3O/pdb), MiSp1 (c; PDB DOI: 10.2210/pdb2MQA/pdb), and AcSp1 from *A. trifasciata*[29] (d; PDB DOI: 10.2210/pdb2MU3/pdb) and *N. antipodiana*[32] (e; PDB DOI: 10.2210/pdb2LYI/pdb). The structures are color coded going from blue at the N-terminus to red at the C-terminus.

1.8. Rationale and Objectives

As detailed above, spider silks are unique biomaterials with incredible mechanical properties. This makes them highly desirable targets for a diverse array of applications, including textiles, body armor, and biomedical applications. To understand how these biomaterials can be applied to solve every-day issues, a complete characterization is needed of the properties that are inherent to each class of spider silk, and the molecular- and supramolecular-level features that are critical for those properties to be achieved and maintained.

The first objective of this work was to develop a method for producing and spinning recombinant pyriform silk proteins to enable mechanical, morphological, and structural analyses (Chapters 2 and 3). This initially used well-established methods for designing, expressing, purifying, and wet-spinning recombinant silk fibres[39, 44], with the requirement of key modifications for my recombinant pyriform silks. These methods were further refined in Chapter 3 to increase expression yields. The second objective was to analyze recombinant pyriform silk mechanical properties (Chapters 2 and 3) and develop an understanding of how this type of silk would behave when decoupled from the attachment disc. The third objective was to obtain an atomic-level solution-state structural assignment (Chapter 4) to interrogate the fibre forming process of pyriform silk and which regions are critical for this (Chapter 5).

Chapter 2

Design, Expression, Purification and Mechanical Characterization of HPy_2, a Recombinant Pyriform Silk Protein

Methods, results, and discussion are based on work presented in the following publication: Simmons, J. R., Xu, L., & Rainey, J. K. (2019). Recombinant pyriform silk fibre mechanics are modulated by wet-spinning conditions. *ACS Biomaterials Science & Engineering*, *5*(10), 4985-4993. All the presented results are my work, the first draft of this manuscript was written by me, and any contributions by coauthor Lingling Xu are specifically noted in the methods.

2.1. Introduction

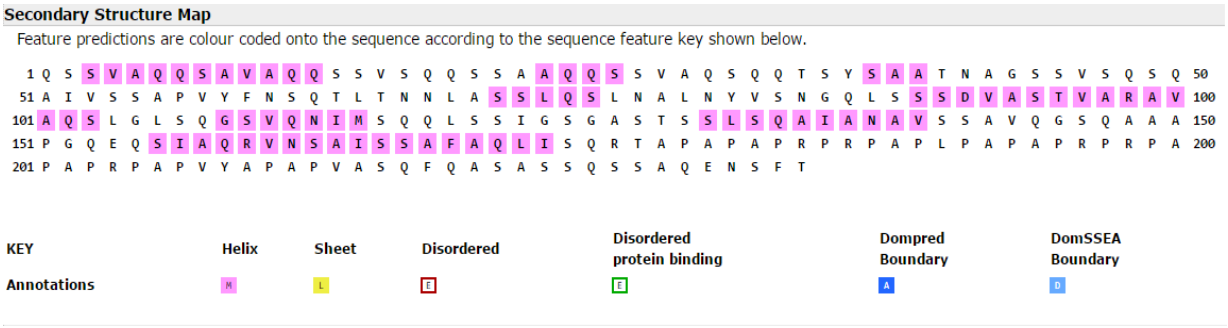
In Chapter 1, I provided a brief overview of what is known about spider silk production and variation based on class of silk. From this discussion, it is clear there are significant gaps in knowledge of structure, mechanical properties, and the relation of the molecular sequences for pyriform spider silks to other silk classes. In this chapter, I start to fill that gap by detailing methods developed to express, purify, and produce silk-like fibres from a recombinant pyriform silk-based protein, HPy_2. This protein construct was used to establish the first mechanical and structural characterization of a recombinant pyriform silk-based protein.

2.2. Materials and Methods

2.2.1. Designing Plasmids for Recombinant Pyriform Silk Expression

A plasmid with an open reading frame encoding one repeat of the central repetitive domain as described in Figure 2.1 was designed for *Escherichia coli* expression. Plasmid design was carried out by Lingling Xu, based upon my input through protein sequence-based secondary structure predictions. Rather than using the EMBOSS/GOR-based secondary structure predictions[45] used to divide regions by Chaw et al.[1], we chose positions with considerations of structural prediction (Figure 2.1), cloning[44], and potential for application in intein *trans* splicing[29, 46]. The one repeat encoding plasmid was ordered in a pUC57 cloning vector (BioBasic Canada Inc.; Markham, ON) with the sequence ggaggagctcttcaAGC before the protein-coding sequence of interest and the sequence AGCagaagagctctgagctgcac after the protein-coding sequence of interest. These sequences encoded a BseRI digestion site before the coding sequence and a BsgI cut site after the coding sequence that, when used with the AatII cut site on the pUC57 vector, allow for seamless ligation of DNA sections at the codons encoding the terminal serine residues. In addition, these sequences encode two SapI cut sites flanking the coding sequence of interest that allows for double digestion of the plasmid, then insertion into a modified pET32 expression vector[44] (Novagen; Darmstadt, Germany) downstream of a His_6 tag with or without a SUMO tag. For optimal transfer to the cloning plasmid, pET32 was treated with Antarctic Phosphatase (37 °C, 30 min, using NEB protocol) to prevent self-ligation.

(a)



(b)

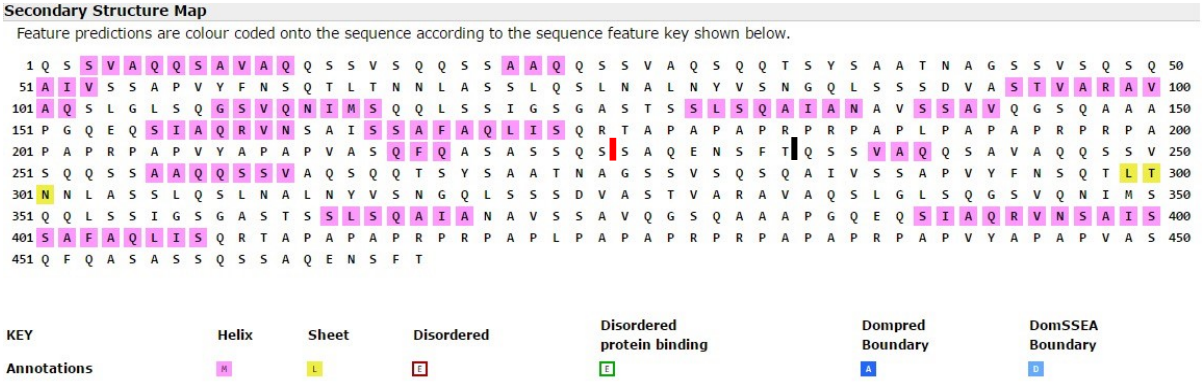


Figure 2.1. PsiPred secondary structural predictions of divisions as reported by Chaw et al.[1] for one block repeat (a) and two-block repeats (b). Divisions as reported by Chaw et al. are shown by black bars, while divisions chosen based on in-house considerations are shown in red. An extra serine was added to the N-terminus of the first block repeat to allow for seamless ligation of pyriform repetitive units.

For all work going forward, N- and C-terminal non-repetitive regions and linkers were not included in the plasmid design following precedents of successfully wet-spun recombinant pyriform[39] and aciniform[47] fibres without inclusion of non-repetitive regions. Focusing on the major constituent of pyriform silk creates the simplest model to understand this recombinant protein while avoiding confounding effects arising from currently unknown

properties of other regions in the full-length pyriform silk protein. For the work presented in this chapter, a double-repeat (Py_2) encoding open reading frame was constructed by restriction enzyme digestion and ligation of two Py_1 sequences as described above. This sequence was then inserted into a pET32 vector[44] containing an N-terminal His_6-tag encoding nucleotide sequence via SapI digestion for protein expression, and the resulting gene and protein are referred to as HPy_2 (Figure 2.2). Identities of all plasmids modified or cloned were confirmed via Sanger sequencing (Genewiz; NJ, USA).

```
  1  MGHHHHHHSS AQENSFTQSS VAQQSAVAQQ SSVSQQSSAA QQSSVAQSQQ  50
 51  TSYSAATNAG SSVSQSQAIV SSAPVYFNSQ TLTNNLASSL QSLNALNYVS  100
101  NGQLSSSDVA STVARAVAQS LGLSQGSVQN IMSQQLSSIG SGASTSSLSQ  150
151  AIANAVSSAV QGSQAAAPGQ EQSIAQRVNS AISSAFAQLI SQRTAPAPAP  200
201  RPRPAPLPAP APRPRPAPAP RPAPVYAPAP VASQFQASAS SQSSAQENSF  250
251  TQSSVAQQSA VAQQSSVSQQ SSAAQQSSVA QSQQTSYSAA TNAGSSVSQS  300
301  QAIVSSAPVY FNSQTLTNNL ASSLQSLNAL NYVSNGQLSS SDVASTVARA  350
351  VAQSLGLSQG SVQNIMSQQL SSIGSGASTS SLSQAIANAV SSAVQGSQAA  400
401  APGQEQSIAQ RVNSAISSAF AQLISQRTAP APAPRPRPAP LPAPAPRPRP  450
451  APAPRPAPVY APAPVASQFQ ASASSQS                           476
```

Figure 2.2. Amino acid sequence of HPy_2. Pink residues indicate the His_6-tag, Red residues indicates repeat 1, blue residues indicates repeat 2, bold residues indicate QSXXXQ motifs, and underlined indicate XP rich motifs. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

2.2.2. Protein Expression and Purification

Large scale expression of recombinant HPy_2 was performed in *E. coli* BL21(DE3) cells based on previously reported methods[44]. Briefly, a starter culture was transferred at ~1:50 v/v to fresh LB medium and grown at 37 °C until an OD_{600} ~0.8 was reached.

Overexpression of HPy_2 was induced by adding 0.64 mM isopropyl β-D-1-thiogalactopyranoside (IPTG; Bio Basic Canada Inc.; Markham, ON), followed by 16 h incubation with shaking at 22.5 °C. Cells were harvested, resuspended in native lysis buffer (QIAexpressionist; Siegen, Hilden, Germany) at 5 mL/g wet cell weight (wcw), and lysed using a French Pressure Cell (American Instrument Company; Silver Spring, MD). The cell lysate was centrifuged at 12 000 rcf (4 °C) for 30 min to pellet insoluble components (including HPy_2 protein).

Two methods were investigated to purify the inclusion body: Ni^{2+}-NTA IMAC (hereafter referred to as Ni^{2+}-IMAC) and inclusion body washes. For Ni^{2+}-IMAC purification, inclusion bodies were resuspended in an 8 M urea buffer (8 M urea, 100 mM NaH_2PO_4 10 mM tris base, pH 8.00 ± 0.05), then diluted to obtain 4 M urea (otherwise identical buffer). Solubilized proteins were purified via Ni^{2+}-IMAC (2-4 columns, ~2.5-3 mL dry bead volume per column) using a wash step with 30 mL of wash buffer per column (4 M urea, 100 mM NaH_2PO_4 10 mM Tris·Cl, 10 mM imidazole, pH 8.00 ± 0.05) followed by 16 mL of elution buffer per column (4 M urea, 100 mM NaH_2PO_4 10 mM Tris·Cl, 250 mM imidazole, pH 8.00 ± 0.05). For inclusion body washes, the pelleted inclusion body was resuspended with 10 mL/g wcw of the chosen inclusion body wash buffer [50 mM potassium phosphate, pH 8.00 ± 0.05 1% Triton X-100 (Fluka; Sigma-Aldrich, St. Louis, MO) and 2 M urea (Sigma-Aldrich)] using a tissue grinder. The solution was allowed to incubate for 1 h at room temperature, then centrifuged (12 000 rcf, 30 min, 4 °C) and the supernatant was discarded. This was repeated, followed by a 10 mL/g wcw water resuspension to remove residual Triton-X (10 min, room temperature) and centrifugation (12 000 rcf, 30 min, 4 °C) to collect the insoluble protein. The purified inclusion body was

resuspended in 5 mL/g wcw of buffer (100 mM sodium phosphate, 10 mM Tris·Cl, 8 M urea, pH 8.00 ± 0.05) and centrifuged (12 000 rcf, 30 min, 4 °C) to solubilize the target HPy_2 protein. The concentration of urea was reduced by dialysis against 50 mM sodium phosphate buffers containing 4 M and 2 M urea (pH 8.00 ± 0.05) to prevent aggregation due to improper refolding, followed by dialysis against water to remove any remaining salt. The protein in water was flash-frozen with liquid nitrogen, then lyophilized (aapptec Sharp Freeze -110, Louisville, KY) for storage as a dry protein powder.

2.2.3. Mass Spectrometry Analysis

Mass spectrometry experiments and data deconvolution were performed by Dr. Alejandro Cohen (Proteomics and Mass Spectrometry Core Facility, Dalhousie University). The identity of HPy_2 was confirmed through two mass spectrometric methods: intact protein analysis by ESI+ with ion trap mass spectrometry analysis and ESI+ on in-gel chymotryptically digested HPy_2. Intact protein analysis was conducted on a sample prepared in potassium phosphate buffer and analyzed by liquid chromatography (LC)-MS/MS using a VelosPRO orbitrap mass spectrometer with an UltiMate 3000 Nano-LC system (ThermoFisher Scientific; Waltham, MA). In-gel digestion was carried out on the most prominent protein band in a Coomassie stained SDS-PAGE, which was believed to correspond to the target protein. The protein was digested for fragment analysis using sequencing grade chymotrypsin (Roche Diagnostics GmbH; Mannheim, Germany) as described by Shevchenko et al.[49]. Results were analyzed using Proteome Discoverer 2.2 (ThermoFisher Scientific) and searched against a variety of fields, including the full-length PySp1 sequence this work was based on[1], *E. coli* proteins, a provided fasta file

corresponding to the amino acid sequence of HPy_2, and the cRAP database of common contaminants (Global Proteome Machine Organization).

2.2.4. Spinning Dope Preparation and Characterization

While HFIP is the most common solvent for solubilization of recombinant spider silk proteins for fibre spinning[50], based on in-house results[47], recombinant protein was solubilized in a 3:1:1 v:v:v mixture (dope solvent) of trifluoroacetic acid (TFA; EMD Millipore Corporation; Darmstadt, Germany), 2,2,2-trifluoroethanol (TFE; Alfa Aesar; Heysham, UK), and water. An appropriate weight per volume ratio was determined qualitatively based on viscosity through small-scale (≤10 µL) tests. For an optimal ratio, the solution needed to appear more viscous than the solvent alone, which was used as an indication of protein-protein interactions in the dope and amenability to wet-spinning, but not so viscous that it would be difficult to extrude through a needle. Protein at the optimal ratio (10% w/v) was incubated in the dope solvent for 30 min at room temperature with intermittent vortexing to allow solubilization. The mixture was centrifuged (20,817 rcf, 30 min, 10 °C) to remove any insoluble particulates, with the soluble portion retained as the spinning dope. To test for protein integrity upon incubation in the acidic dope solvent, samples were prepared for SDS-PAGE analysis before and after each experiment. Dope viscosities were measured at room temperature as previously described[51] using a *micro*Visc HVROC-L portable viscometer equipped with an A05 50 µm flow channel *micro*Visc chip (Rheosense Inc.; San Ramon, CA). Before measurements, the microviscometer was calibrated using deionized water and isopropyl alcohol to ensure no

traces of previously run compounds were evident in the *micro*Visc chip, and the instrument was pre-equilibrated with dope solvent.

2.2.5. Circular Dichroism (CD) Spectroscopy

Two solution-state conditions were chosen to examine the bulk secondary structure of HPy_2: 1% w/v in the dope solvent (~209 μM) and a dilute concentration (6.52 μM) in potassium phosphate buffer (pH 7.50 ± 0.05). The sample in dope solvent was prepared as described in section 2.2.4. The sample in phosphate buffer was prepared from lyophilized protein resuspended in denaturing buffer (8 M urea, 10 mM Tris·Cl, 100 mM sodium phosphate, pH 8.00 ± 0.05) then dialyzed against a decreasing urea gradient (5.1 M, 3.4 M, 1.7 M, 0.8 M, pH 8.00 ± 0.05) to gradually refold the protein before final dialysis against a ~15 mM potassium phosphate buffer (pH 7.50±0.05). Far-UV CD spectra were collected on an Olis DSM20 Circular Spectrophotometer (Bogart, GA) at 22.5 °C ± 0.05 °C with integration time determined as a function of High Volts in Olis SpectralWorks Version 5.888.272. CD spectra were acquired from 250-180 nm with a 1 nm step size using quartz cuvettes of 0.5 mm path length for buffered samples and 0.01 mm path length for the 1% spinning dope (Hellma Canada Limited; Concord, ON). Protein concentrations were determined at 214 nm[52] using the Beer-Lambert law, with blank subtraction performed prior to mean residue ellipticity ([θ]) calculation as the average of two repetitions.

2.2.6. Recombinant Pyriform Fibre Production

HPy_2 resuspended at 10% w/v in 3:1:1 v:v:v TFA:TFE:H_2O was wet-spun by extruding at a rate of 600 μL/h (spinneret inner diameter pf 0.127 mm) into a 90% EtOH coagulation

bath. The coagulation bath served to remove the solvent and promote fibre self-assembly, as previously described[47, 50]. Fibres were spun at 28.2 ± 2.4% humidity and 28.0 ± 2.2 °C (average ± standard deviation), with room conditions carefully monitored and measurements recorded directly after collection. Fibres formed in the coagulation bath were guided by tweezers onto a home-built semi-automated spinning apparatus (Figure 2.3). Fibres were collected without mechanical processing [AS, Figure 2.3(a)] on roller 1 as a baseline measurement for the recombinant silk, with collection speed matching the dope extrusion rate into the coagulation bath. To promote higher mechanical properties based on literature precedents, a targeted set of three post-spin (PS) processing methods were used to examine the effect of stretching and solvent exposure, the latter optimized by stretching AS fibres in a petri dish with the solvent of choice. For a 2× post-spin stretch in air (2× air)[53], AS fibres were guided onto roller 2, rotating at 2× the speed of roller 1 [Figure 2.3(b)]. For a 2× post-spin stretch in air following immersion in 40% EtOH (2× after EtOH), the fibre was guided from roller 1 onto roller 3 (speed equal to roller 1 and emersed in 40% EtOH) and then collected on roller 2 in air [speed 2× that of roller 1 and 3; Figure 2.3(c)]. This condition was to test the effect of solvent exposure, with stretching in solvents known to yield higher strength fibres[50]. Lastly, the AS fibre was guided from roller 1 onto roller 3 (same speed as roller 1) and then stretched in a 40% EtOH bath by passing over roller 4 (speed 2× that of rollers 1 and 3) and collected in air on roller 5 [speed equal to roller 4, Figure 2.3(d)].

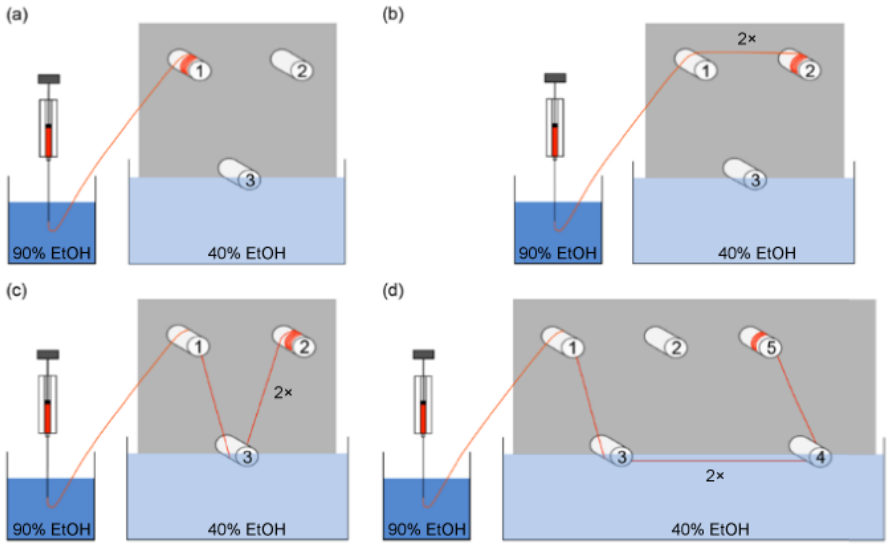


Figure 2.3. Semi-automated wet-spinning apparatus. The spinning dope was loaded into a syringe (illustrated syringe and red dope) and extruded at a constant rate through a spinneret (needle) into a coagulation bath (dark blue bath). Fibre collection was carried out using a series of rollers (1-6) with programmable speeds. Targeted conditions included as-spun (a), post-spin drawing in air (b) vs. in a solvent bath (d; light blue bath), and post-spin drawing in air after immersion in solvent (c). Reprinted by permission from ACS Biomaterials Science & Engineering[48].

2.2.7. Fibre Imaging

Following wet-spinning, fibres were allowed to dry at room temperature and plated on C-shaped paper frames (1 cm per side with a 1x1 cm gap, hereafter referred to as C-cards) by laying across double-sided tape and securing with single-sided tape, following previously described methods[47, 51]. Fibres were optically imaged at 100× magnification (Axio Observer A1; Carl Zeiss Canada; Toronto, ON), with one micrograph acquired at each edge and one in the middle (3 total). These positions were chosen to ensure tested fibres were not stretched while securing to C-cards and were consistent enough to

produce reliable data. Fibre diameters used to calculate engineering stress (force per unit area based on initial material dimensions) were calculated from the average of six measurements per image (18 per sample) using ImageJ[51]. Fibres with visual defects, large (>50%) variability in diameter, or large deviation relative to the condition were rejected for subsequent analysis.

Fibre anisotropy was probed by polarized light microscopy using an Axio Observer A1 equipped with a rotatable stage, a 90° rotatable Polarizer D, and a fixed analyzer slider (Carl Zeiss Canada). Fibres were rotated ~45° relative to the polarized light through angular measurement and sample rotation prior to image capture. Light intensity and camera angle were standardized before any image collection to allow direct comparison between samples.

Scanning electron microscopy (SEM) samples were prepared on SEM stubs following previously reported protocols[51]. For surface imaging, fibres were transferred from a C-card to an SEM stub by laying across three perpendicular strips of conductive tape, then cutting free from the paper frame. Fibres for cross-sectional images were prepared by mounting on C-cards, then submerging in liquid nitrogen for at least a minute to allow a clean fracture of the fibre upon bending. Fractured fibres were affixed vertically to SEM stubs using a thick layer of conductive tape on one or both sides of the remaining paper frame. Samples were sputter-coated (EM ACE200; Leica Microsystems Inc.; Richmond Hill, ON) with a 5 nm-thick layer of Au/Pd particles, and images were collected at 5 000× magnification for surface analysis and 10 000× magnification for cross-sectional analysis (LEO1455VP SEM, Carl Zeiss Canada)

2.2.8. Mechanical Testing of Silk Fibres

Fibres mounted on C-cards were inserted into a home-built apparatus for mechanical testing, as described by Xu et al.[47], at 50 ± 5% humidity and room temperature. On one end of the C-card was a KD Scientific Model 100 syringe pump pulling at a rate of 0.1 mm/s, and on the other end was a ~37 g weight on a Mettler Toledo XS105 Dual Range analytical balance. The weight loss over time as recorded by LabX was used to calculate engineering stress-strain curves. Fibres were excluded if they had visual defects or outlying diameters that, correspondingly, would lead to exceedingly high (diameters below ~27 μm for AS or below ~15 μm for 2× post-spin stretching conditions) or exceedingly low (for diameters above ~40 μm for AS and ~20−21 μm for 2× post-spin stretching conditions) mechanical properties. Data were processed as previously described[44] to determine engineering stress and strain, toughness, and Young's modulus. To provide a set of representative mechanical features as a hallmark of pyriform silk, outlier exclusion was performed. Data are provided in completeness to allow direct comparison to the values obtained following outlier exclusion for full assessment of the validity of this data treatment. Specifically, any fibre containing a non-repeated property more than 2.5 standard deviations away from the average was excluded from the final condition averages until all the averaged data were contained within the 2.5 standard deviations. As this was the first reported set of mechanical properties for a pyriform silk fibre without possible confounding effects arising from the attachment disc, this was performed to ensure the data were presented as a model for what was observed, while still being representative of the set of data obtained.

2.2.9. Raman Spectromicroscopy

Raman spectroscopic measurements were obtained using a diffraction-limited Raman scattering instrument with a design modified from that reported by Gullekson et al.[54] This was composed of an inverted Olympus IX71 optical microscope (Center Valley, PA) coupled to an iHR550 Raman spectrometer (Horiba Jobin Yvon; Edison, NJ). To examine the effect of post-spin processing on fibre structure, AS and 2× in EtOH fibres were tested at room temperature and humidity. Before acquiring data, fibres were exposed to the incident light for ~10-12 min to photobleach the samples and remove artifacts arising from autofluorescence. Raman spectra were acquired (10 scans, 20 s per scan, 900 to 1900 cm^{-1}, pinhole diameter of 400 µm) for three fibres per condition, three spots per fibre using a 60× objective lens. Each day data were collected, a blank was recorded and, if ambient light was detected, baseline subtraction from collected spectra was performed. Fibres were oriented at 0° and 90° relative to the polarized incident light to test for structural differences across long and short fibre axes. Using GRAMS/AI 9.0 (Thermo Scientific, Markham, ON), spectra were baseline corrected, then normalized and x-axis corrected based on the phenylalanine band[18] at 1003 cm^{-1}. Processed spectra were averaged, then a final baseline correction was performed.

2.3. Results and Discussion

2.3.1. Recombinant Pyriform Silk Production and Purification

To optimize HPy_2 expression, small-scale test expressions were conducted at 37 °C and room temperature in LB media with SDS-PAGE samples taken at varying time points (Figure 2.4). The newly appearing band on the gel, believed to correspond to HPy_2, was

more intense for experiments conducted at room temperature, indicating this was the more amenable condition for protein expression. At larger scale, HPy_2 was readily expressed in BL21(DE3) *E. coli* at room temperature, sequestering into the inclusion body after lysis in native buffer (Figure 2.5).

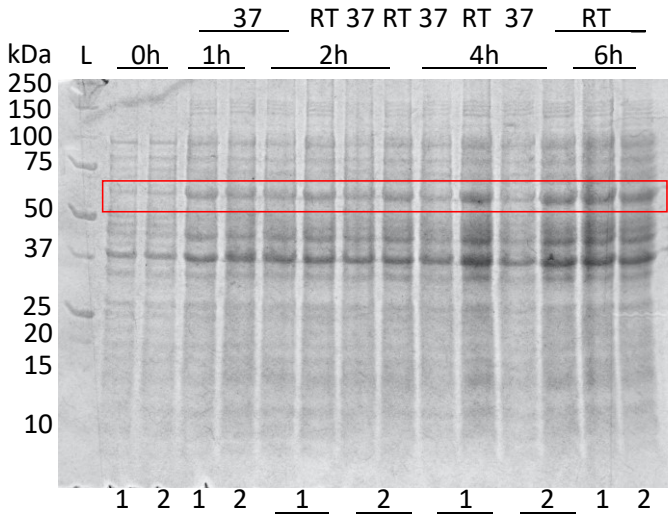


Figure 2.4. SDS-PAGE visualization of HPy_2 test expression in chemically competent BL21(DE3) *E. coli* cells at 37 °C and room temperature. The red box indicates HPy2 band migration, while 1 and 2 indicate lanes from two different colonies for the test expression.

Two purification strategies were investigated in-depth: Ni^{2+}-IMAC based on past purifications by other lab members[44] and inclusion body washes which required optimization. To determine an optimal inclusion body wash protocol, small-scale tests were performed using Triton-X and urea at varying concentrations, as precedents exist for applying both in inclusion body purification[55] (Figure 2.5). From these tests, a wash

buffer of 50 mM potassium phosphate (pH=8) with 1% Triton X-100 and 2 M urea was chosen to minimize solubilization of the target protein.

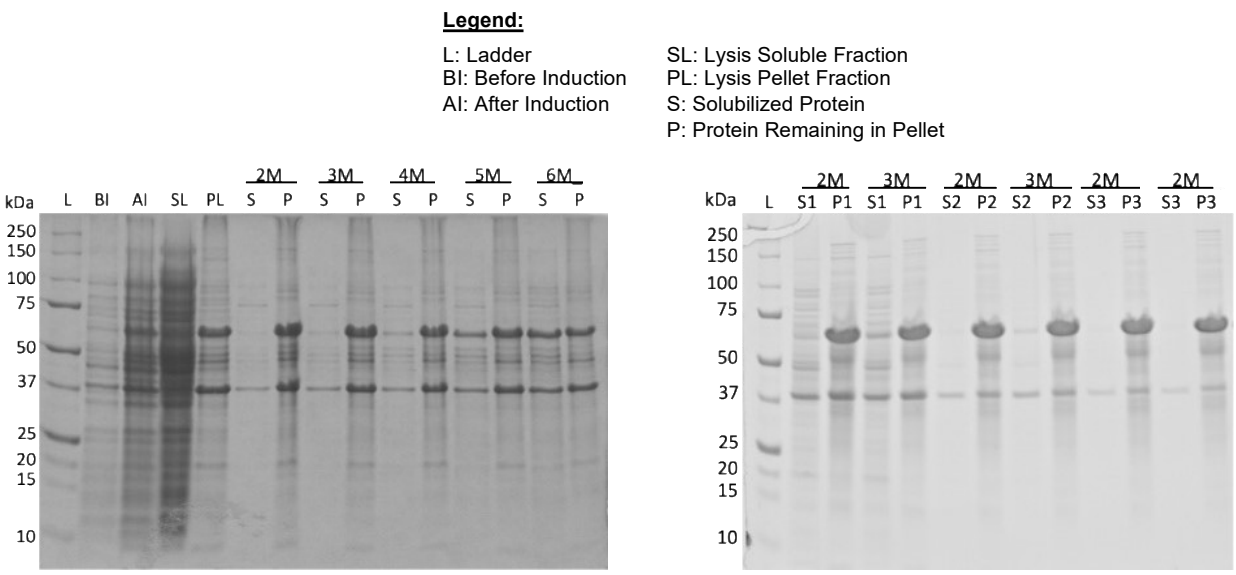


Figure 2.5. SDS-PAGE visualization of HPy_2 inclusion body wash effectiveness with varying concentrations of urea alone (a) and 2 M vs 3 M urea with 1% Triton-X (b). Protein resuspensions were left at room temperature for 30 min prior to centrifugation.

While both led to collection of a qualitatively pure protein (Figure 2.6), inclusion body washes led to a higher protein yield than Ni^{2+}-IMAC (~30-60 mg/L relative to ~24 mg) and inclusion body washes had the advantage of providing a simpler, more efficient workflow relative to Ni^{2+}-IMAC. For these reasons, the method of choice for subsequent purification was inclusion body washes. While proteins purified via Ni^{2+}-IMAC were not applied in the final fibre spinning conditions, they were used for preliminary fibre characterization and optimization of wet-spinning methods from spinning dope prepared at 8% w/v (Figures A1 and A2).

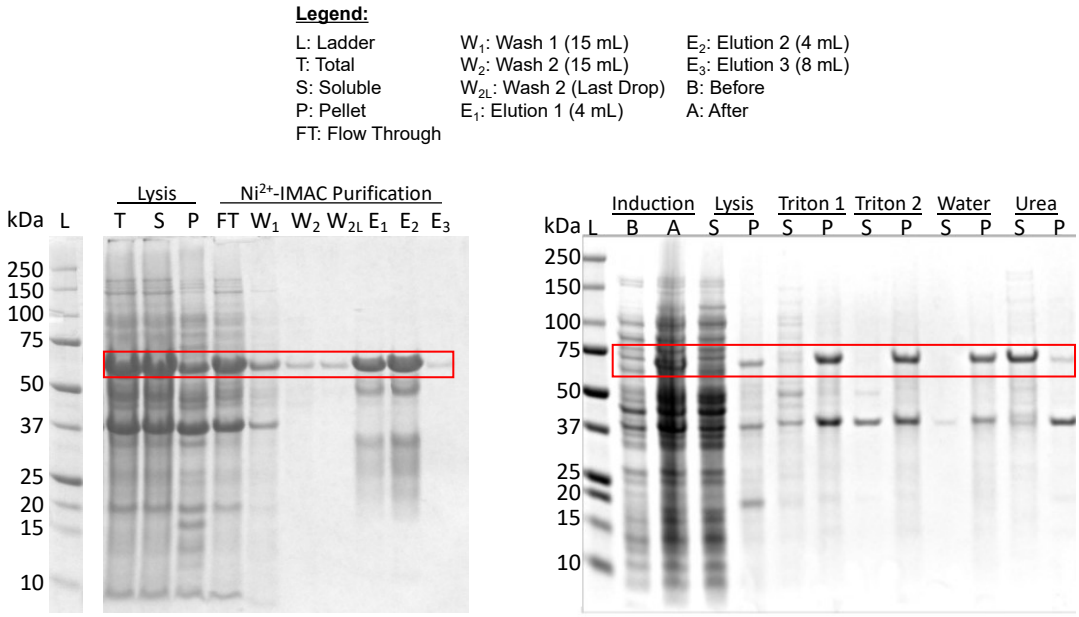


Figure 2.6. Visualized SDS-PAGE for HPy2 purification via Ni^{2+}-IMAC (left) and expression followed by inclusion body washes (right). HPy2 band migration is indicated by the red boxes.

To confirm that the major overexpressed protein resolved by SDS-PAGE with a consistent migration of 65-70 kDa corresponds to HPy_2 (47.9 kDa), mass spectrometry analysis was employed. HPy_2 in phosphate buffer was analyzed via intact protein ESI+ showed a major mass contribution approximately the size of HPy_2 (Figure 2.7). In parallel, the 65-70 kDa SDS-PAGE band was excised for in-gel chymotryptic digestion, followed by ESI+ MS/MS. This resulted in identified peptide fragments that corresponded to various regions of the HPy_2 protein, as illustrated in Figure 2.8.

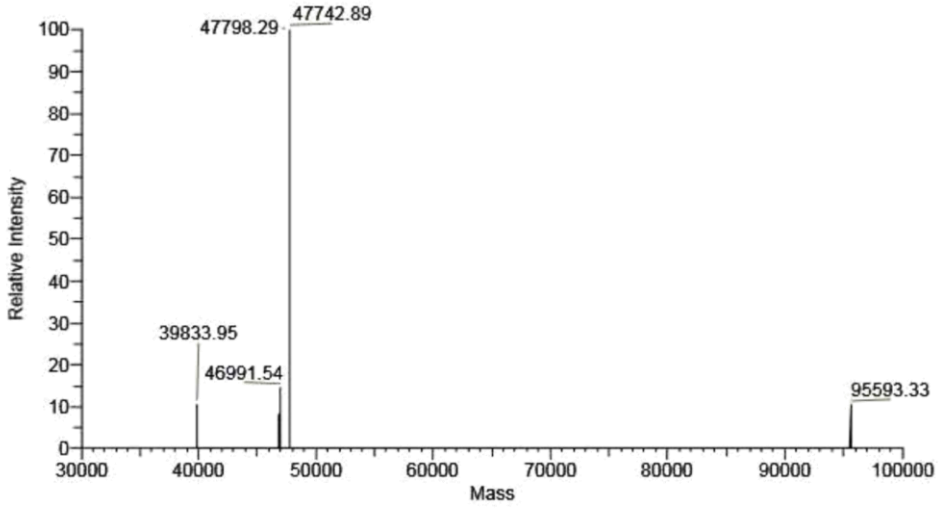


Figure 2.7. Deconvoluted ESI+ LC-MS/MS results. The calculated molecular weight of HPy_2 based on the sequence is 47.9 kDa. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

```
  1  MGHHHHHHSS AQENSFTQSS VAQQSAVAQQ SSVSQQSSAA QQSSVAQSQQ  50
 51  TSYSAATNAG SSVSQSQAIV SSAPVYFNSQ TLTNNLASSL QSLNALNYVS  100
101  NGQLSSSDVA STVARAVAQS LGLSQGSVQN IMSQQLSSIG SGASTSSLSQ  150
151  AIANAVSSAV QGSQAAAPGQ EQSIAQRVNS AISSAFAQLI SQRTAPAPAP  200
201  RPRPAPLPAP APRPRPAPAP RPAPVYAPAP VASQFQASAS SQSSAQENSF  250
251  TQSSVAQQSA VAQQSSVSQQ SSAAQQSSVA QSQQTSYSAA TNAGSSVSQS  300
301  QAIVSSAPVY FNSQTLTNNL ASSLQSLNAL NYVSNGQLSS SDVASTVARA  350
351  VAQSLGLSQG SVQNIMSQQL SSIGSGASTS SLSQAIANAV SSAVQGSQAA  400
401  APGQEQSIAQ RVNSAISSAF AQLISQRTAP APAPRPRPAP LPAPAPRPRP  450
451  APAPRPAPVY APAPVASQFQ ASASSQS                          476
```

Figure 2.8. Peptide fragments identified through in-gel chymotryptic digested protein and ESI+ MS/MS and mapped onto the HPy_2 sequence in red. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

Both MS results were consistent with the potential HPy_2 SDS-PAGE band indeed being HPy_2. The irregular migration observed is likely[56] the result of the high proline content of

HPy_2 (Table 2.1), mainly localized in the PX-rich motif (Figure 2.8). The major contaminating band at 37 kDa was found to be a bacterial protein rather than a protein truncation, based on H6 immunostaining (as will be discussed in Chapter 3) and appearance in other recombinant proteins expressed in our lab.

Table 2.1. HPy_2 composition compared to average protein amino acid composition.

Residue	# in HPy_2	Occurrence in Hpy_2	Average occurrence in proteins[a]
Ala (A)	88	18.50%	8.25%
Arg I	16	3.40%	5.53%
Asn (N)	22	4.60%	4.05%
Asp (D)	2	0.40%	5.46%
Cys I	0	0.00%	1.38%
Gln (Q)	68	14.30%	8.25%
Glu I	4	0.80%	5.53%
Gly (G)	16	3.40%	4.05%
His (H)	6	1.30%	5.46%
Ile (I)	14	2.90%	1.38%
Leu (L)	24	5.00%	9.65%
Lys (K)	0	0.00%	5.81%
Met (M)	3	0.60%	2.41%
Phe (F)	8	1.70%	3.86%
Pro (P)	38	8.00%	4.73%
Ser (S)	109	22.90%	6.62%
Thr (T)	16	3.40%	5.35%
Trp (W)	0	0.00%	1.09%
Tyr (Y)	8	1.70%	2.92%
Val (V)	34	7.10%	6.86%

[a] Average occurrence based upon amino acid composition for the complete UniProtKB/Swiss-Prot protein knowledge base, release 2019_06 (release date: July 3,

2019) as tabulated on http://web.expasy.org/docs/relnotes/relstat.html. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

2.3.2. Spinning Dope Preparation and Characterization

Based on single pyriform repeat precedents[39], widespread usage of the technique[50], and the prospective ability to form well-performing single filament fibres, HPy_2 proteins were targeted for wet-spinning. Following behaviour observed in our group for aciniform silk[47], the protein was resuspended in a 3:1:1 v/v/v TFA/TFE/H_2O solvent at a concentration that led to a qualitatively viscous, but pipettable solution. HPy_2 was dissolved at ~10% w/v in a 3:1:1 v/v/v TFA/TFE/H_2O solvent, hereafter referred to as the dope solvent based on increased viscosity in the solubility tests. Viscometric analysis indicated that the HPy_2 dope was ~2× more viscous than a corresponding dope for the ~60 kDa W_3 protein produced in-house[47] (61.5 ± 6.2 mPa·s at 22.0 ± 0.1 °C and 31.3 ± 0.9 mPa·s at room temperature, respectively). This is consistent with a lower degree of HPy_2 self-assembly or a greater degree of entanglement/formation of smaller assemblies relative to W_3.

To probe the structural state of soluble recombinant pyriform silk proteins we examined bulk secondary structure in phosphate buffer and the spinning dope using far-UV CD spectroscopy (Figure 2.9). Negative bands at 222 nm and 208 nm are consistent with an increase in α-helical character[57] in the spinning dope solution compared to the buffered protein solution. Coupled to this, based upon the observed decrease in ellipticity at 190 nm for the buffer condition relative to the spinning dope, it could indicate that some disordered regions of the protein in phosphate buffer are forming α-helices in the acidic dope solvent. Conversely, the decreased mean residue ellipticity at 190 and 220 nm in

the phosphate buffer relative to the dope may be explained by convolution with characteristic bands[58] arising from polyproline-II helices formed in Pro-rich motifs.

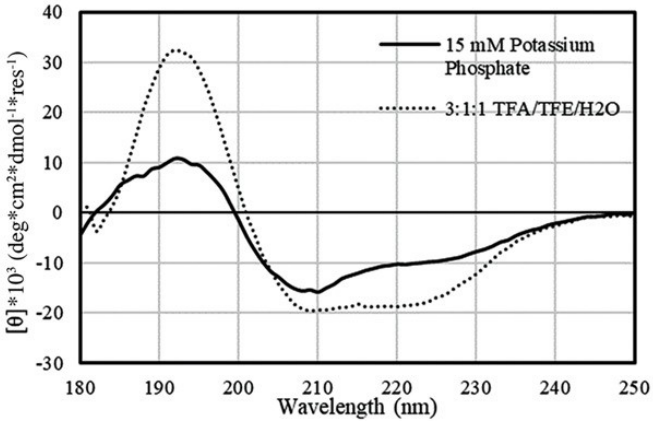


Figure 2.9. Far-UV CD spectra for HPy_2 in phosphate buffer and the dope solvent. Data were collected at 22.50 ± 0.05 °C. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

2.3.3. Silk Fibre Morphology, Structure, and Mechanics

To determine an optimal set of spinning and processing conditions, a variety of conditions were attempted as summarized in Table 2.2. Particularly challenging was a low tolerance to water of the wet-spun fibres, with visible non-uniformities observed when spun at a humidity >30% and an inability to tolerate a water post-spin stretching bath. This indicated an inherent supercontractility to these fibres, as observed in other classes of spider silks[59]. While water was an unsuitable solvent for post-spin stretching HPy_2, IPA and EtOH were found to be suitable by manually stretching AS HPy_2 fibres between two tweezers in a petri dish. By measuring length with a ruler, fibres were found to stretch ~2-3× in either 20% IPA (80% H_2O) or 20% EtOH (80% H_2O), with an increase to ~3-4× when tested with

40% IPA (60% H_2O). Due to the large volume that was required for mechanical testing, the preliminary stage of these tests, and the high similarity of these manual tests at 20% concentration of each alcohol, 40% EtOH was targeted as the stretching solvent for this work. In the future, the difference between post-spin stretching in different alcohols on mechanical properties would be interesting to explore further.

Table 2.2. Summary of conditions tested for mechanical processing.

Successful	Not Successful
As-spun	100% water post-spin stretch
2× air stretch	4x air stretch
2× stretch in 40% EtOH	4x stretch in 40% EtOH
2x stretch after passing through 40% EtOH	2x stretch in EtOH, 2x air stretch
Manually stretching in 40% IPA (small scale)	2x stretch in EtOH, delay, 2x air stretch
	Ammonium sulfate coagulation bath

Of the conditions tested, the four that appeared amenable to HPy_2 wet-spinning with the chosen post-spin stretching bath were targeted and gave a stepwise progression of the different post-spin stretching effects on wet-spun HPy_2 fibres as diagramed in Figure 2.3. Comparing the targeted conditions, optical microscopy demonstrated the expected decrease in fibre diameter upon post-spin stretching (~31 to ~ 17-18 μm; Figure 2.10, Table 2.3), indicating that fibres were irreversibly stretched during the process.

Comparing polarized light microscopy images between the different conditions showed that general post-spin treatment of fibres led to a qualitative increase in anisotropy relative to the AS state (Figure 2.10). Of the post-spin stretching conditions, stretching in air led to the least anisotropy, stretching after EtOH immersion was intermediate, and stretching in EtOH led to the highest degree of anisotropy, indicating a higher degree of structural alignment as the fibres are exposed to the stretching solvent.

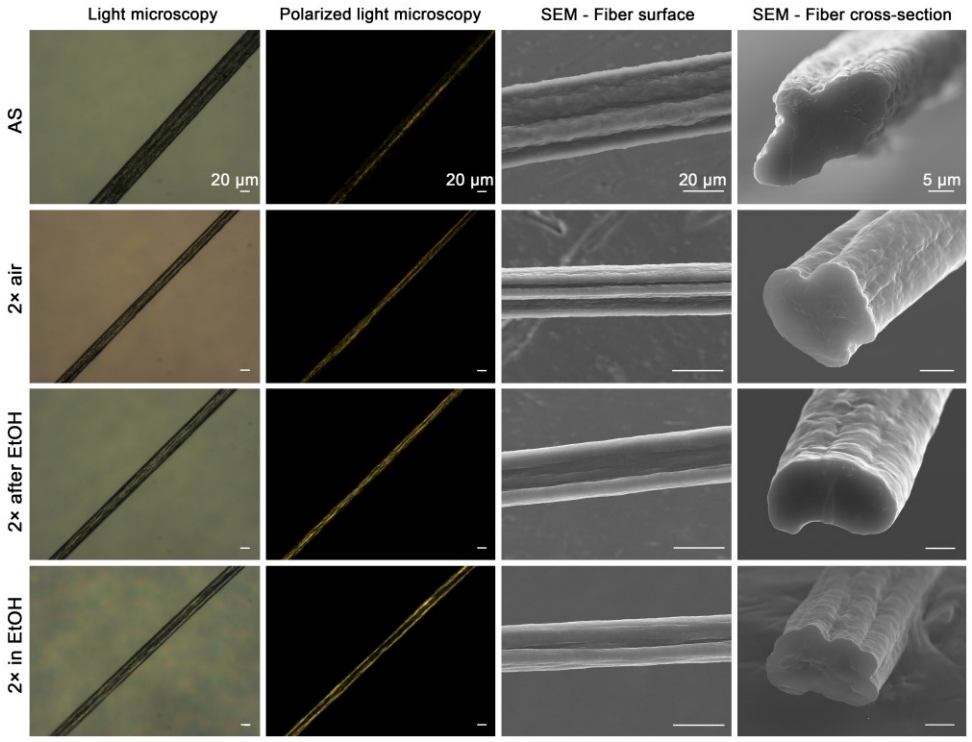


Figure 2.10. Representative micrographs for the targeted recombinant pyriform silk wet-spinning conditions. Scale bars are consistent within each column. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

Mirroring structural implications, the mechanical behavior of wet-spun HPy_2 fibres was highly condition-dependent, with repeatably observed trends per given condition (Table 2.3, Figure 2.11, Figure 2.12). Namely, post-spin stretching increased Young's modulus

and approximately doubled fibre strength at failure, with notable increases in extensibility and further modulation of Young's modulus as solvent exposure increased. Stretching in or following EtOH/H_2O immersion promoted the greatest strength, while stretching in EtOH/H_2O promoted the greatest extensibility. No clear correlation was found between fibre mechanical properties and diameter variations within conditions (Figure 2.11).

Table 2.3. HPy_2 mechanical properties for wet-spinning of the four targeted conditions. Results are reported as mean ± standard deviation for n fibres of a given type. Numbers in brackets indicates mechanical values without excluded data. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

Fibre type	Strength (MPa)	Extensibility (%)	Toughness ($MJ•m^{-3}$)	Young's Modulus (GPa)	Diameter (μm)	n
AS	32.2 ± 7.4	3.26 ± 0.54	0.57 ± 0.19	1.30 ± 0.24	32.9 ± 4.5	15[a]
2× air	53.9 ± 9.7	4.6 ± 1.2	1.70 ± 0.70	1.96 ± 0.32	18.5 ± 1.6	17[b]
	(53.7 ± 9.5)	(4.9 ± 2.0)	(1.88 ± 1.05)	(1.94 ± 0.32)	(18.5 ± 1.5)	(18)
2× after EtOH	71.6 ± 5.6	15.6 ± 5.5	10.0 ± 3.7	2.92 ± 0.33	17.6 ± 1.0	12[c]
	(71.8 ± 10.3)	(21.9 ± 14.8)	(14.8 ± 12.2)	(2.80 ± 0.46)	(17.5 ± 1.3)	(16)
2× in EtOH	65.9 ± 8.6	27.6 ± 6.1	16.7 ± 4.0	2.61 ± 0.34	18.0 ± 1.4	8[d]
	(65.3 ± 8.2)	(25.1 ± 9.3)	(15.1 ± 5.9)	(2.60 ± 0.32)	(18.1 ± 1.3)	(9)

[a] No data sets were excluded due to mechanical properties.

[b] One mechanical data set was excluded due to anomalously high extensibility [~11%; Figure A3(b)]

[c] Two data sets were excluded due to anomalously high extensibilities (~48-62%); 1 due to anomalously low strength (~42 MPa); 1 due to anomalously high toughness (~25 $MJ•m^{-3}$); all anomalous data presented in Figure A3(c)

[d] One mechanical data set was excluded due to anomalously low extensibility [~5%; Figure A3(d)]

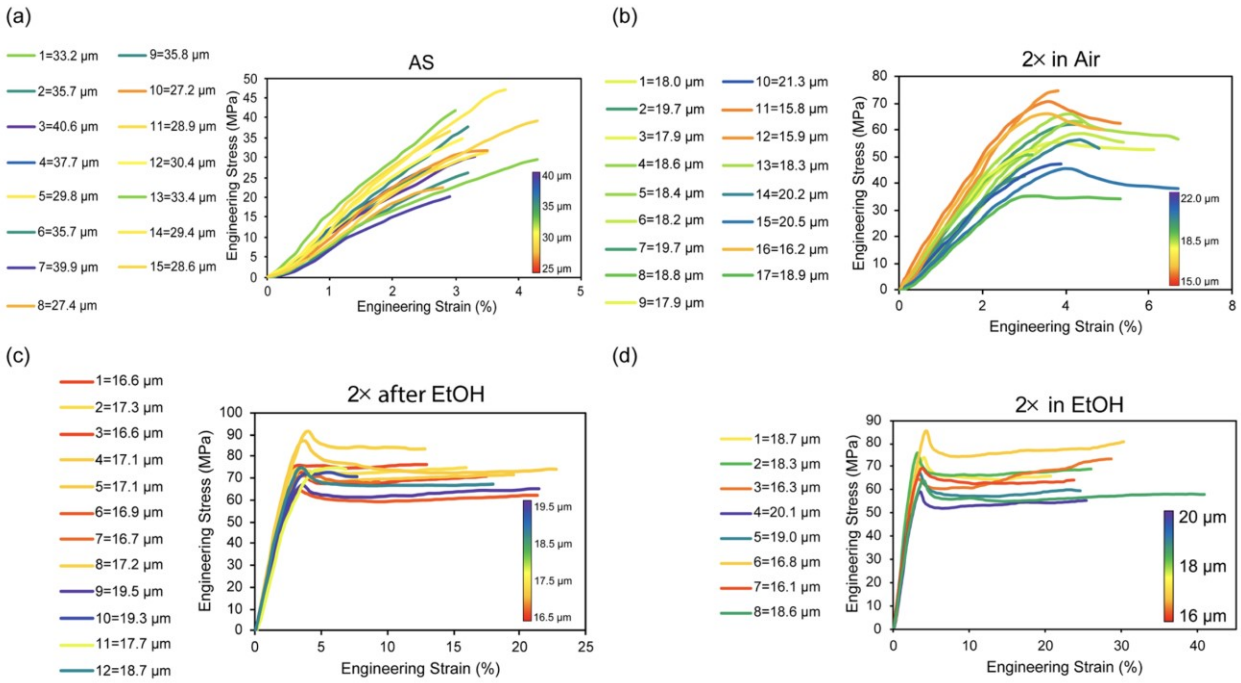


Figure 2.11. Stress-strain curves for individual fibres used for mechanical data averages. Fibres are color-coded by diameter based on the inset color bar, with actual diameters denoted in the legend. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

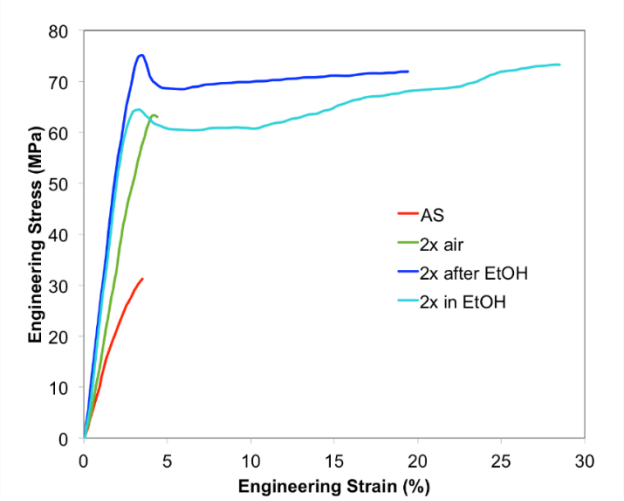


Figure 2.12. Representative stress-strain curves for HPy_2 fibres wet-spun under indicated conditions. It should be noted that engineering strength and strain are plotted (i.e., the values observed without taking into account any cross-sectional deformation occurring during the stretching process.) Reprinted by permission from ACS Biomaterials Science & Engineering[48].

Further insight into post-spin induced structural changes was provided by Raman spectromicroscopy (Figures 2.13 and 2.14). Fibres from conditions showing extremes in anisotropy and mechanical behaviour, AS and 2× in EtOH, were compared as they were expected to exhibit the most extreme degree of structural difference given the general relation between fibre-state structuring and silk mechanics [37]. The presence of amide I features implies predominantly β-sheet structure in fibres, with post-spin stretching in 40% EtOH leading to depletion of α-helical content relative to unprocessed AS fibres. Amide I bands were practically indistinguishable when oriented parallel or perpendicular to the incident light polarization; however, amide III bands exhibited some orientational sensitivity. The observed lack of amide I orientation-based differences in structural

features may reflect a lower degree of structural anisotropy than is observed in other silks such as aciniform silk[46, 47, 51]. To fully characterize this behaviour, however, further investigation with techniques such as X-ray diffraction[60] is required to elucidate the degree of structural anisotropy in these fibres. In short, structural studies indicate distinct structuring in different fibre forms, with enrichment in β-sheet content relative to α-helix.

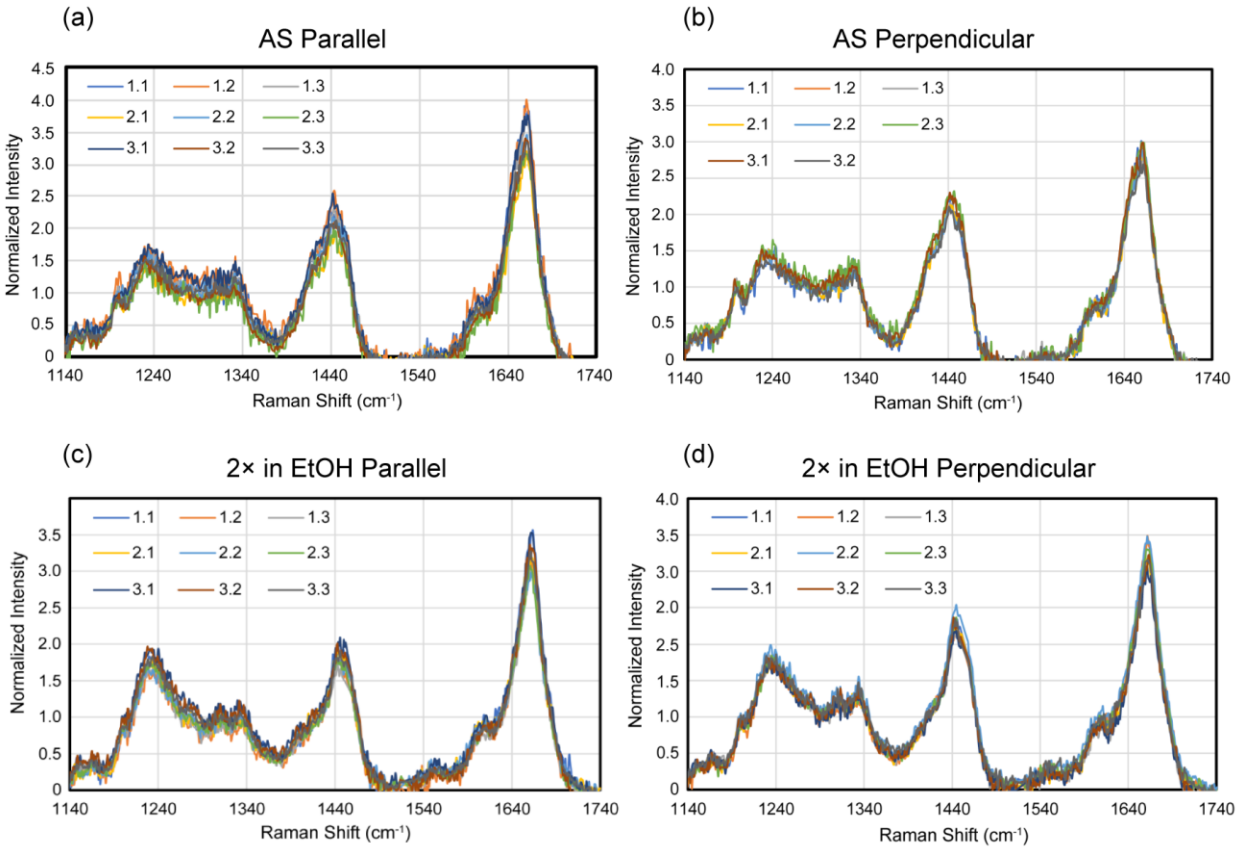


Figure 2.13. Overlay of individual Raman spectromicrographs for each condition tested. One micrograph was excluded from averaging in the AS Perpendicular condition due to highly abnormal amide I and III intensities seeming to indicate a problem with data collection or location chosen. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

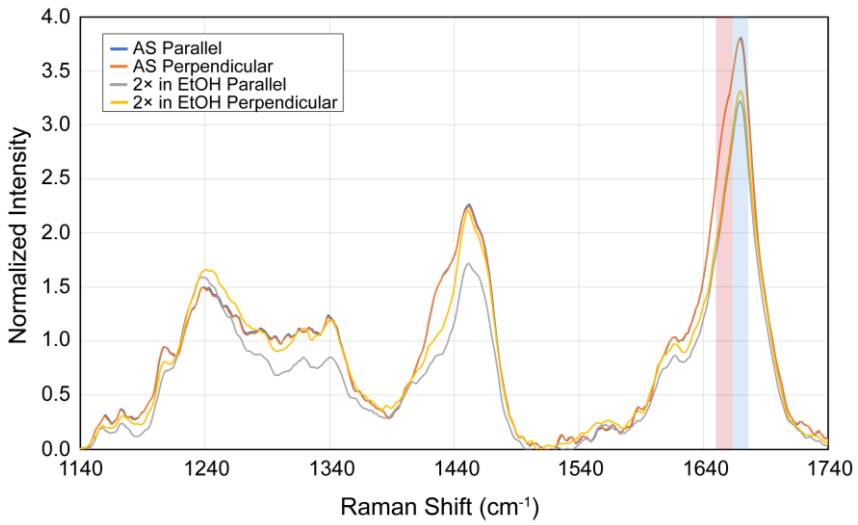


Figure 2.14. Raman spectroscopy results of AS fibres and fibres stretched 2× in EtOH. Typical Raman band regions[18] for α-helix (red) and β-sheet (blue) are shown as shaded bars. Reprinted by permission from ACS Biomaterials Science & Engineering[48].

Mechanical observations are consistent with a model where longer solvent exposure during fibre stretching promotes additional fibre-state structural rearrangements. Specifically, Raman spectroscopy implies α-helical character is being depleted (Figure 2.14), which is likely converted to β-sheet based on the near doubling of engineering stress at failure. Lack of increased β-sheet character could result from non-ideal amide I sensitivity, as demonstrated by discrepancies between amide I vs. amide III changes when comparing orientations of fibres stretched 2× in EtOH. Also contributing to improved mechanical properties is increased ordering obtained by stretching alone and further promoted with solvent exposure as supported by polarized light microscopy (Figure 2.10). Given the noncylindrical morphology of HPy_2 fibres, as observed by SEM (Figure 2.10), it was noted that the longer cross-sectional diameter was closer to the average measured by optical microscopy than the shorter dimension for AS, 2× air stretched, and 2×

stretched in EtOH conditions; conversely, the opposite trend was observed for fibres stretched 2× after EtOH immersion. This indicates that fibres spun under the former conditions likely have a lower cross-sectional area than was determined on the basis of the optical microscopy-derived diameter and are therefore stronger than indicated, while the latter condition may not be as strong as indicated. As it is not possible to use samples for both mechanical testing and SEM cross-sectional analysis to accurately measure noncylindrical geometry, optical microscopy measurements, despite the noted limitations, were employed for diameter estimations for mechanical evaluation of samples used in tensile testing.

2.3.4. Comparison to Other Silks

Based on mechanical analysis, HPy_2 fibres exhibit a favorable combination of strength and extensibility. Since the publication of this work, this was supported by work based on the repetitive and terminal domains of pyriform silk from *Araneus ventricosus*[40], which demonstrated even higher extensibility than found with the methods reported in Chapter 2. This will be revisited in Chapter 3, where further condition-dependent differences for HPy_2 wet-spinning are detailed. Compared to similarly-sized aciniform proteins produced in-house via the same methods[47] (W_3; ~60 kDa), HPy_2 (~48 kDa) fibres exhibited a similar combination of strength and extensibility following a proposal by Lefèvre et al.[18] that aciniform and pyriform silk would have similar properties, but with strength and extensibility being slightly lower for HPy_2. Conversely, compared to a similarly sized (54.6 kDa) dragline silk[61], HPy_2 fibres exhibited comparable strength and greater extensibility,

indicating the potential for a high mechanical performance material with the addition of more repetitive domain units as supported by recently reported work.[40]

2.4. Summary

The work presented in this Chapter provided the first insight into the mechanical properties of pyriform silk fibres. First, it was demonstrated that recombinant silks based on the pyriform silk repetitive domain of *A. argentata* could be produced at good purity and reasonable yields. Soluble HPy_2 proteins were shown to have significant α-helical content, consistent with Raman spectromicroscopy-based structural studies of the pyriform gland[18]. Wet-spinning of this recombinant silk protein construct yielded strong fibres with extensibility that could be tuned on the basis of solvent exposure during post-spin processing. Changes in mechanical properties upon post-spin stretching in 40% EtOH were echoed by a change in anisotropy and observed loss of α-helicity. These results implied the potential for further tunability of mechanical properties based on further EtOH exposure or more stretching of fibres, which might be more amenable to higher-order constructs (e.g., HPy_3). Overall, the work in this chapter provides key insights into methods of recombinant pyriform silk expression and mechanical and structural properties of wet-spun fibres from a protein containing two repeat units of the *A. argentata* PySp1 repetitive domain. In Chapter 3, I further optimize the methods developed for HPy_2 expression to allow for high yield at a laboratory scale.

Chapter 3

Expression Revisited: Using Amino Acid Supplementation to Increase Protein Yield and Express HPy_3

3.1. Introduction

Chapter 2 discussed the protein design, expression, and purification used to give the first mechanical data on pyriform silk fibres, whether from a recombinant source or produced by a spider. As will be discussed in Chapter 3, the expression system introduced and detailed in Chapter 2 was limited to HPy_2 as the largest expressible protein without moving to more complex systems such as terminal domain addition[62] or intein splicing[29]. While exploring selective unlabelling approaches for NMR (as will be discussed in Chapter 5), I noticed a drastic increase in protein yield. This led me to experiment further to determine the minimal required set of amino acids needed to increase protein yield in a chemically defined medium. By fine-tuning the amino acid supplementation scheme to the protein rather than adding all 20 amino acids, I developed a straightforward and more cost-effective M9 supplementation and expression strategy that would potentially allow other researchers to more readily investigate other low expressing proteins. In addition, I found that, when reoptimized, these methods will likely allow for the expression of pure HPy_3 protein.

3.1.1. Amino Acid Supplementation During Expression

At the start of this project, I expressed recombinant pyriform silk using LB media based on previous work in the Rainey lab[44] and the only reported recombinant pyriform silk

expression at the time[39]. This method successfully expressed PySp1-based silk proteins with low numbers of repeat units (i.e., 1 or 2) at workable yields[48], but at higher repeat numbers (i.e., 3), expression yield drastically decreases[44] and truncations become more prevalent. Therefore, new methods would be required to express higher repeat-containing proteins for future work.

By coincidence, while exploring selective unlabelling[63] approaches to enable solid-state NMR spectroscopic characterization of HPy_2, which will be discussed in Chapter 5, it was found that amino acid supplementation to M9 minimal media using supplementation scheme C1 led to a striking increase in protein yield relative to expression in LB medium. Initial qualitative gel results also indicated a lower amount of truncated protein being produced relative to the full length HPy_2, giving additional incentive to explore this method of protein expression. To my surprise, this was a relatively unexplored method in the literature. Typical approaches added most[64] or all of the 20 naturally occurring amino acids[65, 66], which requires creation of a high concentration starting solution, varying the pH to solubilize all amino acids, and also prevents the application of isotope-enrichment for, e.g., NMR spectroscopy. Using instead a specific concoction of amino acids tailored to the protein of interest addresses the challenges with current methods, reduces cost from supplementation of unnecessary components, simplifies the media preparation through judicious choice of amino acids with similar solubility considerations, and allows for a more rigorous approach that can potentially be applied to express a wide variety of recombinant proteins. Over and above these benefits for protein production, this optimized amino acid supplementation approach is shown here to provide a method to

scale up recombinant spider silk protein production, overcoming frequently observed issues of low yields[44, 48].

3.2. Methods

3.2.1. Materials List

All amino acids were obtained from Sigma Aldrich Canada Ltd. (Oakville, Canada) unless otherwise state: L-Glutamine–- Gibco Laboratories (Grand Island, USA); L-Arginine – Acros Organics (New Jersey, USA); L-Alanine – Sigma Aldrich (Milwaukee, USA); Urea, Glycine–- BioShop (Burlington, Canada); NaH_2PO_4 anhydrous – VWR (Fountain Parkway, USA); $NaH_2PO_4 \cdot H_2O$, NaCl, Ammonium Sulfate, $Fe_2(SO_4)_3$, $MgSO_4 \cdot 7H_2O$, K_2HPO_4, KH_2PO_4 – Fisher Scientific (Fair Lawn, USA); $^{15}NH_4Cl$ – Cambridge Isotope Laboratories Inc. (Andover, USA); D_2O, DSS, $Fe_2(SO_4)_3 \cdot 7H_2O$ – Sigma Aldrich (St. Louis, USA); IPTG, Imidazole, NaN_3 – Bio Basic (Markham, Canada).

3.2.2. Protein Expression, Isotopic Labelling, and Purification

As discussed in Section 2.2.1, a plasmid encoding the protein of interest was inserted into chemically competent BL21(DE3) cells, with differences in expression and purification protocol discussed here. Starter cultures grown overnight at 37 °C were transferred to 1.2 L of LB medium at 1:50 v/v and grown at 37 °C until an OD_{600} between 1.0-1.2 was reached. The cells were then pelleted (4000 rcf, 4 °C, 25 min) followed by 2.4:1 v/v transfer to M9 media[67] (100 mM NaH_2PO_4, 40 mM K_2HPO_4, 4 mM $MgSO_4$, 1.8 µM $FeSO_4$, 2 g/L glucose, 1 g/L of $^{15}NH_4Cl$ for isotopic enrichment or 1.2352 g/L $(NH_4)_2SO_4$ for no isotopic enrichment to keep NH_4^+ concentration consistent, and 100

mg/L ampicillin, titrated to pH 7.3 using NaOH). Prior to media transfer, one of nine amino acid cocktails (C1-C9; Table 3.1) was chosen for supplementation and checked for expression yield and purity. For HPy_3, M9 medium was supplemented with 1.5× the concentration of amino acids in C7.

Table 3.1. Amino acid cocktail compositions for expression tested in M9 medium.

Final concentrations of given amino acid (AA) are noted.

AA	Cocktail Composition								
	C1	C2	C3	C4	C5	C6	C7	C8	C9
	[AA] (mM)	[AA] (mM)	[AA] (mM)	[AA] (mM)	[AA] (mM)	[AA] (mM)	[AA] (mM)	[AA] (mM)	[AA] (mM)
Ala	20.7	20.7	20.7	20.7	20.7	20.7	20.7	5	10
Arg	0	0	0	0	0	0	0	5	10
Asn	0	0	0	0	0	0	0	5	10
Asp	0	0	0	0	0	0	0	5	10
Cys	0	0	0	0	0	0	0	5	10
Glu	0	0	0	0	0	0	0	5	10
Gln	27.4	27.4	27.4	27.4	27.4	27.4	27.4	5	10
Gly	6.0	0	0	0	6.0	6.0	0	5	10
His	0	0	0	0	0	0	0	5	10
Ile	21.3	0	0	21.3	21.3	0	0	5	10
Leu	10.7	0	10.7	10.7	10.7	10.7	10.7	5	10
Lys	0	0	0	0	0	0	0	5	10
Met	0	0	0	0	0	0	0	5	10
Phe	0	0	0	0	0	0	0	5	10
Pro	34.7	34.7	34.7	34.7	34.7	34.7	34.7	5	10
Ser	60.9	60.9	60.9	60.9	40.0	40.0	40.0	5	10
Thr	0	0	0	0	0	0	0	5	10
Trp	0	0	0	0	0	0	0	5	10
Tyr	0	0	0	0	0	0	0	5	10
Val	13.7	0	0	0	13.7	13.7	13.7	5	10

The initial amino acid cocktail (C1) was chosen based on a previous selective unlabelling scheme used for aciniform silk (Xu, Tremblay, and Rainey, manuscript in preparation)

and modified using the smaller pyriform silk protein, HPy_1 as discussed in detail in Chapter 5. Modification to C1 were determined through amount of non-specific scrambling observed through NMR analysis, precise reporting of selective amino acid uptake during recombinant expression with bacterial cells[65], and observation of a diagnostic truncation band. Cells resuspended in modified M9 media (500 mL per 2L flask) were incubated with shaking at 37 °C for 1 h to allow uptake of the media components, followed by induction with 0.8 mM IPTG and 20 h expression at room temperature (~22.5 °C).

HPy_2 proteins were purified near identically to the Ni^{2+}-IMAC purification discussed in Section 2.2.2, with a larger scale for Ni^{2+}-IMAC steps. HPy_2 resuspended in 10 mL/g wcw in 8 M urea buffer (8 M urea, 100 mM NaH_2PO_4 10 mM tris base, pH 8.00 ± 0.05), then diluted to obtain 4 M urea (otherwise identical buffer), was purified via Ni^{2+} IMAC (~30 mL dry bead volume) using a wash step with 300 mL of wash buffer (4 M urea, 100 mM NaH_2PO_4 10 mM tris base, 10 mM imidazole, pH 8.00 ± 0.05) followed by three 80 mL elution steps (4 M urea, 100 mM NaH_2PO_4 10 mM tris base, 250 mM imidazole, pH 8.00 ± 0.05). Eluent fractions were dialyzed against 2 M urea, then 1 M urea (both in 100 mM NaH_2PO_4 10 mM tris base, pH 8.00 ± 0.05) for a minimum of 4 h per urea concentration to allow gentle refolding of the protein, followed by dialysis against water to remove all salts prior to lyophilization if intended for fibre spinning. For HPy_3, the same buffer and bead volumes described in Section 2.2.2 were used.

3.2.3. NMR Sample Preparation and NMR Spectroscopy

Samples collected from Ni^{2+} IMAC purification and refolded via gradual urea removal were dialyzed against a ~55.6 mM potassium phosphate buffer (pH 6.00 ± 0.05). The

protein in phosphate buffer was concentrated via centrifugal spin columns (Vivaspin 20, 3 kDa MWCO, Little Chalfont, UK; Amicon Ultra – 15, 3 kDa MWCO, Tullagreen, Ireland) until a concentration reasonable (i.e., at least 100-200 μM) for ^{1}H-^{15}N heteronuclear single quantum coherence (HSCQ) NMR experiments was reached. Protein concentrations were determined as described in Section 2.2.5[48, 52].

To prepare the sample for NMR, a 10 mM solution of DSS/NaN_3 in 100% D_2O was then added 1:9 to the concentrated protein solution, giving a final buffer condition of 50 mM potassium phosphate (pH 6.00 ± 0.05) with 1 mM DSS and 1 mM NaN_3. Samples were stored overnight at 40 ºC after preparation in final NMR buffer conditions to match the experimental NMR condition. On the day of the experiment, prior to loading the sample into a Shigemi tube or standard 5 mm NMR tube, the concentrated protein solution was centrifuged to remove any insoluble aggregates, with the supernatant retained and evaluated in terms of both pH and protein concentration. Typical final protein concentrations ranged from 100-200 μM. All ^{1}H-^{15}N HSQCs were collected on a Bruker AVANCE III 11.7 T NMR spectrometer at 40 ºC (Nuclear Magnetic Resonance Research Resource (NMR^3) facility, Dalhousie University, Halifax, NS), processed using TopSpin 4.0.8, and indirectly referenced to DSS[68].

3.2.4. Dope Preparation, Viscosity, and Fibre Spinning

To determine if the M9 media-based expression methods had an effect on HPy_2 fibre formation, I resuspended lyophilized protein 3:1:1 v/v/v TFA:TFE:H_2O as described in Section 2.2.4 and qualitatively examined viscosity to determine the optimal concentration[48]. Dope viscosities were measured identically as described in Chapter 2 at

22.5 ºC. Fibres were spun at appropriate w/v of solubilized proteins using a 90-100% EtOH coagulation bath, and post-spin stretched using the conditions illustrated in Figure 2.3, as well as stretching 2× after being passed through water. Fibre spinning and data analysis were performed by me and Skylah McLeod van Wagoner, an undergraduate student in the Rainey lab that I mentored during her summer and honours research.

3.3. Results and Discussion

3.3.1. Preparation of Supplemented M9 Media

Each amino acid cocktail was produced through addition of the appropriate amounts of powdered amino acids to M9 medium on the day of use. Cocktails C1-C7 (Table 3.1) were found to have no insoluble amino acids after swirling the flasks and letting them sit for a few minutes at room temperature. Cocktails C8 and C9 (Table 3.1), each using all amino acids, resulted in insoluble amino acid powder that weren't taken up in solution even after the 20 h expression due to low solubility of some amino acids (likely, e.g., glutamic acid) at pH 7.3 as is required for M9 medium. Contrarily, all other supplementation mixtures were readily soluble at the M9 pH, making expression easy to scale as needed.

3.3.2. Expression and NMR Analysis of HPy_2

HPy_2 served as an ideal recombinant protein target for testing the utility of supplementing different combinations of amino acids to the M9 medium based on its relatively low expression yield/litre of LB medium[48]. I initially started with cocktail C1 based on the increased yield observed during selective unlabelling studies (as will be detailed in

Chapter 5), then iteratively adjusted this to determine the minimal set of amino acids needed to achieve higher expression of a pure protein (Table 3.1). For all experiments to test expression, cell density as measured by OD_{600} (with a 1/10 dilution to accurately determine OD_{600} values) were made to be similar after incubating in M9 medium for 1 h before induction of expression (OD_{600} of ~2.5-3.0). For all experiments using the different amino acid supplementation schemes C1-C9, cell densities were found to be similar at the end of the expression process (OD_{600} of ~6.0-8.0) as tabulated in Table 3.2.

Table 3.2. OD_{600} values during growth and expression. Colour coded numbers represent independent experimental replicates.

Cocktail Used	OD_{600} Before Induction	OD_{600} After expression (20 h unless otherwise indicated)
No Supplementation	2.8; 2.9	3.7; 5.1
C1	2.1; 2.5; 2.9	5.4; 6.4; 8.2
C2	2.6	6.8
C3	2.7	7.2
C4	2.8	6.2
C5	2.7	7.0
C6	2.7	7.6
C7	2.9; 2.8; 2.7	6.5; 6.0; 6.2
C8	2.8	7.4
C9	2.4	6.4

After initial results for HPy_2 expression in M9 medium with supplementation scheme C1 showed markedly increased yields (>240 mg/L; Table 3.3) compared to what was observed for expression in LB medium (~24 mg/L with Ni^{2+}-IMAC purification)[48], I wanted to test how my supplementation approach compared to addition of all amino acids as tested by other groups[65, 66]. I assessed this by adding 5 mM (C8) and 10 mM (C9) of all

20 natural amino acids based on previously reported upper thresholds of effectiveness from Kumar et al.[65], with each of these mixtures resulting in similar expression levels to our in-house mixture. Knowing that our mixture was at least as effective as the best alternative, I then decided to explore the potential to further reduce the number of added amino acids being supplemented while still achieving a high protein yield. I first attempted addition of only the four most abundant amino acids in the protein (Table 2.1), with cocktail C2 (Table 3.1). However, based on the observation of a major truncation product via SDS-PAGE (Figures 3.1 and 3.2), I concluded that addition of only the four most abundant amino acids was insufficient to increase production of the full-length protein while producing a readily purifiable protein.

Table 3.3. Growth and expression in M9 media for cells containing HPy_2 encoding plasmid. Colour coded numbers represent independent experimental replicates.

Cocktail	Wet Cell Weight	Protein Yield (mg)	Protein Yield (mg/L)	Protein Yield (mg/g wet cell weight)
No Supplementation	4.0; 5.2	-; 31.1	-; 62.2	-; 6.0
C1	5.3; 6.7; 7.7	140.8; 165; 87.5	281.6; 330; 255.6*	26.6; 24.6; 16.5*
C2	6.1	113.7	287.3*	23.6*
C3	6.7	94.1	221.9*	16.6*
C4	6.7	68.7	169.1*	12.6*
C5	6.6	79.1	118.3*	14.2*
C6	7.1	74.2	197.1*	13.9*
C7	7.0; 5.4; 5.9	94.5; 126.5; 113.4	242.7*; 253; 226.8	17.3*; 23.4; 19.2
C8	6.4	117.7	235.4	18.4
C9	7.4	93.2	264.9*	17.8*

* indicates the value was calculated based on the mass measured while taking into account sample loss or samples allocated for NMR purposes

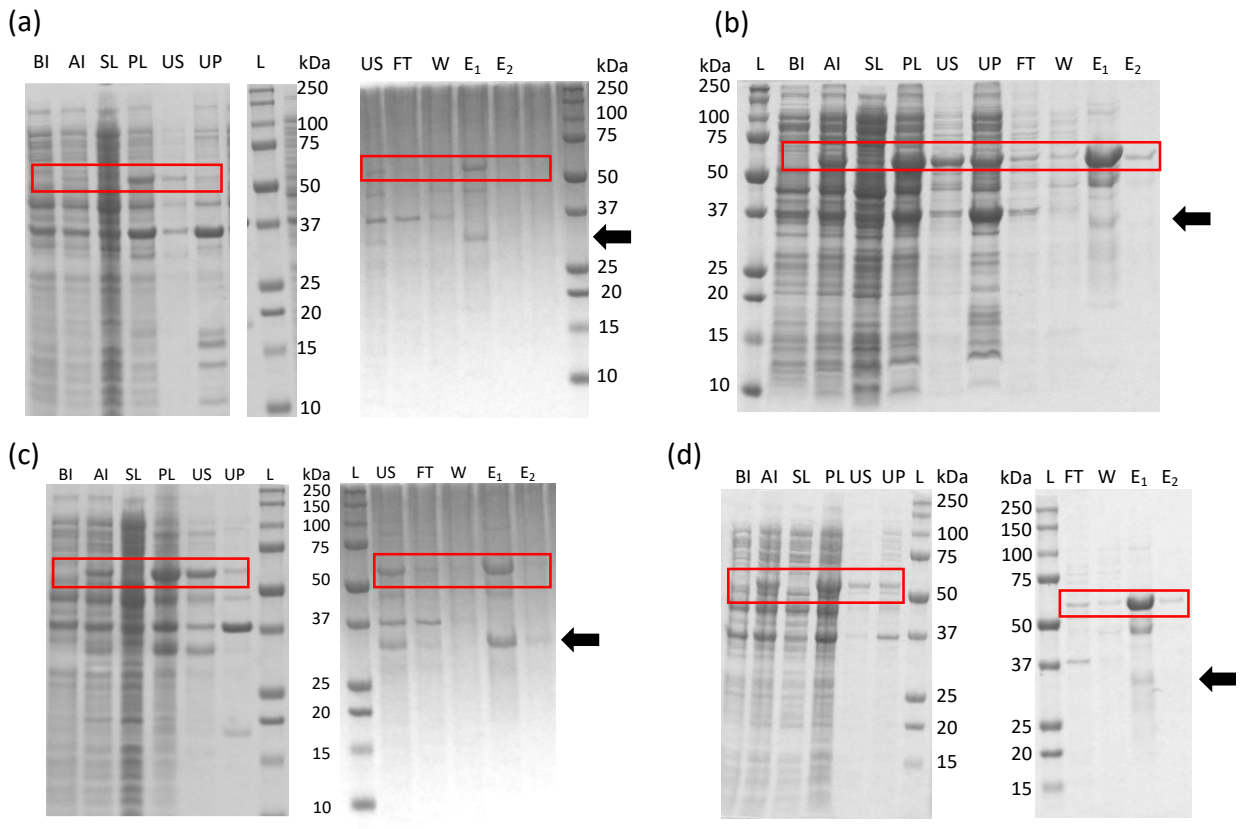


Figure 3.1. HPy_2 expression and purification with and amino acid supplementation in M9 media. Shown are full SDS-PAGE gel results for expression and purification without amino acid supplementation (A), with supplementation scheme C1 (B), with supplementation scheme C2 (C), and with supplementation scheme C7 (D). Red boxes indicate the bands of interest. The labelling is as follows: BI=before induction, AI=after induction, SL=soluble portion of lysis, PL=insoluble portion of lysis, US=soluble protein after 8 M urea resuspension, then dilution to 4 M, UP=protein not soluble in urea, FT=Ni^{2+} IMAC flow through, W=Ni^{2+} IMAC wash, E= Ni^{2+} IMAC elution. The black arrow indicates the band used to assess supplementation efficacy.

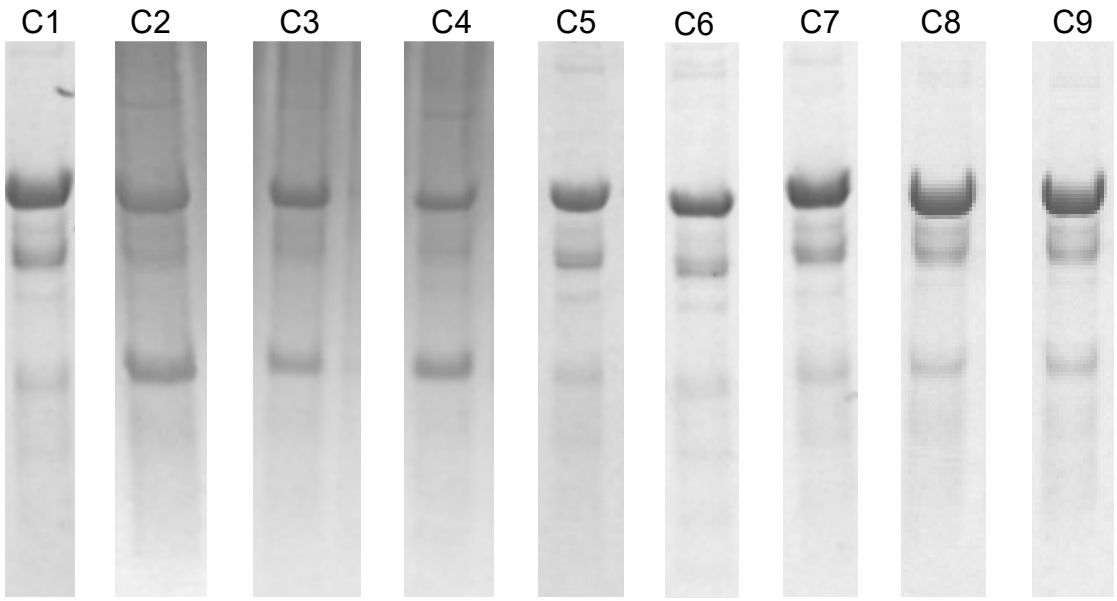


Figure 3.2. Comparison of E1 from each HPy_2 expression condition.

I then decided to test whether adding leucine alone (cocktail C3; 11 mM) or both leucine and isoleucine (cocktail C4; 11 mM and 21 mM respectively) would alleviate issues leading to excess truncations, with neither leading to the same purity level achieved by C1 (Figure 3.2). Working backwards from C1 to minimize amino acids added, I made small changes such as lowering the level of serine added (cocktail C5) and removing isoleucine and glycine (cocktail C7; 21 mM and 6 mM respectively). It was found that I could remove isoleucine and glycine while lowering the amount of serine without detrimental effect to protein yield or purity. A summary of all expression results is shown in Tables 3.2 and 3.3 and Figure 3.2.

Comparing protein yields, a striking difference can be seen between yields previously observed for HPy_2 in LB medium with Ni^{2+} IMAC purification (~24 mg/L)[48] and those observed in M9 media regardless of supplementation scheme. Without the addition of amino acids, protein expression doubled in M9 relative to LB which is likely a result of the buffering capacity of M9[69], but contained higher truncation levels as observed from non-optimal supplementation schemes [Figure 3.1(a)]. An increased yield indicates a positive effect from expression in M9, with the caveat that this led to higher cell stress from amino acid starvation consistent with the observed higher levels of truncated protein. Addition of amino acids helped to prevent this, leading to at least three-fold increase in protein yield for all supplementation mixtures. This difference was also seen when comparing cell weight, which contained less variability taking into account differences in cell growth. Most yields were consistently around 16-19 mg/g wet cell weight when expression media was supplemented with amino acids, indicating a level of tunability to express other target recombinant proteins with different amino acid compositions.

3.3.3. NMR Analysis of Residue Incorporation via Unlabeling

Considering the influence of types of amino acids supplemented based on selective unlabelling techniques that originally inspired this study, we adapted the NMR approach for analyzing the effectiveness of selective amino acid unlabelling (discussed in more depth in Chapter 5) to probe the efficacy of selective amino acid utilization. To investigate the extent of amino acid incorporation into recombinant HPy_2, ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) experiments were carried out[70]. This experiment shows a peak for all H-N correlations in a protein or peptide, thereby giving a unique fingerprint for

a protein. Due to the presence of a prominent truncation band diagnostic for lack of a desired amino acid, I looked more broadly at the general intensity of the ^{1}H-^{15}N backbone amide signals given a similar protein concentration rather than changes in specific residues. Taking the spectral intensity as a whole gives an idea of the relative level of non-specific scrambling occurring during expression, thereby giving an idea of whether or not a noticeable excess of a specific amino acid was supplemented. This became particularly apparent when comparing the concentration of serine at 40 mM and 61 mM (Figure 3.3) to see if a higher signal to noise ratio was obtained. This showed increased non-specific scrambling with the higher amounts of supplemented amino acids, which was used as part of the rational to lower serine concentration for supplementation scheme C7.

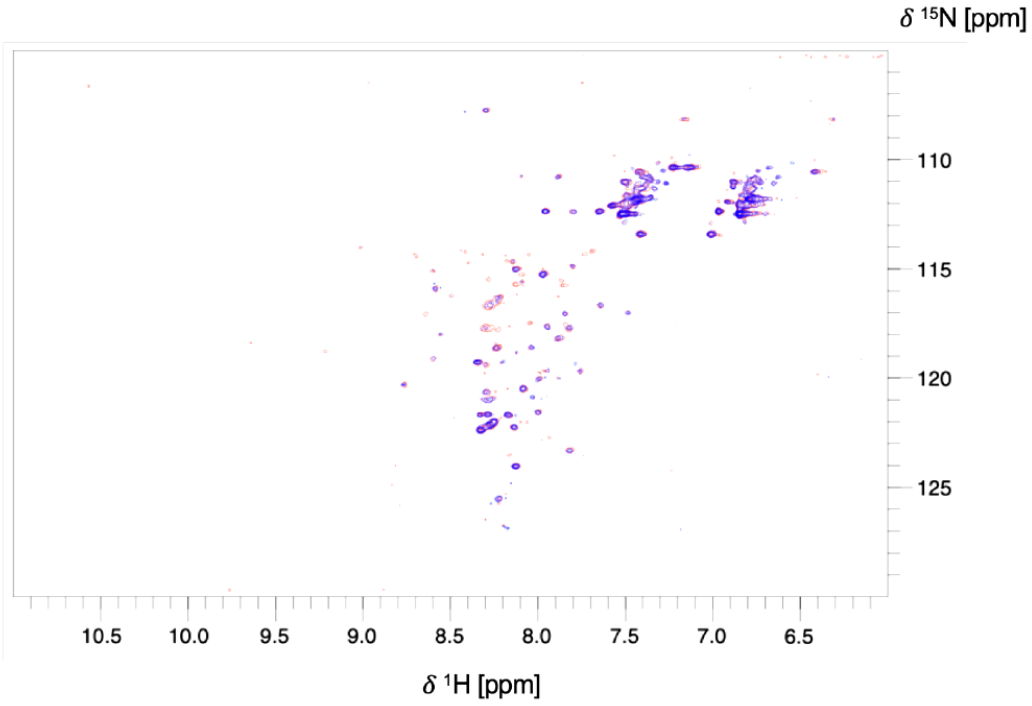


Figure 3.3. ^{1}H-^{15}N HSQC comparison of protein expressed with supplementation schemes C1 (blue) and C5 (red). Both were ~150 µM and collected identically.

3.3.4. Dope Solution and Fibre Characterization

From the moment HPy_2 expressed with amino acid supplementation was solubilized in the dope solvent, there was a clear difference in properties relative to what was observed for LB-expressed HPy_2. Dope solutions were found to be much less viscous than the solution discussed in Chapter 2 (Table 3.4) and more similar to viscosities observed for comparable recombinant aciniform silk dope solutions[47]. This indicates a higher level of protein-protein interactions in a dope made from M9 expressed protein (hereafter referred to as M9-HPy_2 to distinguish from LB expressed HPy_2, LB-HPy_2), which could result from a higher amount of target protein in the mixture relative to truncations or bacterial proteins.

Table 3.4. Viscosity of dope solvents prepared using LB expressed protein vs. protein expressed in supplemented M9.

Dope (% w/v)	LB Expressed HPy_2 Viscosity (mPa*s)	M9 Expressed HPy_2 Viscosity (mPa*s)
15% HPy_2	-	26.9 ± 0.5
10% HPy_2	61.49 ± 6.15[1]	16.7 ± 0.3
7.5% HPy_2	-	11.2 ± 0.2
5% HPy_2	-	7.5 ± 1
2.5% HPy_2	-	4.13 ± 0.08
1% HPy_2	3.795 ± 0.014	-
Dope Solvent	2.3 ± 0.1	2.26 ±0.05

[1]As reported in Simmons et al.[48]

Beyond solution viscosity and presumed self-assembly implications, M9-HPy_2 was more tolerant of high EtOH coagulation baths and exposure to H_2O during post-spin processing than LB-HPy_2 and did not appear to have the same supercontractile properties in the

fibrous state. This allowed the use of higher concentration EtOH baths as a stronger coagulant for the increased protein solutions, and, while stretching in 100% H_2O was still not possible, submersion in 100% H_2O was tolerated.

Three spinning dope conditions were targeted for fibre spinning to test these new capabilities: 15%, 20%, and 25% w/v. Of these conditions, fibres from 15% dope were difficult to spin due to low viscosity, fibres from 25% were difficult to spin due to high viscosity, and fibres from 20% dope were relatively straightforward to spin and collect, focusing our studies on 20% w/v dopes. Compiled mechanical results for fibres spun from 20% w/v dopes are shown in Table 3.5, while 15% and 25% w/v spun fibres are shown in Table A1. It should be noted that AS mechanical results are not included in this table, as the fibres adhered to the roller to the point of being untestable.

Table 3.5. Tabulated mechanical results for fibres spun from 20% w/v M9-HPy$_2$ protein

% HPy2	% EtOH Coagulation Bath	Post-spinning condition	Temp (ºC)	% Humidity	Diameter (µm)	Stress at failure (MPa)	Ultimate Tensile Strength (MPa)	% Strain	Toughness (MJ/m3)	Young's Modulus (GPa)	Number of Fibres (N)
20%	95%	2× in air	33.5	24.2	22.4 ± 2.9	46.6 ± 7.5	51.2 ± 9.2	69.6 ± 6.6	30.4 ± 30.0	1.57 ± 0.32	16
20%	95%	2× in air; $^{15}N/^{13}C$ labelled	30-33	24-28	17.6 ± 2.0* (17.9 ± 1.8)	66.8 ± 6* (62.7 ± 9.9)	67.2 ± 5.6* (64.8 ± 6.5)	166 ± 30* (128 ± 75)	90.5 ± 17.5* (69.2 ± 42.2)	1.89 ± 0.09* (1.79 ± 0.23)	6* (8)
20%	95%	2× in 40% EtOH	34.2	27.0	22.5 ± 2.5	67.4 ± 16.1	70.1 ± 17.8	57.4 ± 7.1	34.9 ± 46.4	2.23 ± 0.50	11
20%	95%	2× before H2O	30-35	27.0	20.9 ± 4.3	51.3 ± 14.2	52.6 ±13.3	154 ± 45	68.8 ± 25.9	1.78 ± 0.44	16

*data averaged without two data sets due to anomalous start or strain. Averages with excluded data are shown in brackets.

All data sets are shown in Figure 3.4.

As shown in Table 3.5, fibres spun from M9-HPy_2 were far more extensible than fibres spun from LB-HPy_2, regardless of the condition tested. The highest extensibility from LB-HPy_2 fibres (27.6 ± 6.1%) was only obtained after stretching fibres in a solvent bath. In contrast, M9-HPy_2 fibres could stretch over 75% of the original length, even without solvent exposure during post-spin stretching (Table 3.5), indicating an inherent extensibility in these wet-spun fibres.

For initial conditions other than 2× before H_2O, this gain of extensibility appeared to come at the cost of strength (Table A1). Fibres were also consistently larger than expected for these initial tests, with diameters >25 μm when stretched 2× compared to Chapter 2 results showing fibres between 15-20 μm. Later wet-spinning with the same setup gave fibres that were consistently smaller in diameter, and with correspondingly higher strength at failure, more in line with what was reported in Chapter 2. One noticeable difference was that there was an increased temperature of the room (> 30 ºC) during spinning, which could have contributed to this change in mechanical properties.

Surprisingly, fibre mechanical properties appeared to be less tunable for M9-HPy_2, with each post-spin stretching condition giving similar averages for fibre mechanical properties. Lack of post-spin induced changes in mechanical properties indicates that no significant structural rearrangement is occurring that would lead to enhanced mechanical properties. M9-HPy_2 fibres also appeared to be less consistent along the stretching profile (Figure 3.4), potentially indicative of smaller-scale structural reorganizations than inferred to occur at the yield point.

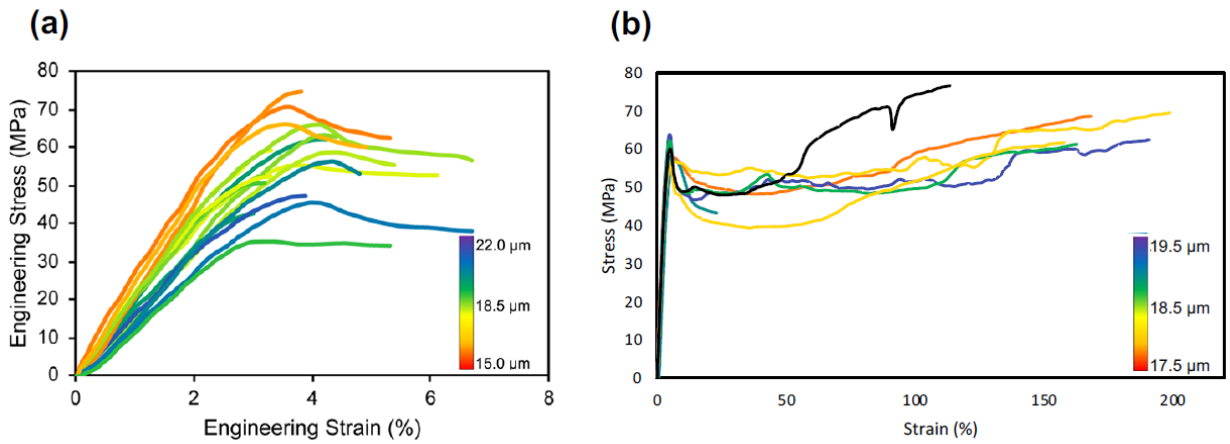


Figure 3.4. Comparison of mechanical properties for best 2× air stretched conditions for each expression method. Shown are fibres spun from the LB expressed protein (10% HPy_2, 90% EtOH coagulation bath; a) and from the M9 expressed protein ($^{15}N/^{13}C$ labels, 20% HPy_2, 95% EtOH coagulation bath; b). The black line in panel b indicates a fiber 13.6 µm in diameter. Panel (a) Reprinted by permission from ACS Biomaterials Science & Engineering[48] .

3.3.5. HPy_3 Expression

When expressing recombinant PySp1-based proteins in LB medium, the largest protein that could be expressed was HPy_2. This limitation was based on multiple truncation products relative to intact protein for HPy_3 as visualized by a InVision His-Tag In-Gel Stain using manufacturers specifications (Figure 3.5), thereby decreasing the purity of the full length HPy_3 protein that could be collected. While expression methods initially used for HPy_2 were found to be unsuccessful for HPy_3 expression, by employing supplementation scheme C7 for HPy_3 expression at 1.5× the amino acid concentrations used for HPy_2

(carried out by Skylah McLeod Van Wagoner), the first promising results for HPy_3 expression were obtained (Figure 3.6).

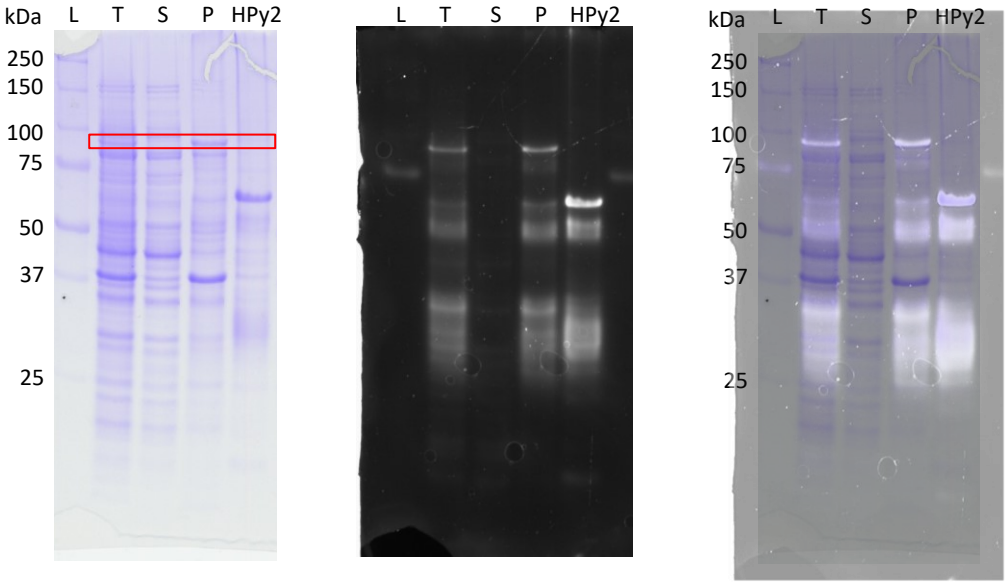


Figure 3.5. HPy_3 expression in LB (left), His_6 antibody staining (middle; InVision His-Tag In-Gel Stain, Invitrogen), and overlay (right). The full-length target protein is indicated in a red box, and L,T,S,P refer to the loaded protein ladder, total lysate, soluble lysate fraction and pellet lysate fraction respectively.

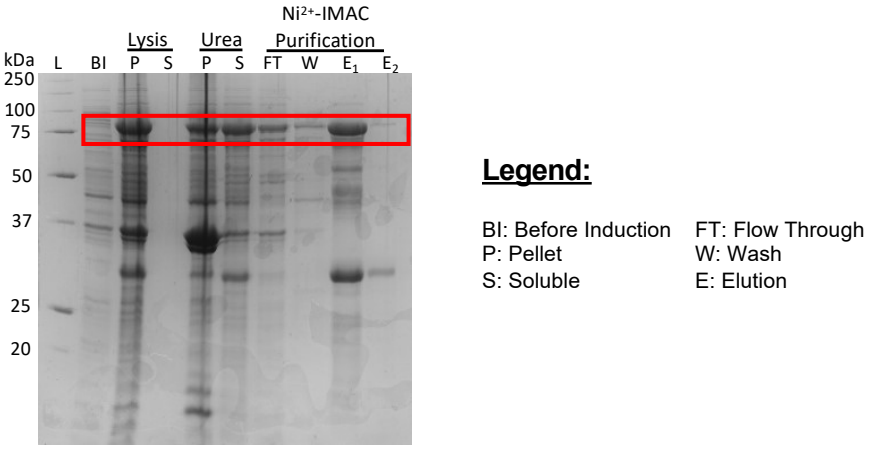


Figure 3.6. Expression and purification of HPy_3, based on C7 amino acid supplementation. The full-length target protein is indicated in a red box.

While the protein purified from this expression was not pure enough for fibre spinning, the major impurity was a single truncation instead of a smear of various truncations, giving a very promising method of HPy_3 expression. The major impurity migration is consistent with the truncation that was minimized during optimization of HPy_2 expression, indicating that either some additional amino acids and/or elevated amino acid concentrations may be needed relative to the C7 supplementation scheme. Knowing this, the supplementation scheme can be reoptimized for HPy_3 or, if unsuccessful, size exclusion chromatography should now be possible based on the difference in apparent molecular weight.

3.4. Summary

The work presented in this Chapter provides a method of expressing recombinant silk proteins at a six- to ten-fold increase in yield relative to expression in LB medium, without the need for specialized equipment such as a bioreactor. When combined with bioreactor-based techniques, this could further promote protein overexpression. Furthermore, this combination of amino acids was also found to decrease the number of protein truncations as observed by SDS-PAGE, increasing final protein purity relative to protein expressed in other media such as LB media or M9 with inadequate amino acid supplementation. This led to far more extensible fibres than observed previously[48] while maintaining mechanical strength in similar diameter fibres. Having now extensively characterized expression methods and mechanical properties for recombinant pyriform silk, I wanted to start investigating what underlying structures give rise to these mechanical properties. In Chapters 4 and 5, I approach this through a combination of techniques to probe the solution- and fibre-states.

Chapter 4

Backbone and Side Chain Chemical Shift Assignment and Secondary Structure Assessment

4.1. Introduction

Chapters 2 and 3 discussed the characterization of HPy_2 pyriform silk mechanical properties, with Chapter 2 focused on initial methodologies and Chapter 3 focused on fine-tuning expression techniques in a chemically-defined medium to increase expression yield. This led to lower levels of truncation products being observed during expression and also fibres that were both more extensible and more tolerant of water during fibre spinning. Having developed an understanding of how these fibres performed under tension and refining the methods of protein preparation in Chapters 2 and 3, I developed approaches to determine the atomic-level solution-state structure of this protein as presented in this Chapter. Atomic-level structural analysis provides a groundwork for understanding how the mechanical properties of pyriform silk fibres are obtained and is a fundamental starting point in elucidating the protein-fibre structural transition that occurs. This analysis was approached using NMR spectroscopy, requiring chemical shift assignment for ^{1}H, ^{13}C and ^{15}N nuclei as a function of position within the amino acid sequence and focusing on both full-length HPy_1 and various truncations of the protein in order to maximize solution-state NMR data quality and assignment completeness for the structured portion of HPy_1.

4.1.1. Protein NMR Spectroscopy

To understand the atomic-level structure of proteins under a given condition, modern NMR methods take advantage of isotopic enrichment of proteins during expression and observation of chemical shifts for different nuclei. One of the most common NMR experiments for probing proteins is the ^{1}H-^{15}N HSQC[70], which provides a two-dimensional ^{1}H-^{15}N correlation spectrum that represents all ^{1}H-^{15}N bonds within the protein. In the two-dimensional form, the resulting correlation spectrum pattern is highly sensitive to changes in secondary structure or local environment, so this spectrum is often referred to as the "fingerprint" of a protein.

To determine the structure of a protein, in the triple-resonance spectroscopy-based backbone walk approach of biomolecular NMR spectroscopy, a series of 3D experiments is also collected, which correlate ^{1}H, ^{13}C, and/or ^{15}N chemical shifts in highly specific and predictable manners[71]. In these experiments, covalently attached or proximal NMR active nuclei can be correlated by relying on characteristic J-coupling constants, allowing us to generate highly predictable multi-dimensional spectral correlation maps. The collection of experiments that probe residue *i* and residue $i_{+/-1}$, with experiments that only probe residue $i_{+/-1}$ creates connectivity between the observed chemical shifts and allows for sequential assignment in what is referred to as a backbone walk. Experiments that "see" the side-chain protons and carbons (HCCH-TOCSY[72], $^{15}N/^{13}C$-NOESY[73]) and protons attached to neighbouring carbon atoms (HCCH-COSY[72]) can then be used to assign side-chain peaks. A representation of the information gained from each experiment is shown in Figure 4.1.

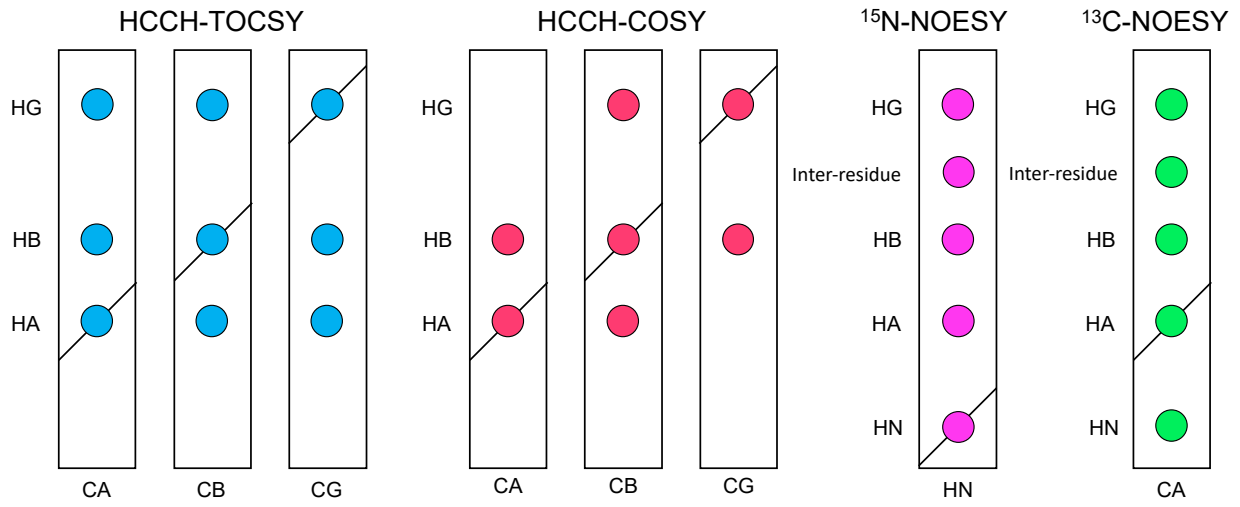


Figure 4.1. Diagram of representative correlations observed in experiments used for side-chain assignments. HCCH-TOCSY, HCCH-COSY, and ^{13}C-NOESY correlations are observed for all ^{1}H-^{13}C bonded nuclei, with the exception being prolines for HCCH-TOCSY and HCCH-COSY spectra.

In this chapter, I use these spectra to assign chemical shifts for nuclei in a truncated HPy$_1$ protein (Δ9-70+Δ206-242 HPy$_1$). These chemical shifts indicate potential secondary structure when compared to random coil chemical shifts for respective residues and, when combined with DANGLE restraints[74], generate chemical-shift based inferences of secondary structure. These can further be refined by assignment of inter-residue NOE contacts that are characteristic of secondary structure such as $i_{\pm3}$ and $i_{\pm4}$ contacts, as carried out in Chapter 5. This approach, along with assignment of medium- and long-range NOE contacts, allows for the calculation of the atomic-level solution state structure of a protein.

4.2. Materials and Methods

4.2.1. Plasmid Construction

HPy_1 (sequence as in Figure 4.1) was initially used as the target for NMR studies. To obtain plasmids for desired HPy_1 deletion products, reverse PCR was performed on the HPy_1 vector in pUC57, giving plasmids encoding two types of respective N- and C-terminal deletions (Δ9-60, Δ9-70, Δ196-242, Δ206-242), and a combination of the least structurally perturbing deletions as discussed in Section 4.3 (Δ9-70+Δ206-242). The protein sequences for each deletion construct are shown in Figures 4.2 and 4.3.

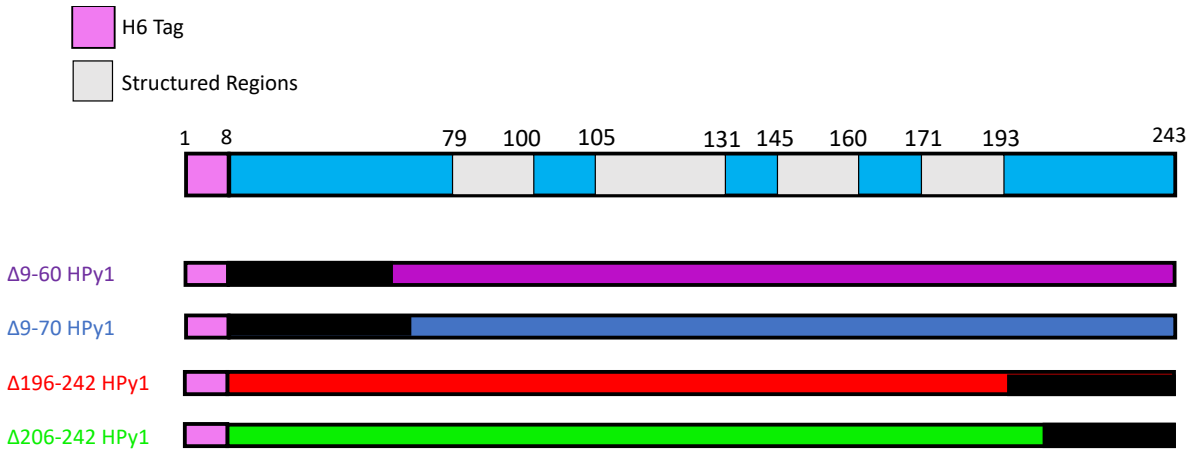


Figure 4.2. Schematic of HPy_1 and truncated HPy_1 proteins.

(a) HPy_1

```
  1  MGHHHHHHSS AQENSFTQSS VAQQSAVAQQ SSVSQQSSAA QQSSVAQSQQ TSYSAATNAG  60
 61  SSVSQSQAIV SSAPVYFNSQ TLTNNLASSL QSLNALNYVS NGQLSSSDVA STVARAVAQS  120
121  LGLSQGSVQN IMSQQLSSIG SGASTSSLSQ AIANAVSSAV QGSQAAAPGQ EQSIAQRVNS  180
181  AISSAFAQLI SQRTAPAPAP RPRPAPLPAP APRPRPAPAP RPAPVYAPAP VASQFQASAS  240
241  SQ®                                                               243
```

(b) Δ9-60

```
  1  MGHHHHHHSS VSQSQAIVSS APVYFNSQTL TNNLASSLQS LNALNYVSNG QLSSSDVAST  60
 61  VARAVAQSLG LSQGSVQNIM SQQLSSIGSG ASTSSLSQAI ANAVSSAVQG SQAAAPGQEQ  120
121  SIAQRVNSAI SSAFAQLISQ RTAPAPAPRP RPAPLPAPAP RPRPAPAPRP APVYAPAPVA  180
181  SQFQASAS®S                                                        191
```

(c) Δ9-70

```
  1  MGHHHHHHSS APVYFNSQTL TNNLASSLQS LNALNYVSNG QLSSSDVAST VARAVAQSLG  60
 61  LSQGSVQNIM SQQLSSIGSG ASTSSLSQAI ANAVSSAVQG SQAAAPGQEQ SIAQRVNSAI  120
121  SSAFAQLISQ RTAPAPAPRP RPAPLPAPAP RPRPAPAPRP APVYAPAPVA SQFQA®SQ  180
181  S                                                                 181
```

(d) Δ196-242

```
  1  MGHHHHHHSS AQENSFTQSS VAQQSAVAQQ SSVSQQSSAA QQSSVAQSQQ TSYSAATNAG  60
 61  SSVSQSQAIV SSAPVYFNSQ TLTNNLASSL QSLNALNYVS NGQLSSSDVA STVARAVAQS  120
121  LGLSQGSVQN IMSQQLSSIG SGASTSSLSQ AIANAVSSAV QGSQAAAPGQ EQSIAQRVNS  180
181  AISSAFAQLI SQRTAS                                                 196
```

(e) Δ206-242

```
  1  MGHHHHHHSS AQENSFTQSS VAQQSAVAQQ SSVSQQSSAA QQSSVAQSQQ TSYSAATNAG  60
 61  SSVSQSQAIV SSAPVYFNSQ TLTNNLASSL QSLNALNYVS NGQLSSSDVA STVARAVAQS  120
121  LGLSQGSVQN IMSQQLSSIG SGASTSSLSQ AIANAVSSAV QGSQAAAPGQ EQSIAQRVNS  180
181  AISSAFAQLI SQRTAPAPAP RPRPAS                                      206
```

(f) Δ9-70, Δ206-242

```
  1  MGHHHHHHSS APVYFNSQTL TNNLASSLQS LNALNYVSNG QLSSSDVAST VARAVAQSLG  60
 51  LSQGSVQNIM SQQLSSIGSG ASTSSLSQAI ANAVSSAVQG SQAAAPGQEQ SIAQRVNSAI  120
120  SSAFAQLISQ RTAPAPAPRP RPAS                                        144
```

Figure 4.3. HPy_1 protein sequences for each deletion tested.

4.2.2. Protein Expression, Purification, and NMR Sample Preparation

HPy_1 and truncated HPy_1 proteins were expressed in *E. coli* BL21(DE3) cells and purified similarly to methods described in Section 3.2.2, while HPy_2 was expressed and purified exactly as detailed in Section 3.2.2. without amino acid supplementation. For HPy_1 proteins, starter cultures grown overnight at 37 °C were transferred to 1.2 L of LB medium at 1:50 v/v and grown at 37 °C until an OD_{600} between 1.0-1.2 was reached. The cells were then pelleted (4000 rcf, 4 °C, 25 min) followed by 2.4:1 v/v transfer to M9 media supplemented with $^{15}NH_4Cl$ and glucose with or without ^{13}C-labelling, as appropriate for a given experimental target. Cells resuspended in modified M9 media were incubated with shaking at 37 °C for 1 h to allow uptake of the media components, followed by induction with 0.8 mM IPTG and expression for 2 h at 37 °C with shaking. Cells were lysed via French Pressure Cell in native lysis buffer, with soluble proteins were purified via Ni^{2+} IMAC (2 columns, 2.5-3 mL dry Ni^{2+}-NTA beads each) using native wash and elution buffers (QIAexpressionist protocol; Siegen, Hilden, Germany) at 1/10 the volumes described in Section 3.2.2 to match the dry bead volume. NMR samples were prepared in 50 mM potassium phosphate buffer, 1 mM DSS, and 1 mM NaN_3, exactly as described in Section 3.2.3. NMR samples were stored at 37-40 ºC (HPy_1) or at room temperature (Δ9-70, Δ206-242-HPy_1).

4.2.3. Data acquisition and Spectral Processing

Spectra were collected on a Bruker AVANCE III 11.7 T spectrometer at Dalhousie's NMR^3 Facility, a Bruker AVANCE III 16.4 T spectrometer at the National Research Council (HPy_1 triple resonance spectra used for chemical shift assignment; Halifax, NS), an

Bruker AVANCE III 18.8 T spectrometer at the MARS Facility in Toronto, Ontario (^{13}C rooted spectra which were collected and processed by Dr. Geneviève Seabrooke), and the 18.8 T Bruker Ascend spectrometer at the Québec/Eastern Canada High Field NMR Facility in Montréal, Québec (triple resonance spectra for Δ9-70+Δ206-242-HPy$_1$ which were collected by Dr. Tara Sprules, with all experiments except ^{15}N/^{13}C-NOESYs processed by me). Full data collection parameters are detailed in the Appendix (Tables A2 and A3). Spectra were processed using either NMRpipe[75] (using the NMRbox resource[76]) or TopSpin, and analyzed using CcpNMR Analysis 2.2.1[77]. If processed using TopSpin, spectra were indirectly referenced to DSS, while NMRpipe referencing to water as a function of temperature was used for all triple-resonance data collected with non-uniform sampling (NUS). As NMRpipe references spectra based upon the water chemical shift at a manually input temperature, which is highly sensitive to sample acquisition temperature[78], both referencing methods should give the same result. Data reconstruction for experiments acquired using NUS was performed through Iterative Soft Thresholding[79] for full length HPy$_1$, while SMILE[80] was used for NUS reconstruction of Δ9-70+Δ206-242-HPy$_1$ data.

4.2.4. Data Analysis and Validation

Chemical shifts were assigned following the backbone walk strategy described in section 4.1.1, starting with backbone residues on the full-length HPy$_1$, then carrying assignments over to Δ9-70+Δ206-242-HPy$_1$. The key experiments for each type of assignment are delineated in Table 4.1. Initial side chain ^{1}H assignment was attempted for HPy$_1$ based on characteristic chemical shifts in the ^{1}H-^{15}N HSQC NOESY spectra, with assignments

rigorously assigned using Δ9-70+Δ206-242-HPy$_1$. Through CcpNmr Analysis, chemical shifts from these assignments were used to predict secondary structure based on differences in chemical shift relative to random coil[81, 82] and DANGLE[74] calculated dihedral angle restraints.

Table 4.1. Key experiments used for backbone assignment and side-chain assignment.

Backbone	Side-Chain
^{1}H-^{15}N HSQC	^{1}H-^{13}C HSQC
^{1}H-^{15}N HSQC NOESY	^{1}H-^{15}N HSQC NOESY
HNCO	^{1}H-^{13}C HSQC NOESY
HN(CA)CO	HCCH-COSY
CBCA(CO)NH	HCCH-TOCSY
HNCACB	

Various quality control checks were performed to ensure that the data were analyzed and assigned correctly. Starting with backbone assignment quality, ^{1}H-^{15}N HSQC spectra collected from various selectively-unlabelled proteins in Chapter 3 and truncated proteins were compared with final residue assignments to ensure no errors were made in the identity of backbone assignments. Assignment quality of the backbone chemical shifts and data referencing were further validated using the PANAV algorithm[83]. For quality control of regions predicted to contain secondary structure, a {^{1}H}-^{15}N-Heteronuclear

NOE (Het-NOE) experiment was conducted on the full-length HPy_1 to determine which backbone residues are in rigid or dynamic regions.

4.3. Results

4.3.1. Determination of the Best Solution-State NMR Conditions to Use

To determine a standard set of conditions for triple-resonance NMR analysis – which requires sample stability for at least a several d–y period – I had to compare buffer, pH, and temperature for spectral collection. This was approached by comparing HPy_1 in three different buffer conditions (acetate buffer, pH 5.00 ± 0.05; potassium phosphate buffer, pH 6.0; and potassium phosphate buffer, pH 7.50 ± 0.05). This was designed to determine a pH that allowed for the resolution of sharp NMR signals and ensure the protein was not precipitating and was structurally unperturbed during the time it would take for collection of triple resonance experiments.

HPy_1 was found to rapidly precipitate in acetate buffer at pH 5.0 prior to protein concentration, indicating this condition was unsuitable for the timeframes required for triple-resonance experiments. At pH 7.5 in phosphate buffer, a much lower signal to noise ratio was observed in 1D proton experiments relative to the pH 6 sample, indicating the higher pH would not be amenable to 3D spectra collection (Figure 4.4). When in potassium phosphate buffer at pH 6, ^{1}H-^{15}N HSQC signals appeared well resolved with no changes over time [Figure 4.5(b)] and no obvious protein precipitation, indicating the protein was stable in this buffer and at this pH.

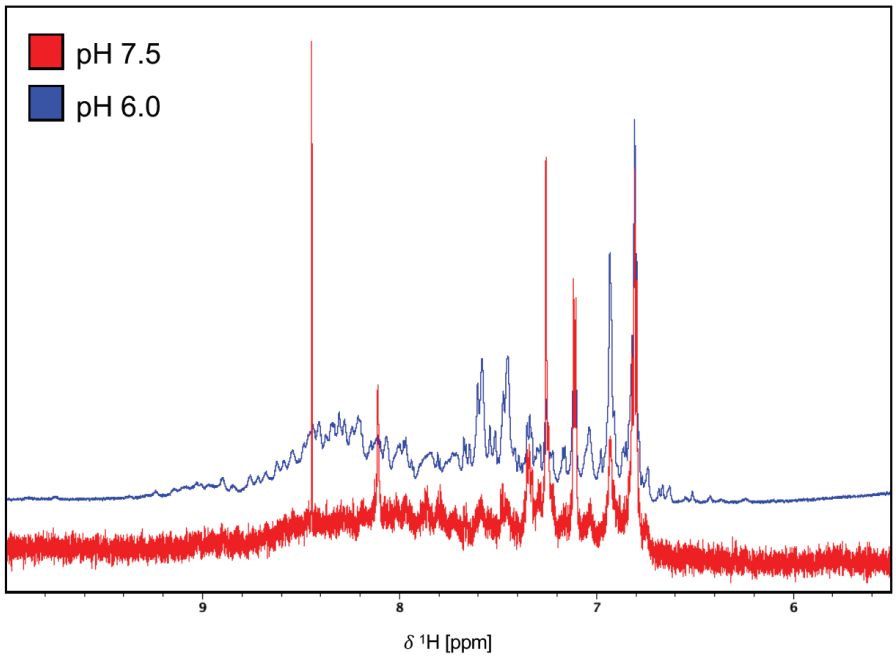


Figure 4.4. Comparison of HPy_1 1D ^1H NMR spectra 50 mM potassium phosphate buffer at pH 6.0 (blue) and pH 7.5 (red). Both samples were at ~150 µM each and the pH 7.5 condition is scaled at 13.45× relative to the pH 6.0 condition. Both were collected using the zgpr pulse program, with 16 scans and 32786 data points.

Proceeding to determine the optimal temperature in this buffer, after comparing ^1H-^{15}N HSQCs acquired at 30 ºC vs. 40 ºC, I noticed increased signal-to-noise for backbone peaks of highly shifted residues at 40 ºC, such as the ones at 9.8 ppm and 10.4 ppm in the proton dimension [Figure 4.5(a)]. Knowing that this would likely translate to a higher likelihood of assignment of these residues and seeing that the protein is highly stable at 40 ºC [Figure 4.3(b)], I used this temperature for all spectra collected to assign chemical shifts. Considering the size of HPy_1 (24.5 kDa), transverse relaxation optimized spectroscopy (TROSY) versions of the ^1H-^{15}N HSQC experiments were conducted to see if an enhanced resolution was obtained. No substantial improvement in spectral quality

was observed when using TROSY and a decrease in signal-to-noise was observed relative to the non-TROSY spectra (Figures 4.6 and 4.7), so I proceeded to collect data from non-TROSY versions of the experiments.

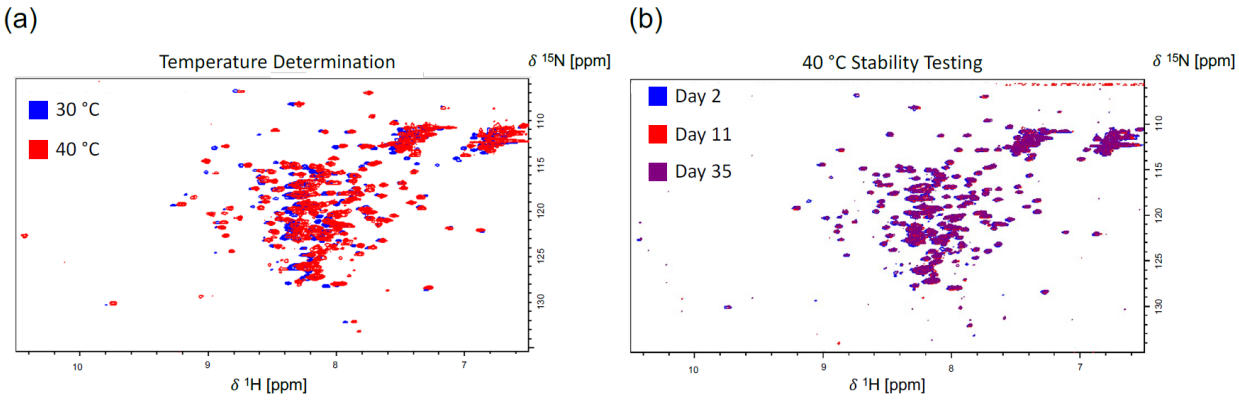


Figure 4.5. Comparison of ^{1}H-^{15}N HSQC experiments at different temperatures (a) and stability of protein with storage at 40 ºC (b). Spectra in (a) were manually offset to more clearly show signal enhancement at 40 ºC.

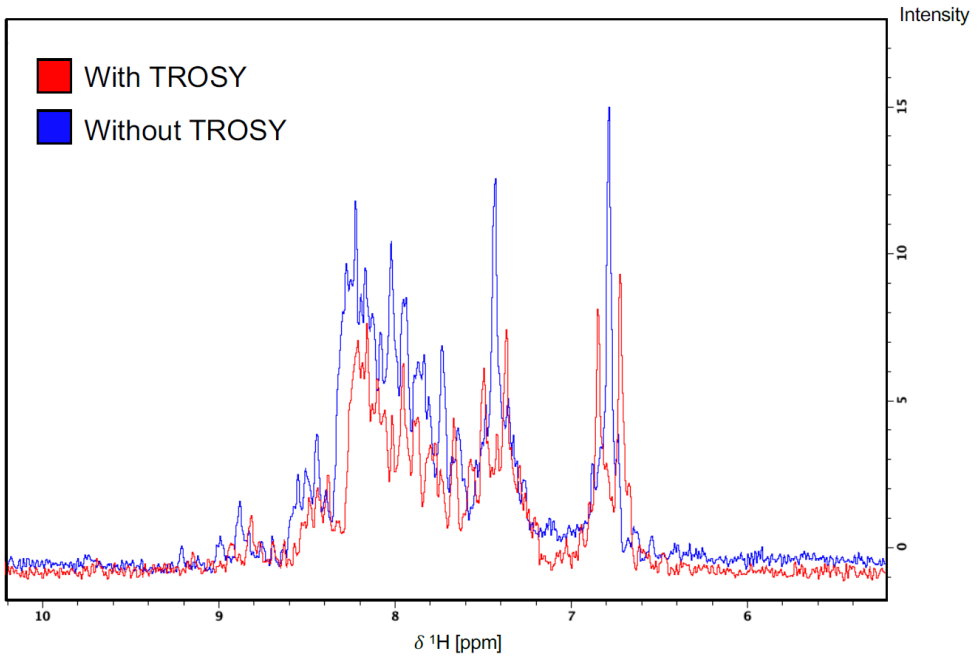


Figure 4.6. Overlay of the first increments of ^{1}H-^{15}N spectra of HPy_1 collected with (red) and without (blue) TROSY.

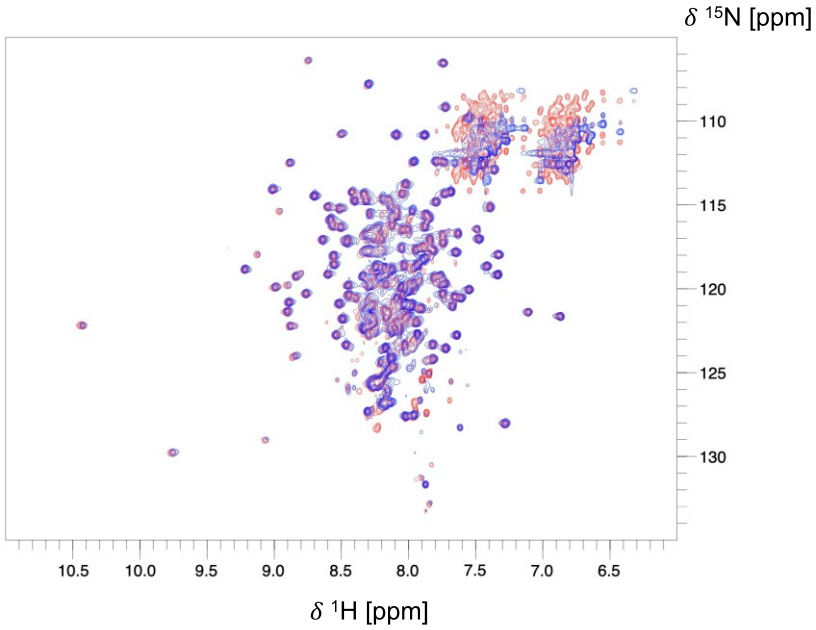


Figure 4.7. Comparison of ^{1}H-^{15}N HSQC TROSY (red) vs. standard (blue) HSQC experiments.

4.3.2. Testing HPy_1 as a Pyriform Repetitive Domain Model

Having found appropriate conditions for preparing HPy_1 and HPy_2 for NMR, the two protein constructs were compared to determine if adding extra repeat units led to a perturbation of the folded structure. To do this, ^{15}N-labelled HPy_2 proteins were expressed in M9 without amino acid supplementation, purified, and prepared for NMR spectra collection as discussed in Section 3.2. A ^{1}H-^{15}N HSCQ spectrum collected for HPy_2 was overlayed with HPy_1 to evaluate if differences are observed in residue chemical shifts. As shown in Figure 4.8, the profiles for these proteins overlay with each other with no extra ^{1}H-^{15}N correlations present, indicating no change has occurred from the one repeat to the larger protein.

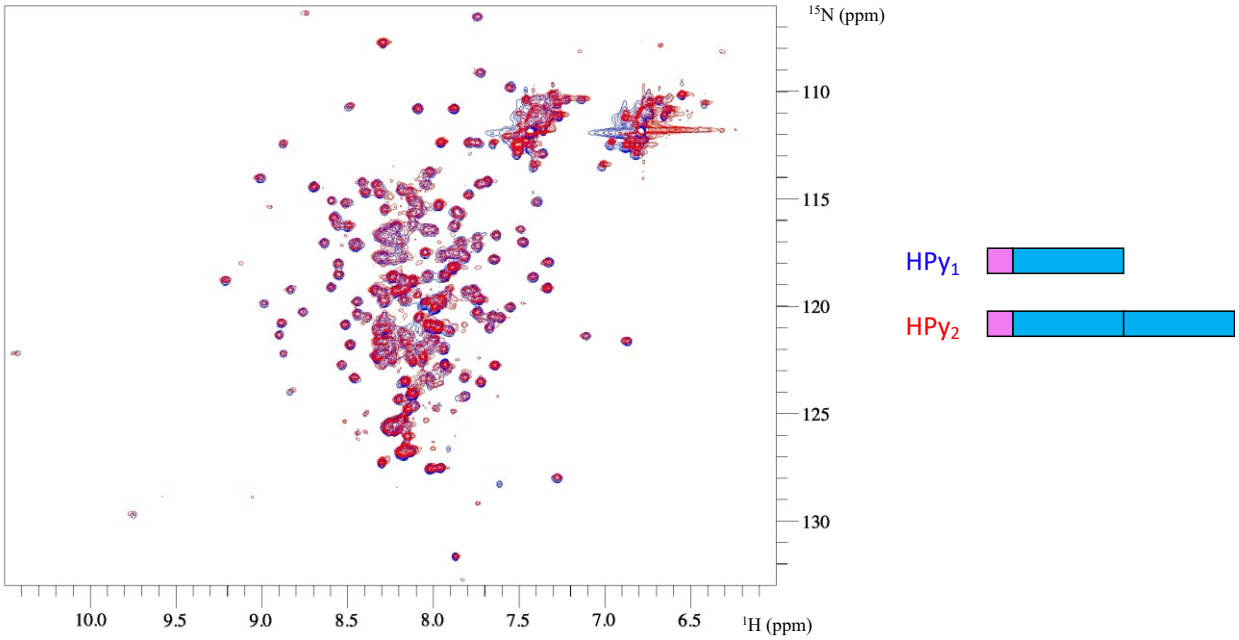


Figure 4.8. Overlay of HPy_1 and HPy_2 ^{1}H-^{15}N HSQC spectra.

4.3.3. Assignment and Challenges in Assigning the Full-Length HPy_1 Protein

Starting with the ^{1}H-^{15}N HSQC spectra (Figure 4.8), a combination of disperse and highly overlapped signals were observed, which were initially inferred to result from structured and disordered areas respectively in the protein. To assign which residues correspond to which peaks in the ^{1}H-^{15}N HSQC, sets of 3D spectra were analyzed using the backbone walk approach described in section 4.1.1. Specifically, the HNCACB and HN(CA)CO spectra were assigned for backbone chemical shifts, while the CBCA(CO)NH and HNCO spectra identification the i_{-1} residues for each assigned shift. By initially targeting residues with easily identifiable CA and CB chemical shifts[84] such as alanine, glycine, threonine, valine, and serine, correlations were built between chemical shifts of backbone residues and many of the backbone residues were able to be assigned as shown in Figure 4.9. Some notable exceptions were a highly disordered region which corresponded to Ala and Arg residues in the PX-rich motif, leading to an inability assign those residues (Table 2) and the side chain amides which were not visible in the collected 3D spectra. From these data, I assigned ~79% of backbone+H+HA chemical shifts (starting rom residue 10S) and preliminarily assigned 34% of assignable side chain protons based on ^{1}H-^{15}N HSQC NOESY chemical shifts (Table 4.2). From these assignments, I then calculated DANGLE-based dihedral angle restraints, and generated a secondary structure prediction for each residue in HPy_1 (Figure 4.10)

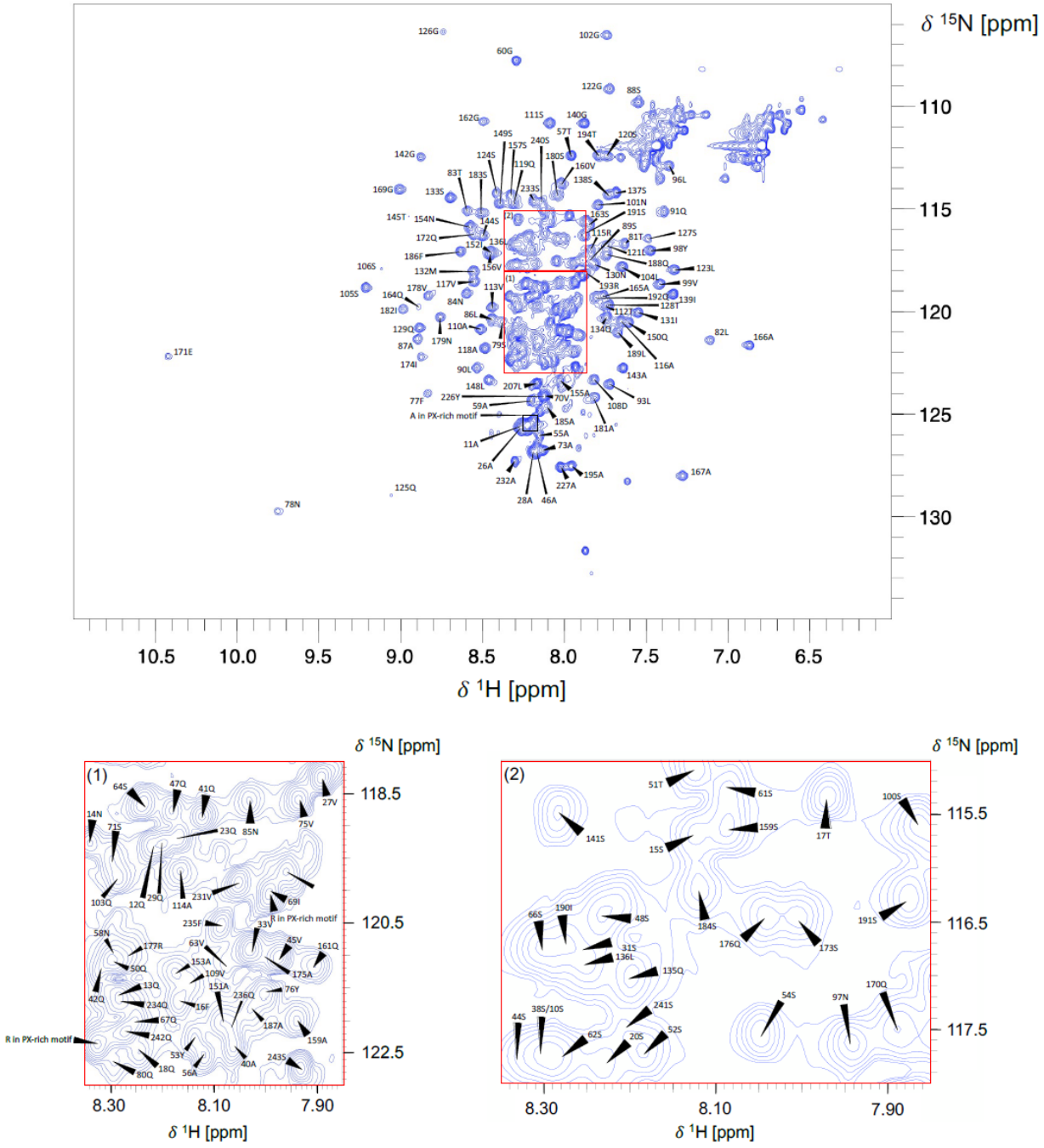


Figure 4.9. ^{1}H-^{15}N HSQC spectra of HPy$_1$ annotated with chemical shift assignments.

Table 4.2. Assignment statistics for HPy_1 starting at the first assignable residue, 10S.

Category	Available	Assigned	% Assigned	Min Shift	Mean Shift	Max Shift
Element C	973	572	58.8	16.3	92.6	181.2
Element H	1240	605	48.8	0.7	4.6	10.4
Element N	303	190	62.7	106.4	119.4	136.6
(Backbone)N+H	449	375	83.5	None	None	None
Backbone+H+HA	1159	918	79.2	None	None	None
Backbone	702	584	83.2	None	None	None
Side Chain H	783	271	34.6	0.7	2.3	8.0
Side Chain non-H	574	178	31.0	16.3	40.5	70.2
Type N	234	190	81.2	106.4	119.4	136.6
Type ND	11	0	0	None	None	None
Type NE	42	0	0	None	None	None
Type NH	16	0	0	None	None	None
Type H	215	185	86.0	6.9	8.1	10.4
Type HA	242	149	61.6	3.7	43	4.8
Type HB	376	163	43.4	1.3	2.6	4.3
Type HG	203	74	36.5	0.8	1.6	2.5
Type HD	115	28	24.3	0.7	1.4	8.0
Type HE	85	6	7.1	7.1	7.6	7.8
Type HZ	4	0	0	None	None	None
Type C	234	193	82.5	173.5	176.6	181.2
Type CA	234	201	85.9	45.0	58.0	67.4
Type CB	226	178	78.8	16.3	40.5	70.2
Type CG	152	0	0	None	None	None
Type CD	102	0	0	None	None	None
Type CE	9	0	0	None	None	None
Type CZ	16	0	0	None	None	None

Table 4.2. Continued

Category	Available	Assigned	% Assigned	Min Shift	Mean Shift	Max Shift
Residue Ala	44	35	79.5	None	None	None
Residue Arg	8	3	37.5	None	None	None
Residue Asn	11	10	90.9	None	None	None
Residue Asp	1	1	100	None	None	None
Residue Gln	34	30	88.2	None	None	None
Residue Glu	2	2	100	None	None	None
Residue Gly	8	8	100	None	None	None
Residue Ile	7	7	100	None	None	None
Residue Leu	12	12	100	None	None	None
Residue Met	1	1	100	None	None	None
Residue Phe	4	4	100	None	None	None
Residue Pro	19	8	42.1	None	None	None
Residue Ser	54	52	96.3	None	None	None
Residue Thr	8	8	100	None	None	None
Residue Tyr	4	4	100	None	None	None
Residue Val	17	16	94.1	None	None	None
All Residues	234	201	85.9	None	None	None
Spin Systems	261	255	97.7	None	None	None

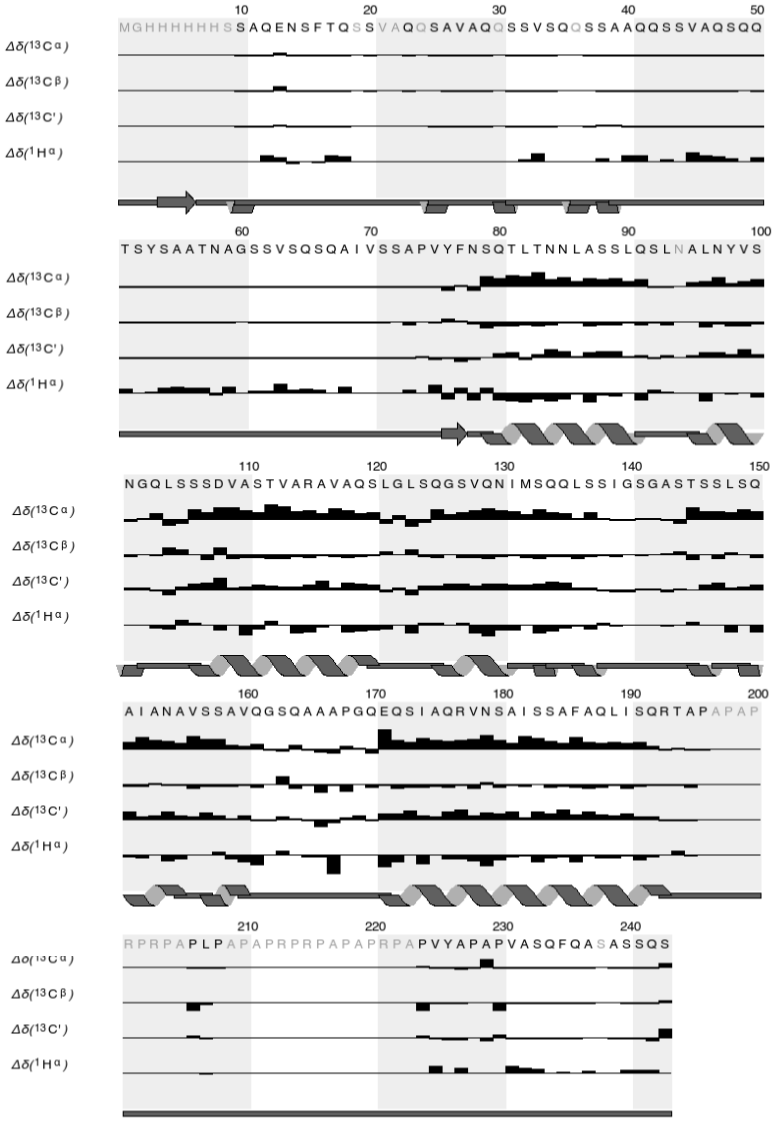


Figure 4.10. Chemical shift index (CSI)[81] histogram for HPy$_1$. This illustrates the predicted secondary structure of Δ9-70+Δ206-242-HPy$_1$ based on HA**,** CA, CB, and C' chemical shifts and DANGLE restraints. CSI plots and DANGLE generated secondary structure prediction were generated through CcpNmr Analysis[74, 77].

From the assigned chemical shifts and calculated dihedral angle restraints, HPy_1 was predicted to contain 5-6 α-helices and long, disordered regions at the N- and C-termini (~75 and ~40 residues respectively). This is in agreement with high chemical shift overlap observed in 2- and 3D spectra for residues in these regions (Figure 4.12) and inferences from a {^{1}H}-^{15}N-Heteronuclear NOE spectra[85] (Figure 4.11). In the collected spectra, most of the region between the predicted first and last helix including areas with no predicted structure (e.g., 142G) are found to be assignable to residues inferred to be in more rigid environments, suggesting this region is highly structured while the overlapped N-and C-terminal regions are more flexible and likely disordered as indicated in the annotated Figure 4.11.

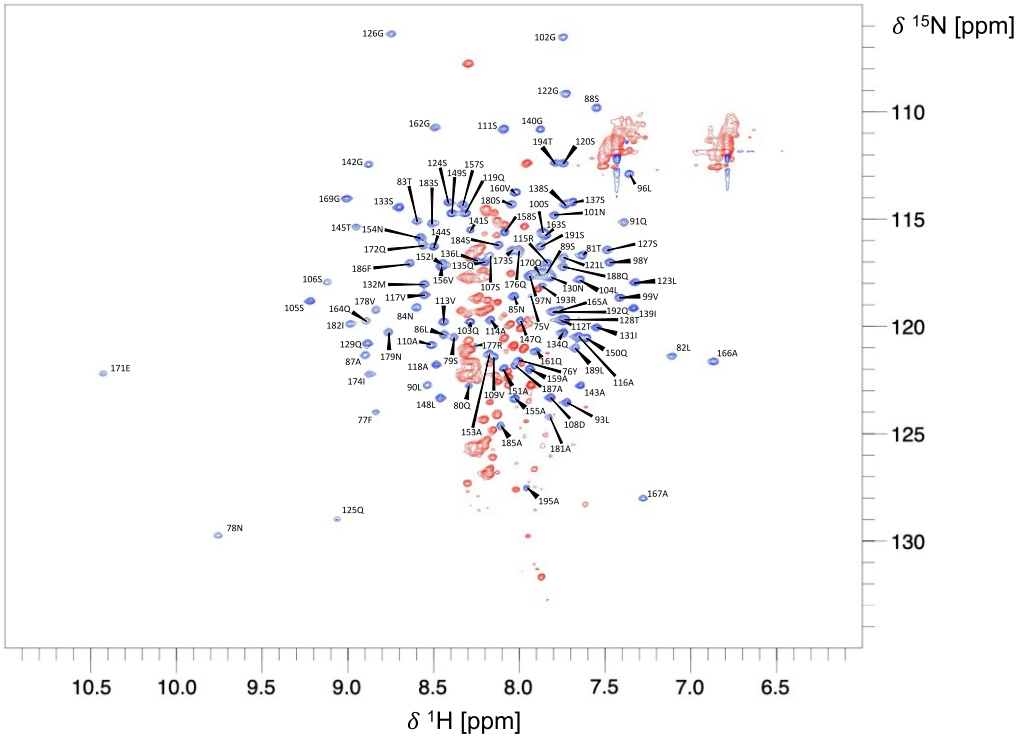


Figure 4.11. {^{1}H}-^{15}N-Heteronuclear NOE spectra, indicating residues in rigid (blue) and dynamic (red) regions. Residues indicated to be in more rigid regions are annotated.

While much of the HPy_1 backbone was assigned using the HNCACB, HN(CA)CO CBCA(CO)NH, HNCO and ^{1}H-^{15}N HSQC NOESY, significant challenges were encountered that impeded further chemical shift assignments. This difficulty can be linked to the size of the protein, the disordered N-terminal region, and the long, proline-rich region near the C-terminus. Combined, these three factors led to overlapping resonances in all data sets, the inability to observe Gln and Asn side chain peaks in triple resonance data, and the inability to readily resolve peaks in the ^{1}H-^{15}N NOESY spectrum (Figure 4.12).

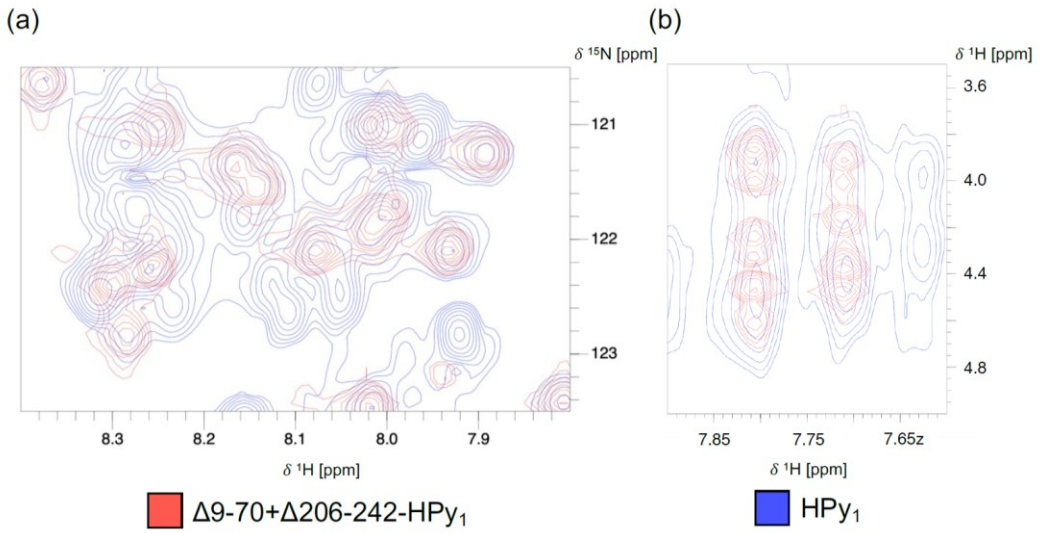


Figure 4.12. Challenges in assigning atoms in HPy_1 relative to Δ9-70+Δ206-242-HPy_1 as demonstrated by overlapped resonances in the ^{1}H-^{15}N HSQC (a) and broadening of peaks in the ^{1}H-^{15}N HSQC NOESY (b).

Beyond disorder and size, proline residues cannot directly be observed in the data collected due to experimental reliance on the amide proton, which prolines in the protein would not have. Due to this limitation of the collected spectra, ^{13}C-rooted experiments

were carried out to determine if this region was disordered, or potentially formed a polyproline II helix. These ^{13}C-rooted spectra are analyzed analogously to the previously mentioned sets of spectra, except the correlations along the backbone are between residues *i* and i_{+1}. Within these experiments, residues in the PX-rich motif were still unable to be readily assigned due to high overlap (Figure 4.13). Taking all of the mentioned difficulties of data analysis into account, from here we moved to truncated versions of the HPy_1 protein.

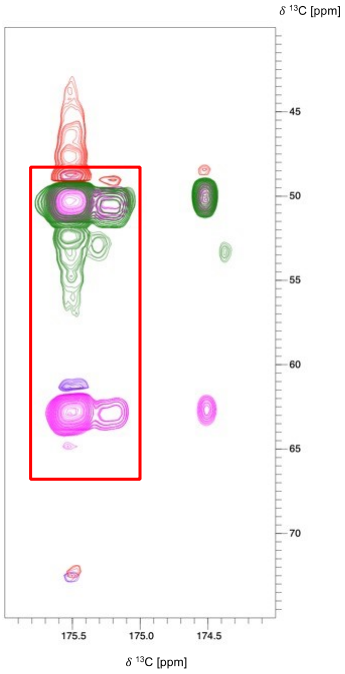


Figure 4.13. Illustration of overlap in ^{13}C-rooted spectra. ^{13}c-rooted HCaNCo (pink for positive peaks, purple for negative peaks) and HCaCoN (green for positive peaks and red for negative peaks) showing one of the overlapped proline resonances at 175.5 ppm (Add Ala/Arg).

4.3.4. Protein Truncations for Further NMR Assignments

Based on structure predicted from chemical shift assignments, two N- and C-terminal truncations were attempted to test how close the protein can be truncated to the core structured region: Δ9-60, Δ9-70, Δ196-242, Δ206-242. ^{1}H-^{15}N HSQC spectra showed that Δ9-60, Δ9-70, and Δ206-242 were non-perturbing to the sequentially closest helix-associated residues (79S for N-terminal truncations and 193R for C-terminal truncations), indicating that any of the three would be good candidates for residue truncations. Conversely, Δ196-242 led to a chemical shift perturbation for 193R, indicating this was too close to the start of the last helix to be a suitable truncation. These results are summarized in Figure 4.14.

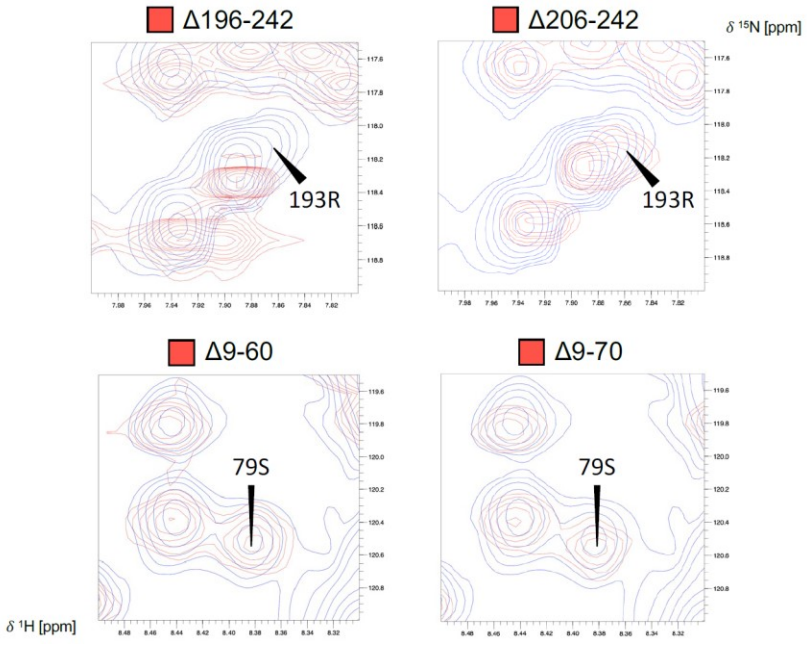
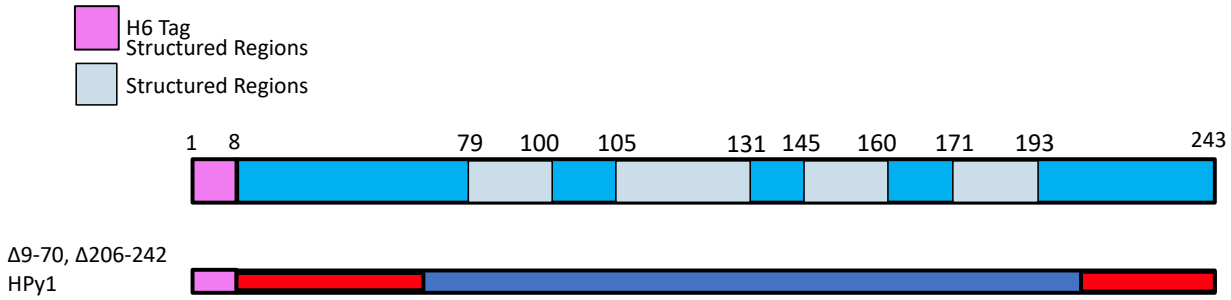


Figure 4.14. ^{1}H-^{15}N HSQC overlays of respective truncations and full-length HPy1, blow out depicting the sequentially closest structured residue. Full length HPy_1 is shown in blue for each spectrum, with the closest residue relative to the truncation that is predicted to have structure annotated.

Based upon these results, aΔ9-70+Δ206-242-HPy1 was designed (Figure 4.11) and expressed, as it appeared this combination would narrow the protein down to the structured core without perturbing structured residues. The doubly-truncated protein is diagramed in Figure 4.15. Comparing HSQC spectra showed no structural perturbation through the application of both truncations in theΔ9-70+Δ206-242-HPy1 protein and much more dispersed peaks (Figure 4.16). Compared to the full-length HPy_1, Δ9-70+Δ206-242-HPy_1 was found to have more uniform-looking peaks in 3-dimensional data sets and have much less overlap in the ^{1}H-^{15}N HSQC-NOESY [Figure 4.12(b)], allowing for much cleaner assignments and visualization of more peaks than initially seemed present.




```
  1  MGHHHHHHSS AQENSFTQSS VAQQSAVAQQ SSVSQQSSAA QQSSVAQSQQ   50
 51  TSYSAATNAG SSVSQSQAIV SSAPVYFNSQ TLTNNLASSL QSLNALNYVS   100
101  NGQLSSSDVA STVARAVAQS LGLSQGSVQN IMSQQLSSIG SGASTSSLSQ   150
151  AIANAVSSAV QGSQAAAPGQ EQSIAQRVNS AISSAFAQLI SQRTAPAPAP   200
201  RPRPAPLPAP APRPRPAPAP RPAPVYAPAP VASQFQASAS SQS          243
```

Figure 4.15. Δ9-70+Δ206-242-HPy_1 schematic compared to the full-length HPy_1 (top) and deletions shown in red on the HPy_1 sequence (bottom).

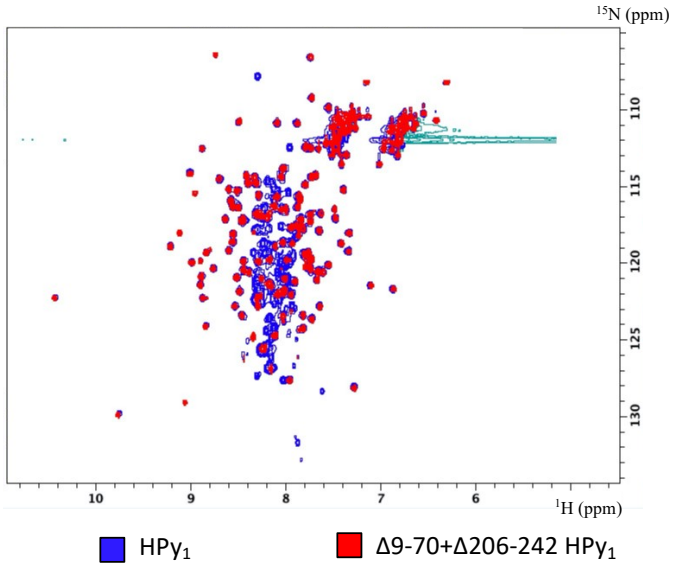


Figure 4.16. Overlay of ^{1}H-^{15}N HSQC chemical shift fingerprints of HPy1 andΔ9-70+Δ206-242-HPy1.

4.3.5. Assignment of Δ9-70+Δ206-242-HPy$_1$

While backbone assignments were transferred from the HPy$_1$ (Figure 4.8, Table A4) to the Δ9-70+Δ206-242-HPy$_1$ spectra (Figure 4.17), side chain assignments were rigorously assigned using the newly collected NMR spectra of Δ9-70+Δ206-242-HPy$_1$. Specifically, side chain carbon and proton chemical shifts were assigned by using observed ^{1}H-^{15}N HSQC NOESY signals and assigned CA and CB chemical shifts, and navigating to these correlations in the HCCH-TOCSY and HCCH-COSY spectra. By observing which signals are present in the COSY and TOCSY spectra, and navigating along the ^{13}C chemical shift dimension to find when these sets of ^{1}H and ^{13}C correlations are present for a given residue, this allows for robust assignment of each side chain chemical shift, and identification of signals corresponding to intra- vs. inter-residue correlations. These assignments, if correct, should also correspond to correlations in the ^{1}H-^{13}C HSQC and

^{1}H-^{13}C HSQC NOESY spectra. This was found to be the case, and most ^{1}H-^{13}C HSQC signals were assigned with unassigned signals believed to be linked to unassignable residues in the Δ9-70+Δ206-242-HPy$_1$ protein (e.g., proline side chain carbons). From here forward, any mention of residue position, unless otherwise specified, refers to Δ9-70+Δ206-242-HPy$_1$ and not the full-length protein.

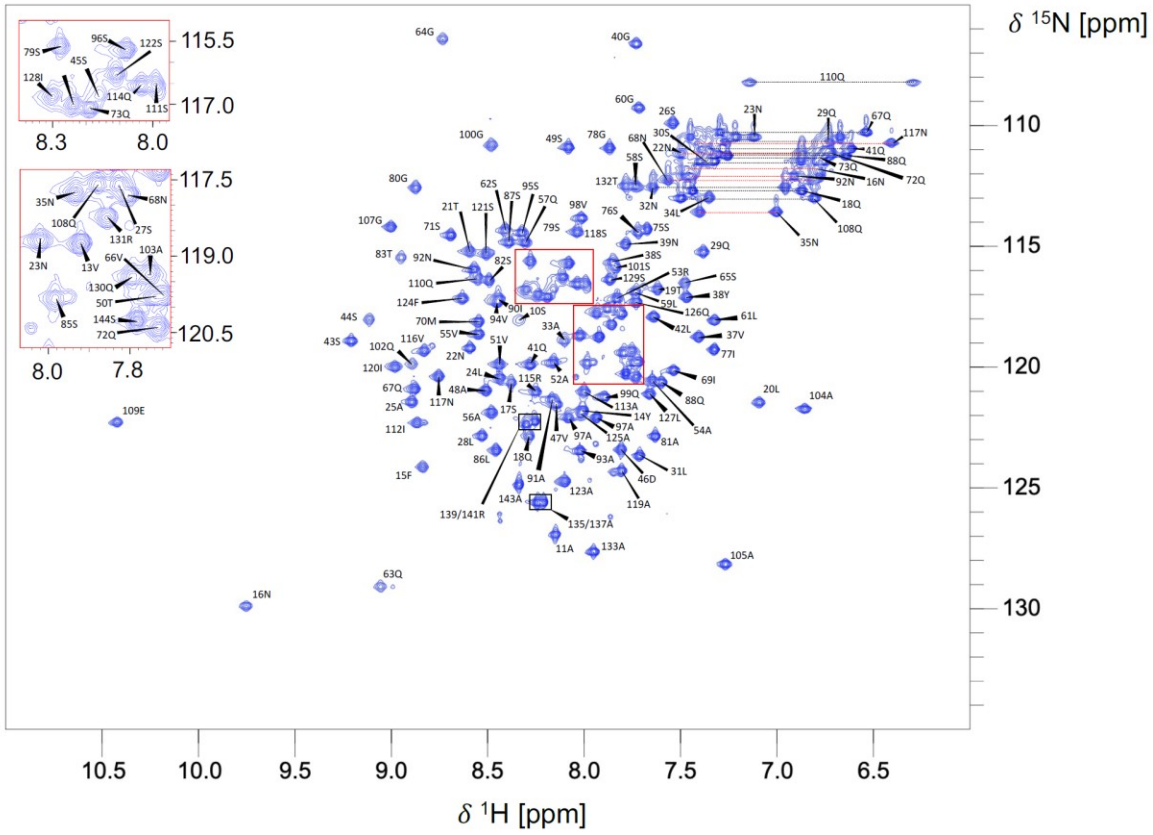


Figure 4.17. ^{1}H-^{15}N HSQC spectra of Δ9-70+Δ206-242-HPy$_1$ annotated with chemical shift assignments. Dotted lines represent assignments for Asn (red) and Gln (black) side chain amide pairs.

Full atom assignments are presented in Table A4. Combining all assignments mentioned so far and further data cleanup in the process of assigning inter-residue NOE contacts as

discussed in Chapter 5, starting at 10S I assigned 97% of backbone+HA+H atoms, 82% of side chain protons and 80% of other assignable side chain atoms in Δ9-70+Δ206-242-HPy_1 (Table 4.3). Assignments and chemical shift referencing were supported by PANAV validation. PANAV indicated chemical shift offsets of -0.30 ppm, -0.47 ppm, 0.14 ppm, and 0.50 ppm for C', CA, CB, and N respectively, which are reasonable within error when considering CA and CB are both assigned on the HNCACB spectra. CONA scores generated during PANAV analysis were 99.25% and 100% for 3-residue scans and 4-6-residue scans respectively, with the only flagged assignment in the 3-residue scan being 109E CA. All chemical shifts assigned for E109 were higher or lower than average for standard glutamic acid residues (Table A4), but the assignment was correctly rooted in the backbone walk data and consistent with Sanger sequencing results of the plasmid and the number of observed ^{1}H and ^{13}C chemical shifts during the side-chain assignment.

Table 4.3. Assignment statistics for Δ9-70+Δ206-242-HPy$_1$ starting at first assigned residue, 10S.

Category	Available	Assigned	% Assigned	Min Shift	Mean Shift	Max Shift
Element C	563	519	92.2	12.9	83.4	181.3
Element H	713	629	88.2	0.5	4.2	10.4
Element N	175	144	82.3	106.4	117.7	129.9
(Backbone)N+H	263	254	96.6	None	None	None
Backbone+H+HA	675	658	97.5	None	None	None
Backbone	405	395	97.5	None	None	None
Side Chain H	443	366	82.6	0.5	2.8	7.6
Side Chain non-H	333	268	80.5	12.9	51.3	180.5
Type N	135	127	94.1	106.4	118.6	129.9
Type ND	9	8	88.9	110.5	111.8	113.6
Type NE	21	9	42.9	108.2	111.0	113.0
Type NH	10	0	0	None	None	None
Type H	128	127	99.2	6.9	8.1	10.4
Type HA	142	136	95.8	3.3	4.1	4.8
Type HB	213	190	89.2	0.5	2.7	4.6
Type HG	112	94	83.9	0.5	1.7	2.6
Type HD	74	58	78.4	0.5	3.4	7.6
Type HE	42	22	52.4	6.3	7.0	7.5
Type HZ	2	2	100	7.0	7.1	7.2
Type C	135	133	98.5	173.5	177.1	181.3
Type CA	135	135	100	45.2	58.4	67.5
Type CB	128	126	98.4	16.3	39.5	70.2
Type CG	90	75	83.3	17.6	42.7	177.8
Type CD	61	44	72.1	12.9	66.2	180.5
Type CE	5	4	80	118.0	124.4	130.8
Type CZ	9	2	22.2	128.4	129.0	129.6

Table 4.3. Continued

Category	Available	Assigned	% Assigned	Min Shift	Mean Shift	Max Shift
Residue Ala	23	23	100	None	None	None
Residue Arg	5	5	100	None	None	None
Residue Asn	9	9	100	None	None	None
Residue Asp	1	1	100	None	None	None
Residue Gln	16	16	100	None	None	None
Residue Glu	1	1	100	None	None	None
Residue Gly	7	7	100	None	None	None
Residue Ile	6	6	100	None	None	None
Residue Leu	11	11	100	None	None	None
Residue Met	1	1	100	None	None	None
Residue Phe	2	2	100	None	None	None
Residue Pro	7	7	100	None	None	None
Residue Ser	30	30	100	None	None	None
Residue Thr	5	5	100	None	None	None
Residue Tyr	2	2	100	None	None	None
Residue Val	9	9	100	None	None	None
All Residues	135	135	100	None	None	None
Spin Systems	312	159	51.0	None	None	None

Secondary structure predictions for fully assigned Δ9-70+Δ206-242-HPy_1 (including NOE contacts discussed in Chapter 5) indicated the protein contains 6 α-helical regions (17S-30S, 32N-37S, 44S-58S, 63Q-76S, 83T-98V, and 109E-131R), with 17S and 131R (79S and193R in HPy_1) being the start and end residues for the first and last helices, respectively (Figure 4.18). The starting and ending residues for the first and last helix are consistent with early secondary structure predictions for the full-length HPy_1, indicating the original predictions were representative of the structured core.

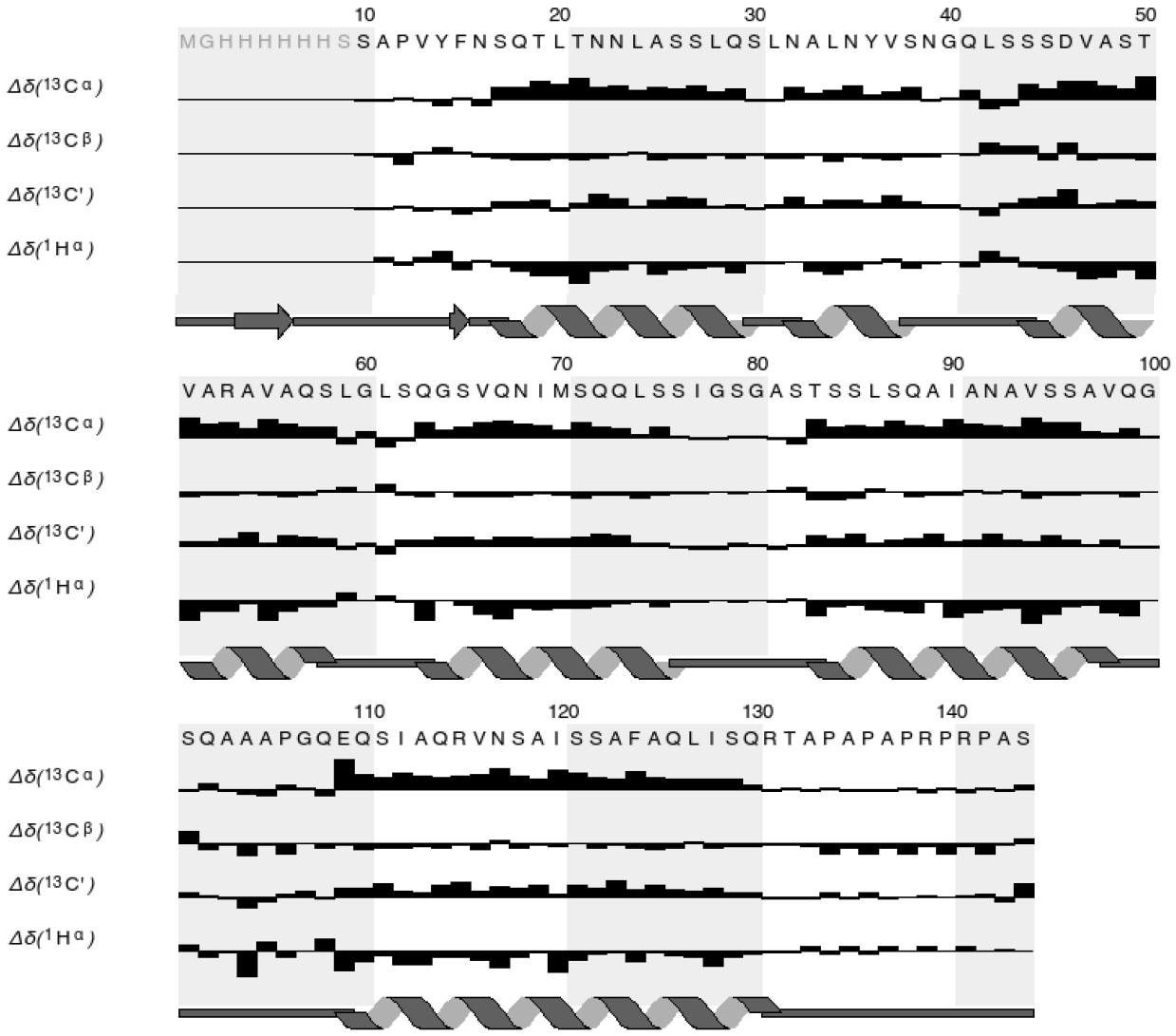


Figure 4.18. CSI[81] histogram for Δ9-70+Δ206-242-HPy$_1$. This illustrates the predicted secondary structure of Δ9-70+Δ206-242-HPy$_1$ based on HA**,** CA, CB, and C' chemical shifts and DANGLE restraints. CSI plots and DANGLE generated secondary structure prediction were generated through CcpNmr Analysis[74, 77].

4.4. Discussion

In the work discussed in Chapter 4, I was able to successfully assign the backbone of full length HPy_1, show that the repetitive domain units of *A. argentata* PySp1 are modular, and identify a truncated version of HPy_1 that allowed for near complete atom assignment in the core structured region. Spectra of wild-type HPy_1 were found to be challenging to resolve, particularly in the highly repetitive Gln-Ser rich motif near the N-terminus and the PX-rich motif near the C-terminus. Still, HPy_1 was found to serve as a model for the central repetitive domain of *A. argentata* PySp1 based on the lack of chemical shift perturbation in HPy_2. From assignments of the HPy_1 N- and C-terminal motifs, it was found that the Gln-Ser rich motif and the PX-rich motif were potentially disordered based on CSI and DANGLE bioinformatic analysis and high overlap of NMR chemical shifts in these regions. After removing these two regions, backbone amide resonances were unperturbed, indicating no interaction between the N- and C-terminal tails and the rest of the protein, and that the doubly-truncated protein was still representative of the central repetitive domain. This indicates the N- and C- terminal motifs likely act as spacer units between core structured regions in each central domain repetitive unit, having no obvious interactions with the structured residues and preserving structure between repeats. Alternatively, the C-terminal region could be forming a polyproline-II helix that is loosely packed, meaning that it does not interact (in a stable manner) with the core protein structure, while the N-terminal disordered tail potentially acts as a spacer between repeat units. The strongest support for PP-II structure in this region is the CD line shape for HPy_2 (Figure 2.10), which has a decreased mean residue ellipticity at 195 nm and 220 nm relative to what is expected from mean residue ellipticity at 195[86]. While the data were

collected at pH 6.0 vs the previously collected CD spectra, results obtained by Skylah McLeod van Wagoner, collected under my supervision, showed a more characteristic α-helical spectra[57] for Δ9-70+Δ206-242-HPy$_1$, giving further support to this theory (Figure 4.19). Regardless of the structural appearance in the C-terminal region, lack of interaction with the structured core was apparent with no perturbation in the truncated protein.

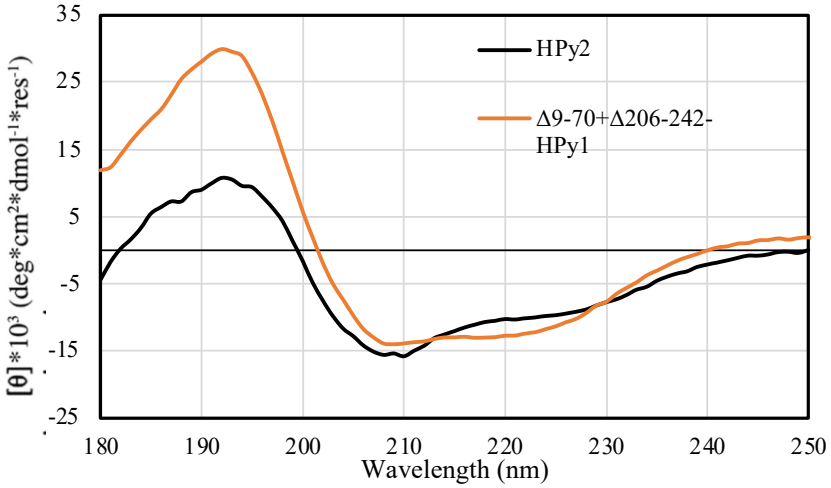


Figure 4.19. Overlay of CD spectra for HPy2 (potassium phosphate buffer, pH 7.5) and Δ9-70+Δ206-242-HPy$_1$ (potassium phosphate buffer, pH 6.0).

Examining chemical shift assignments showed particularly abnormal values for E109, which appeared to be correctly assigned but shifted relative to glutamic acid in a structured or disordered environment. With no indication of mutation in the DNA sequence, a high CONA score indicating the location in the backbone assignment is correct, and the number of observed peaks being reasonable for glutamic acid, this was indicative of a residue in a highly perturbing environment such as proximity to an aromatic residue or to multiple charged side chains that could be lead to a positive or negative electrostatic

environment. Further insight was obtained through solving the high-resolution NMR-derived structure of the protein, which will be discussed in Chapter 5.

Compared to other reported repetitive domain structures for TuSp1, MaSp1, and AcSp1, the secondary structure of Δ9-70+Δ206-242-HPy_1 appears to be most similar to the TuSp1 structures[43] in the core, containing 6 helices that are separated by short (<10 amino acid) segments, with a long (>20 amino acid), C-terminal helix. The reported MaSp1 structure (10.2210/pdb2MQA/pdb) shows this same pattern, but with a much shorter C-terminal disorder region that's only 13 residues long. This short, disordered region, assuming structure is conserved between repeats, could indicate a difference in behaviour of that C-terminal helix. While the helical regions of the TuSp1 and MaSp1 proteins appear to correlate with my inferred secondary structuring for the HPy_1 protein, they both contain short (<20 amino acid) disordered regions at the N- and C-termini rather than long, disordered regions. In this regard, the HPy_1 protein appears similar to the repetitive domain of the *Argiope trifasciata* aciniform silk previously solved in our group[29], which contains a short N-terminal linker, but a long (>60 amino acid) C-terminal linker. This longer C-linker indicates that, while the core of HPy_1 is potentially more tightly clustered than that of the W_1 unit, there could still be some similarity in the fibre forming process in relation to the role and rearrangement of linker regions. To better understand how the HPy_1 structure compares to other solved silk protein structures, in the next Chapter I focus on solving the high-resolution NMR-derived structure of the protein, as well as the atomic-level through micron-level structural features of pyriform silk fibres to gain more insight into the spun fibres and the transformation from the soluble state to the insoluble silk fibre.

Chapter 5

Atomic-Level Structure of HPy_1 and Fibre-State Structural and Hierarchical Assembly Characterization

5.1. Introduction

In Chapter 4, I described the near complete backbone and side-chain chemical shift assignment of HPy_1 and the systematic approach where I designed and moved to the smaller, structurally-conserved Δ9-70+Δ206-242-HPy_1. This construct allowed me to assign nearly all of the ^{1}H, ^{13}C and ^{15}N chemical shifts and to, in turn, assign intra- and inter-residue ^{1}H-^{1}H NOE contacts, which I will use in this chapter to determine the atomic-level structural ensemble of Δ9-70+Δ206-242-HPy_1. Temperature-based denaturation studies were conducted to develop an understanding of what regions of the protein were initially perturbed as an indication of what regions might be unfolded first during fibre formation. Lastly, solid-state NMR spectroscopy was applied to provide insight into the structural transition that takes place from the soluble to fibre states, and fibre microstructures were studied via microscopy.

5.1.1. Atomic-Level Structural Ensemble Calculation using Solution-State NMR data

A variety of programs have been developed and fine-tuned to calculate atomic-level structures of proteins from solution-state NMR restraints, some of the most popular programs being Crystallography and NMR System (CNS)[87], XPLOR-NIH[88], ARIA2[89]

(hereafter referred to as ARIA), and CYANA[90]. In my work, the chemical shift lists, distance restraints, and dihedral angle restraints were input into ARIA, which uses CNS[87] and a simulated annealing strategy[91] to perform structural calculations. ARIA was used based on its ability to assign unknown peaks in the spectra and because NOE restraints, dihedral angle restraints, and picked but incompletely assigned NOESY spectral peak data from CcpNmr Analysis can be input directly into ARIA for structural calculation setup. In this process, ARIA starts by using the chemical shift list to derive all possible NOE cross-peaks. Next, peak volumes are used to calculate distance restraints, which assumes single NOE contributions per peak, thereby relating volume as a function of $1/r^6$ between the two contributing nuclei[92]. Ambiguous assignments (assignments that can be caused by more than one NOE contact) are converted into ambiguous distance restraints, thereby allowing for the potential of all assigned members in calculated distance restraints. Following each ensemble calculation, carried out using CNS through a set of ARIA-produced input files, ARIA then filters out unlikely assignments and noise peaks[91] iteratively based upon a user defined cut-off number of structures (i.e., the lowest energy x structures of y structures that were calculated in the ensemble). This is used to refine the distance restraint list for the next round of structural calculation in CNS, producing another ensemble. After calculating these structural ensembles and iteratively refining restraints, the lowest energy structures from the final iteration can be water refined[93] to reduce artifacts[89] such as unrealistic side-chain packing, thereby improving the quality of the final structural ensemble[94]. These structures can then be compared for convergence and reasonability, and the NOESY assignments that are retained, removed, determined, and assigned by ARIA are manually validated using, e.g., CcpNmr Analysis.

5.1.2. Solid-State NMR Spectroscopy for Protein Structural Evaluation

As shown previously in this thesis, spider silk proteins are known to contain a mix of α-helical and random coil character in the soluble state, then undergo an α→β conversion upon fibrillogenesis[95]. This conversion leads to the high mechanical strength of silk fibres[5]. The process under which this transition occurs is poorly understood and the degree of α→β conversion is silk-type dependent[18], with pyriform remaining poorly characterized. A major challenge comes from the fact that there is a lack of atomic-level structural data for both the soluble and fibre states of most classes of silk protein. Along with calculating the HPy_1 atomic-level structure, I wanted to investigate the change in the structure of the pyriform silk protein repetitive unit through solid-state NMR spectroscopy and to develop a baseline understanding of which amino acids and regions of the Py unit fold may change structure during the fibre forming process. To probe this at the atomic level, I turned to solid-state NMR. This technique is used when molecular tumbling in a sample is attenuated, such as in solid-state samples, so that transverse relaxation is very rapid and leads to significant NMR line broadening[96] and making solution-state NMR techniques unsuitable. In my solid-state NMR work, I have applied three primary experimental techniques: magic angle spinning (MAS)[97], cross-polarization (CP)[98], and dynamic nuclear polarization (DNP)[99].

First, I applied MAS and CP-based solid state NMR spectroscopy techniques. Briefly, samples are packed densely into a small rotor and spun inside the NMR magnet at the so-called “magic angle” (54.7°) at a frequency higher than the linewidth[97]. This sample spinning technique is used to minimize the influence of dipolar coupling, quadrupolar coupling, and chemical shift anisotropy upon the slow-tumbling sample. The MAS

technique thus substantially narrows the observed NMR resonances for a solid-state sample, increasing both sensitivity and resolution while allowing the isotropic chemical shifts for a given nuclear spin type to be determined. CP-based techniques provide another method of minimizing line-broadening by transferring magnetization from a highly abundant high sensitivity nucleus like ^{1}H to the nuclei of interest through dipolar coupling[100]. This minimizes the effect of long longitudinal relaxation observed in other nuclei, allowing for much more rapid spectral acquisition. For cross-polarization to occur the two nuclei that are dipolar coupled need to satisfy Hartmann-Hahn matching conditions[100], meaning the product of the applied magnetic field and gyromagnetic ratio of the first nucleus (i.e. ^{1}H) is equal to the product of the applied magnetic field and gyromagnetic ratio for the second nucleus (e.g., ^{15}N or ^{13}C).

Once spinning, 1D ^{13}C and ^{15}N CP-MAS experiments are conducted, where the polarization transfer is from ^{1}H, and the resulting spectra are then used to gain a variety of information about the sample[101], including but not limited to which resonances are observed relative to what would be expected for a given sample and overall signal to noise. Experiments such as 2D cross-polarization dipolar assisted rotational resonance (CP-DARR)[102] experiments can be conducted to obtain further information. Like the solution state NOESY experiment which relies on dipolar cross-relaxation[85], CP-DARR experiments can show intra- and inter-residue contacts depending on the mixing time used, although here the mechanism of transfer differs and relies instead on a ^{13}C-^{1}H recoupling interaction[102]. This allows for a robust method of examining near-neighbour contacts to assign amino acid spin systems and ^{13}C-^{13}C contacts that are indicative of secondary structures being formed by isotopically labelled residues.

Another technique that I have applied and that is being increasingly applied in solid-state NMR spectroscopy is dynamic nuclear polarization (DNP). This technique relies on the fact that the spin-polarization induced by a magnetic field is orders of magnitude greater for unpaired electrons than nuclei at a given set of NMR conditions[103]. By immersing the sample in a liquid containing an appropriate polarizing agent (e.g., TEMPO[104], AMUPol[105] and AsymPolPOK[106]; Figure 5.1), cooling to low (well below 0 °C) temperatures, and saturating the electron paramagnetic resonance at a frequency specific to the polarizing agent[107], polarization can be efficiently transferred from the unpaired electrons of the polarizing agent to neighbouring nuclei, providing a greatly enhanced signal-to-noise ratio. This enhanced polarization can then be transferred to other NMR active nuclei through cross-polarization, giving on the order of a 10-100 fold enhancement in sensitivity for each atom probed.

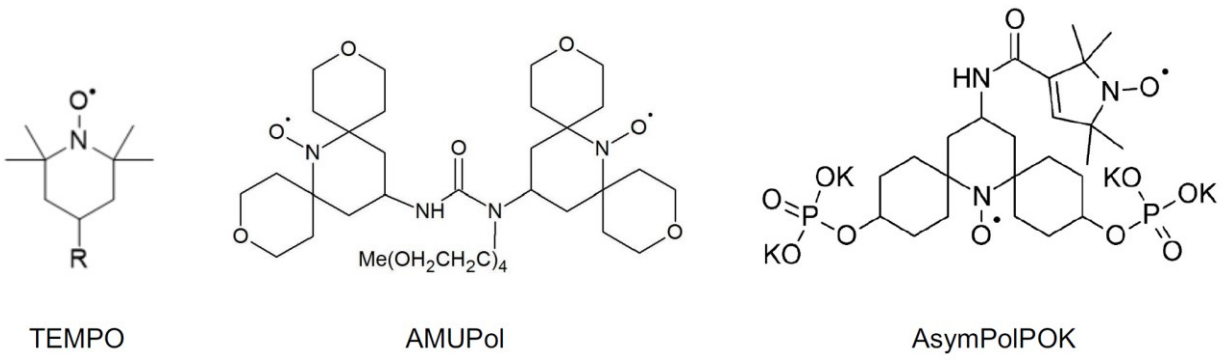


Figure 5.1. Structures of TEMPO, AMUPol, and AymPolPOK.

5.1.3 Non-Linear Optical Microscopy

Non-linear optical imaging is a powerful tool for microscopic analysis that uses non-linear interactions between light and materials to generate images. These techniques are

generally non-destructive and provide label-free approaches to study various systems including cell membranes[108] and collagen fibres[109]. While autofluorescence of silk fibres can be studied through two-photon excited fluorescence[110], the work I present in this thesis has focused on second and third harmonic generation (SHG and THG). Both SHG and THG also have precedents in spider silk analysis[111-113]. These harmonic generation techniques refer to a light-material interaction where two (SHG) or three (THG) photons of the same frequency interact and are combined. The resulting photon will have X times the frequency and 1/X times the wavelength, where X equals 2 or 3 for SHG or THG, respectively. The THG phenomenon is induced at interfaces within a sample between regions that have different refractive indices[114], such as an air-material interface. The SHG phenomenon, alternatively, gives a highly sample-specific signal. To obtain an SHG signal, a sample must be non-centrosymmetric such as containing slightly offset β-sheets[111]. Thus, silks provide an ideal target for SHG analysis.

5.2. Materials and Methods

5.2.1. Δ9-70+Δ206-242-HPy$_1$ Structure Calculation and Validation

NOE-derived distance restraints for Δ9-70+Δ206-242-HPy$_1$ were assigned in CcpNMR Analysis 2.4.2[77] for a ^{15}N-edited NOESY-HSQC spectrum and a ^{13}C-edited NOESY-HSQC spectrum. Ambiguous NOEs were iteratively refined using ARIA 2.3[115], with maximal distance restraints of 6 Å. Following four rounds of NOE restraint assignment and iterative refinement through ARIA, coupled with manual assessment of each set of restraint assignments, dihedral angle restraints were included as determined by DANGLE[74] through its implementation in CcpNmr Analysis 2.4.2[77]. A summary of the

iteration conditions is shown in Table 5.1. The NOE peak assignments determined by ARIA were manually assessed between each round of calculations to either accept or reject each suggested assignment based both on evaluation of the agreement of the peak with my assigned chemical shifts and on the calculated interatomic distance in a given water-refined structural ensemble. A schematic of the iterative ARIA process is illustrated in Figure 5.2.

Table 5.1. Summary of ARIA parameters for structural ensemble calculations.

	Iteration Number		
	1-4	5-10	11-13
Dihedral Angle Restraints Used	No	Yes	Yes
Number of Ensembles	8	8	8
Structures Per Ensembles 1-7	20	20	40
Structures for Ensemble 8	20	20	100
Structures Used for Refinement of Successive Ensembles	7	7	15
Number of Water Refined Structures from Ensemble 8	10	10	20

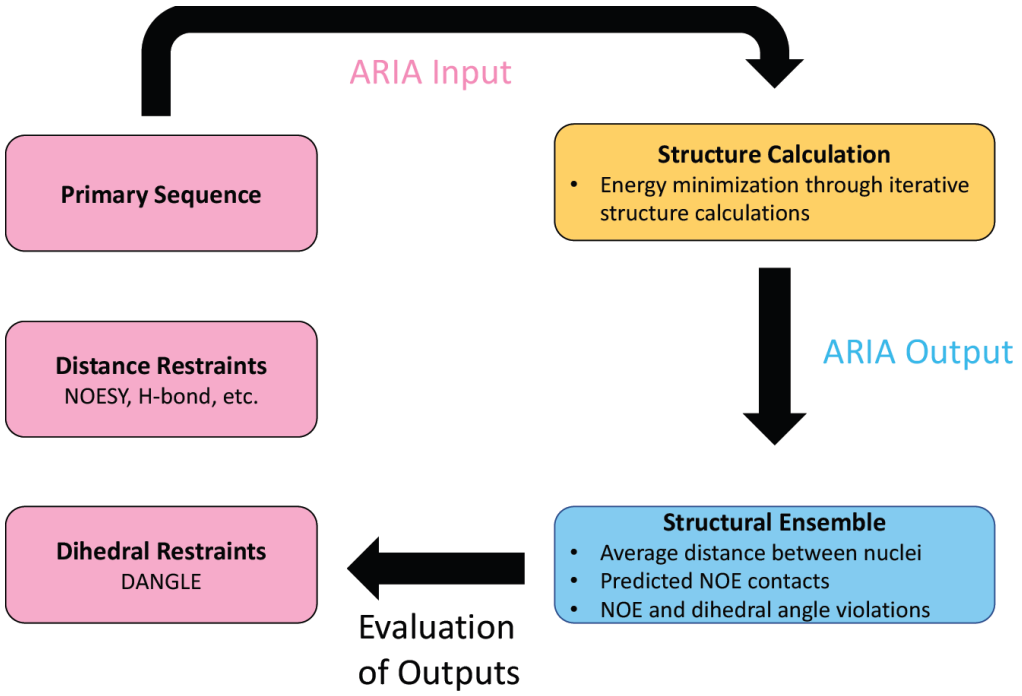


Figure 5.2. Schematic of iterative ARIA calculation strategy.

Between iterations, a number of quality control checks were performed, starting with evaluation of NOE violations based on distance restraints with the help of in-house scripts to extract the number of violations from the analysis output. From the current structural ensemble (iteration 13), the 20 water-refined structures were concatenated into a single PDB file and input into the online PSVS[116] server (https://montelionelab.chem.rpi.edu/PSVS/PSVS/) for structure evaluation with secondary structures calculated using DSSP[117]. Further, DANGLE violations were sorted from the ARIA output data using an in-house tcl script and assessed to confirm agreement of phi and psi angles in the water refined structural ensemble.

5.2.2. AlphaFold and CS-ROSETTA Structural Predictions

Two independent methods of predicting tertiary fold were compared directly to the structural ensemble resulting from iterative NMR restraint refinement: CS-ROSETTA[118] and AlphaFold[119]. In the former case, with CS-ROSETTA, structures were predicted by exporting the backbone chemical shift assignments for the full-length HPy_1 and manually removing assignments from varying segments of the N- and C-terminal structured regions. This allowed for a comparative determination of the region of the HPy_1 that appeared to be most suitable for chemical shift-based structural prediction by CS-ROSETTA. Data files were converted to TALOS format for data entry into the web application (https://csrosetta.bmrb.wisc.edu/csrosetta/submit) focusing on backbone chemical shift assignments from residues 71S to 196P of the full-length HPy_1, with 5000 structures generated while specifying not to modify the structure via flexible tail handling. AlphaFold structural predictions were performed via the web application (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb). Both the HPy_1 and Δ9-70+Δ206-242-HPy_1 amino acid sequences were independently provided as structural prediction input to the AlphaFold2.ipynb web application and run with msa mode=Mmseqs2 (UniRef+Environmental), pair mode=unpaired+paired, and 3 recycles per predicted structure.

5.2.3. Tracking Chemical Shift Perturbation as a Function of Temperature

Preliminary temperature-based perturbation experiments were performed on Δ9-70+Δ206-242-HPy_1 to evaluate the potential of region-specific differences in stability and

secondary structure units[120, 121]. Data on temperature stability were collected and processed by me, with data analysis performed under my mentorship by Skylah McLeod van Wagoner. Sensitivity enhanced ^{1}H-^{15}N HSQC experiments were acquired at 40.0 °C, 45.0 °C, and 50.0 °C using an 11.7 T Avance NMR spectrometer equipped with a 5 mm BBFO SmartProbe. Samples were prepared in susceptibility-matched Shigemi tubes and allowed to equilibrate at the desired experimental temperature for 5 min prior to data collection. ^{1}H-^{15}N HSQC experiments used to track temperature-based chemical shift perturbations were collected with the following parameters: initial Δ9-70+Δ206-242-HPy$_1$ concentration of ~175 μM; D1 = 1 s; NS = 4 (40 °C, 45 °C), 8 (50 °C); sweep width (ppm) =14.0019 (^{1}H), 30 (^{15}N), number of data points = 2048 (^{1}H), 256 (^{15}N).

Data were processed via NMRpipe[75] (using the NMRbox resource[76]) and then imported into CcpNmr Analysis[77] for analysis relative to the assignments determined at 40 °C. Combined chemical shift changes for backbone residues were tracked via the Shift Differences function in CcpNmr Analysis and exported for calculation of the combined chemical shift displacement[122] for each residue at position *j* in the spectra of interest ($\Delta\delta_{comb,j}$) as follows:

$$\Delta\delta_{comb,j} = \sqrt{\frac{1}{N_a}\sum_{i=1}^{N_a}(w_i\Delta\delta_{ji})^2} \tag{5.1}$$

where N_a accounts for the types of atoms involved and is generally ignored when the same for all residues, w_i is a weighting factor for nuclei *i* accounting for the gyromagnetic ratio of each nucleus of interest (1.000 for ^{1}H, 0.102 for ^{15}N). Simplifying this for ^{1}H-^{15}N correlations, equation 5.1 reduces to 0.102 $\Delta\delta_{ji}$.

5.2.4. Investigating the C-Terminal PX-rich motif

To investigate the role of the C-terminal proline-rich region, a truncated version of HPy_1 was cloned via PCR to remove residues 196-230 from HPy_1 (protein sequence shown in Figure 5.3). The truncated protein was expressed, purified, and prepared for NMR analysis identically to all other HPy_1 proteins as described in Section 4.2. ^{1}H-^{15}N HSQC experiments were conducted at 40.0 °C using a 16.4 T Avance III spectrometer equipped with a 5 mm indirect detection TCI cryoprobe. To further understand what was observed in these experiments, AlphaFold predictions were carried out in an identical manner to that described in section 5.2.2.

```
  1  MGHHHHHHSS AQENSFTQSS VAQQSAVAQQ SSVSQQSSAA QQSSVAQSQQ TSYSAATNAG  60
 61  SSVSQSQAIV SSAPVYFNSQ TLTNNLASSL QSLNALNYVS NGQLSSSDVA STVARAVAQS  120
121  LGLSQGSVQN IMSQQLSSIG SGASTSSLSQ AIANAVSSAV QGSQAAAPGQ EQSIAQRVNS  180
181  AISSAFAQLI SQRTAVASQF QASASSQS                                     208
```

Figure 5.3. Protein sequence for Δ196-230 HPy_1

5.2.5. Selective $^{15}N/^{13}C$ Unlabelling for Solid-State NMR

Selectively unlabelled HPy_2 was expressed and in the same manner as described in Section 3.2 and the same C1 cocktail as used in Chapter 3, with the specific amino acid supplementation scheme used as follows. Initial concentrations of amino acids were determined based on experiments on aciniform silk (L. Xu, M-L. Tremblay, X-Q. Liu, and J.K. Rainey, unpublished work). In this work, recombinant aciniform proteins were expressed in ^{15}N-containing M9 medium supplemented with unlabelled amino acids. Unlabelling efficiency was tracked on the basis of assigned chemical shifts[71] using ^{1}H-^{15}N

HSQC experiments to evaluate the degree of residual labelling observed for residues of types that corresponded to the supplemented amino acids.

Considering the most prevalent amino acids in HPy_1, cross-metabolism[63], and unique chemical shifts, an initial amino acid supplementation condition was developed for HPy_1 (Table 5.2, condition B1). The concentration in each set of amino acids was optimized for HPy_1 by testing unlabelling efficiency via ^{1}H-^{15}N HSQC spectra. In choosing which amino acids to add to the media, I aimed for a set of amino acids that would give a wide distribution throughout the protein (Figure 5.4), have distinctive chemical shifts in solid-state NMR, avoid deleterious cross-metabolism, and can be tracked with some location specificity. Tested conditions are shown in Table 5.1.

(a)

```
  1  MGHHHHHHSS  AQENSFTQSS  VAQQSAVAQQ  SSVSQQSSAA  QQSSVAQSQQ  TSYSAATNAG  60
 61  SSVSQSQAIV  SSAPVYFNSQ  TLTNNLASSL  QSLNALNYVS  NGQLSSSDVA  STVARAVAQS  120
121  LGLSQGSVQN  IMSQQLSSIG  SGASTSSLSQ  AIANAVSSAV  QGSQAAAPGQ  EQSIAQRVNS  180
181  AISSAFAQLI  SQRTAPAPAP  RPRPAPLPAP  APRPRPAPAP  RPAPVYAPAP  VASQFQASAS  240
241  SQS                                                                    243
```

(b)

```
  1  MGHHHHHHSS  AQENSFTQSS  VAQQSAVAQQ  SSVSQQSSAA  QQSSVAQSQQ  TSYSAATNAG  60
 61  SSVSQSQAIV  SSAPVYFNSQ  TLTNNLASSL  QSLNALNYVS  NGQLSSSDVA  STVARAVAQS  120
121  LGLSQGSVQN  IMSQQLSSIG  SGASTSSLSQ  AIANAVSSAV  QGSQAAAPGQ  EQSIAQRVNS  180
181  AISSAFAQLI  SQRTAPAPAP  RPRPAPLPAP  APRPRPAPAP  RPAPVYAPAP  VASQFQASAS  240
241  SQS                                                                    243
```

Figure 5.4. Targeted distribution of isotopically labelled (red) residues for the first (a) and final supplementation scheme (b) denoted as B1 and B2 respectively.

Table 5.2. Final concentrations of amino acids supplemented for protein overexpression in M9 media for solid-state NMR.

AA	Cocktail		
	B1	B2	C1
	[AA] (mM)	[AA] (mM)	[AA] (mM)
Ala	10.3	10.3	20.7
Arg	0	0	0
Asn	0	0	0
Asp	0	0	0
Cys	0	0	0
Glu	3.4	0	0
Gln	13.7	13.7	27.4
Gly	2.4	3.2	6.0
His	6.4	0	0
Ile	10.7	10.7	21.3
Leu	5.3	5.3	10.7
Lys	0	0	0
Met	3.4	0	0
Phe	0	0	0
Pro	17.4	17.4	34.7
Ser	25.5	30.5	60.9
Thr	0	0	0
Trp	0	0	0
Tyr	0	0	0
Val	6.8	6.8	13.7

5.2.6. Fibre Spinning and Evaluation of Selectively Unlabelled HPy_2

Selectively unlabelled HPy_2 protein samples were wet-spun using a 3:1:1 TFA:TFE:H_2O dope solvent, as described in Section 2.2. To obtain a fibre with more pronounced

anisotropy than anticipated for an AS silk fibre, 2× post-spin stretching in air was targeted as a highly consistent spinning condition on the basis of my prior work (Chapter 2). Fibres were then imaged and mechanically tested as described in Sections 2.2.7 and 2.2.9 to ensure that the spun product was self-consistent and reasonable compared to previously collected data.

5.2.7. Solid-State NMR Data Acquisition, Processing, and Analysis

Selectively unlabelled HPy_2 fibres from batches that showed tensile testing results consistent with the condition of interest (i.e., consistent with those reported in Chapter 2) were evaluated by MAS solid-state NMR spectroscopy. All conventional solid-state NMR experiments were conducted by Dr. Ivan Hung while DNP sample preparation and experiments were carried out by Dr. Frederic Mentick-Vigier, both at the National High Magnetic Field Laboratory (Florida State University, Tallahassee, FL). Fibres were packed into a Revolution NMR 3.2 mm pencil-style rotor and spun at the magic angle (13 kHz spinning frequency, 20 °C) using an 18.8 T Bruker Avance III spectrometer (Bruker Biospin) equipped with a home-built triple-resonance low-E 3.2 mm MAS probe similar to previously reported probes[123]. To enhance the signal-to-noise ratio, the sample was repacked into a Bruker Stator 3.2 mm standard wall sapphire rotor and dynamic nuclear polarization experiments were collected in 4:4:1 $D_2O/H_2O/d_8$-glycerol using AsymPolPOK as the polarizing agent. The sample was cooled to -173 °C, and spectra were collected, using a 18.8 T Avance III spectrometer equipped with a Bruker 3.2 mm LTMAS probe and application of MAS at a frequency of 10.5 kHz. The high-frequency peak of adamantane (at 40.48 ppm) was used as an external reference for all spectra. ^{13}C CP-

MAS spectra peaks were annotated to determine if all targeted residue types were accounted for and if any extra residues were observed. DARR[102] experiments were peak picked in CcpNMR analysis[77], and peaks were assigned with the assistance of the PLUQin algorthim[124], which calculates the probability of a given correlation and secondary structure based on standard values for different residues in each type of secondary structure (Figure 5.5). Spectral acquisition parameters are described in Table S5.1

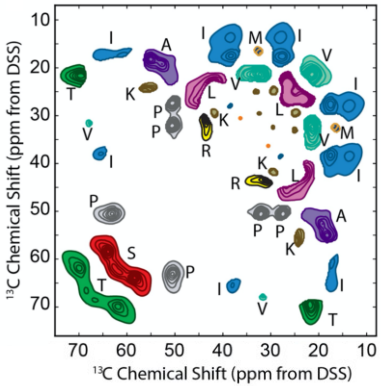


Figure 5.5. ^{13}C-^{13}C chemical shift regions where correlations for one type of residue are >10 times more likely to occur than all other residues. Reprinted by permission from Springer Nature, Journal of Biomolecular NMR, Conformationally selective multidimensional chemical shift ranges in proteins from a PACSY database purged using intrinsic quality criteria, Fritzching K. et al., 2016[124] .

5.2.8. Microstructural Characterization of the Assembled Fibre

Recognizing that the assembled fibres are a macroscopic assembly of proteins and, therefore a more complex system than atomic-level information can fully elucidate, I performed preliminary experiments to further probe the arrangement of this assembly. SHG and THG measurements were collected using a home-built apparatus as previously

reported in detail[125]. SHG and THG experiments were conducted by Dr. Richard Cisek and me at Saint Mary's University, with Dr. Cisek operating the instrument. Briefly, SHG data were collected using a femtosecond pulsed laser (FemtoLux3, Ekspla) at 1030 nm wavelength, 5 MHz repetition rate, ~250 fs per pulse, which was coupled to a custom-built nonlinear optical microscope. Images were created by raster scanning the laser beam across the sample using a pair of galvanometric scan mirrors (ScannerMAX, Pangolin Laser Systems Inc.). The laser was focused using a 0.9 numerical aperture air immersion objective lens (Plan-Apochromat 20×, Carl Zeiss AG) and SHG signal was collected using a custom built 0.85 numerical aperture collection lens (Omex Technologies) and an interference filter to separate the SHG signal from the laser light (87-789, Edmund Optics Inc.). Polarization-in, polarization out (PIPO) spectra were collected using a linear polarizer (after the collection objective lens) followed by a rotating half-wave plate, the latter placed before the excitation objective lens to allow for rotation of the incident laser polarization. Each measurement consisted of a combination of 8 emission polarization angles for each of 8 half-wave plate angles and one scan of the initial condition (65 scans total). These images were then summed in ImageJ as a Z-projection. THG measurements were collected on the same equipment and a application of a different filter to remove the laser light (65–129, Edmund Optics Inc.).

5.3. Results

5.3.1. First Atomic-Level Structural Ensemble of a Soluble Recombinant Pyriform Protein

Having determined in Chapter 4 that Δ9-70+Δ206-242-HPy_1 is the structured core of the HPy_1 unit, chemical shift assignments and a set of both assigned and unassigned ^{1}H-^{15}N HSQC NOESY and ^{1}H-^{13}C HSQC NOESY peaks were applied in ARIA for iterative folding calculations as described in Section 5.1.1. As would be expected for early rounds of structural restraint refinement, the initial iterations carried out prior to inclusion of dihedral angle restrictions produced a mostly disordered, non-converged structure (Figure 5.6; iterations 1 and 4). By introducing dihedral angle restraints as a folding restriction, a more ordered structural ensemble was determined, however the generated structures led to bending of the 23-residue helix-6 and eventual packing on the top of the bundle (Figure 5.7). Bending of helix-6 as calculated appeared to be an unlikely architecture, probably resulting from ARIA having spuriously assigned some NOE restraints that were causing this helical bend, so all NOE peaks that were assigned to residues within the bent region were manually reassigned to become ambiguous where reasonable to determine if the ARIA had indeed forced convergence to this structure with one or more incorrect NOE assignments. This straightened out the helix and aligned it beside the helical core (Figure 5.7), with connections that were fully supported by the data with respect to my chemical shift assignments, and led to greatly improved convergence of the structural ensemble. Select iterations to reach the final structure are diagramed in Figures 5.6, with tabulated statistics for the structural assignment and last ensemble of 20 structures shown in Figure 5.8, Figure 5.9, Table 5.3 and Table A5.

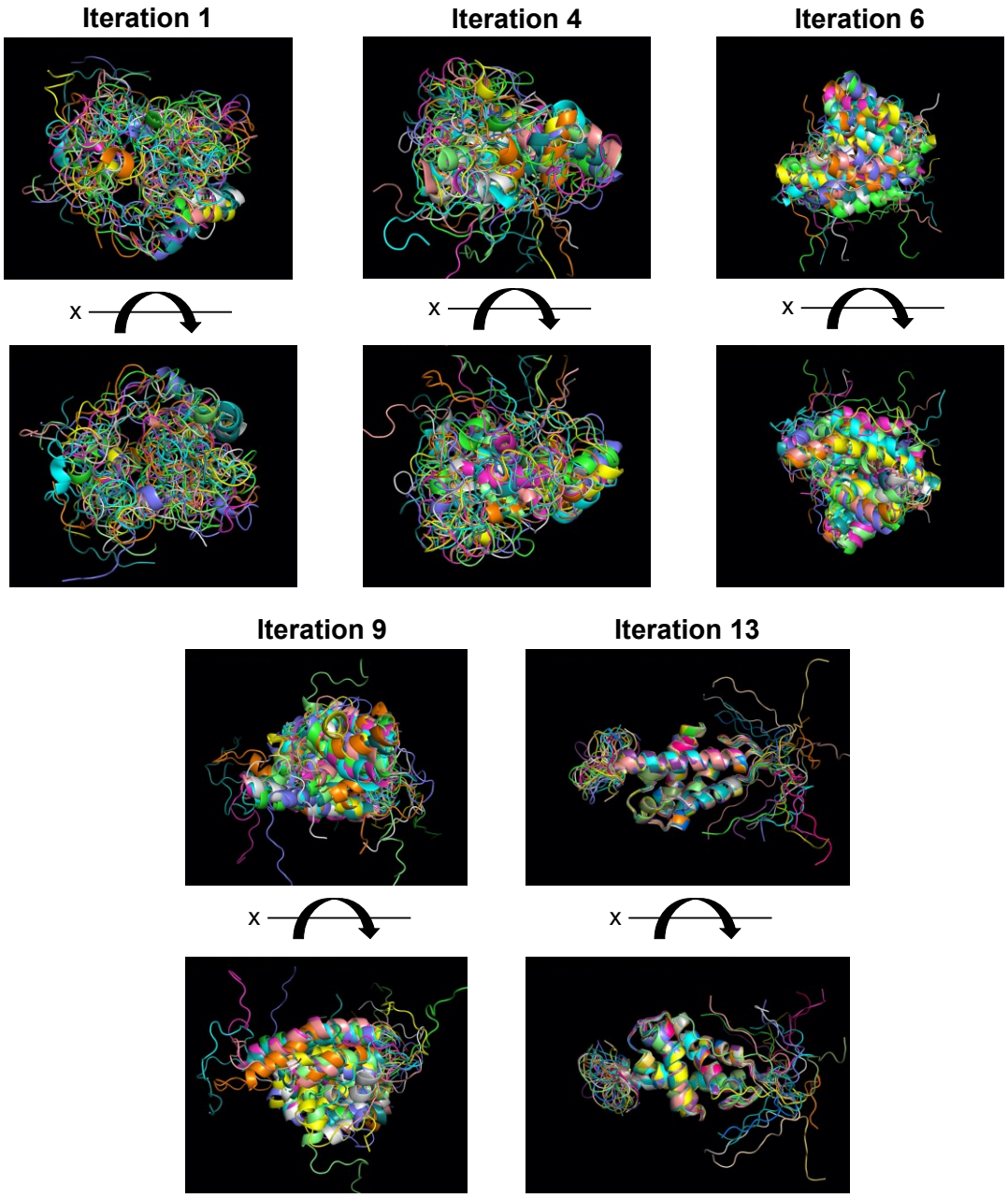


Figure 5.6. Progression of ARIA calculated Δ9-70+Δ206-242-HPy$_1$ structures with selected iterations and 180° x-axis rotation. Shown are the first iteration, the last iteration before DANGLE restraint addition (iteration 4), after two rounds with DANGLE restraints (iteration 6), an intermediate iteration (iteration 9), and the current structure (iteration 13). All structures are water refined, with 10 structures shown for iterations 1, 4, 6, and 9. For iteration 13, 20 structures are shown.

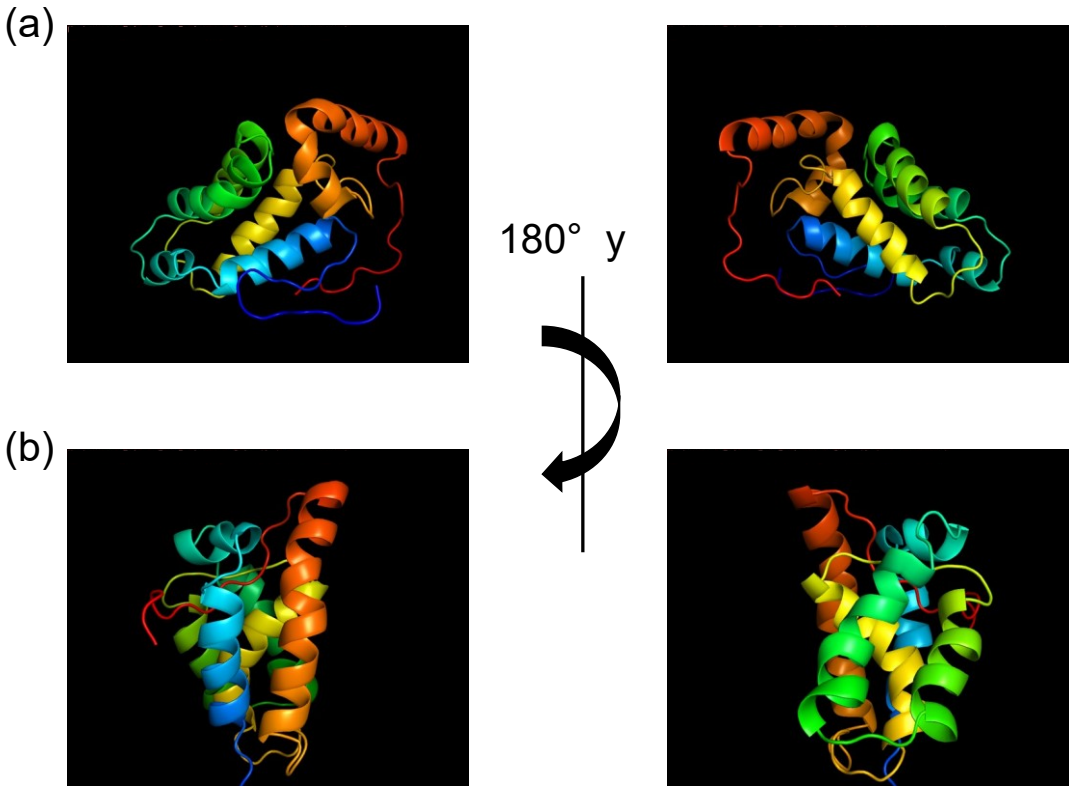


Figure 5.7. Bent helix-6 (a) vs well-packed helix-6 (b).

Table 5.3. Tabulated statistics for the current ARIA ensemble of 20 structures.

Distance restraints	
Intra-residue	1624
Sequential ($\|i - j\| = 1$)	1112
Medium-range ($1 < \|i - j\| < 5$)	938
Long–range ($\|i - j\| \geq 5$)	532
Total	4206

Average Dihedral violations > 5 deg	2.7 ± 1.76	
Average distance restraint violations per structure	violations > 0.5 Å	0.65 ± 0.79
Average distance restraint violations per structure	violations > 0.3 Å	4.25 ± 1.47

Energies (kcal/mol)	
Total	-4847.58 ± 130.605
Van der Waals	-1191.68 ± 11.2048

PSVS Secondary Structure Elements:

alpha helices: 17A-28A, 35A-38A, 44A-59A, 63A-77A, 83A-98A, 109A-131A

beta strands: None

RMSD	*All residues*	*Ordered residues*[126]	*Residues 17-133*[127]
All backbone atoms	4.6 Å	1.5 Å	0.5 Å
All heavy atoms	5.0 Å	1.6 Å	0.8 Å

Number of close contacts (within 1.6 Å for H ats, 2.2 Å for heavy ats): 15

RMS deviation for bond angles: 0.8°

RMS deviation for bond lengths: 0.006 Å

Ramachandran Plot Summary for residues 17-133[127] from Procheck

Most favoured regions	*Additionally allowed regions*	*Generously allowed regions*	*Disallowed regions*
90.60%	9.30%	0.00%	0.10%

Global quality scores

Program	*Verify3D*	*ProsaII (-ve)*	*Procheck (phi-psi)*[127]	*Procheck (all)*[127]	*MolProbity Clashscore*
Raw score	0.13	0.81	-0.04	-0.21	49.92
Z-score[1]	-5.3	0.66	0.16	-1.24	-7.04

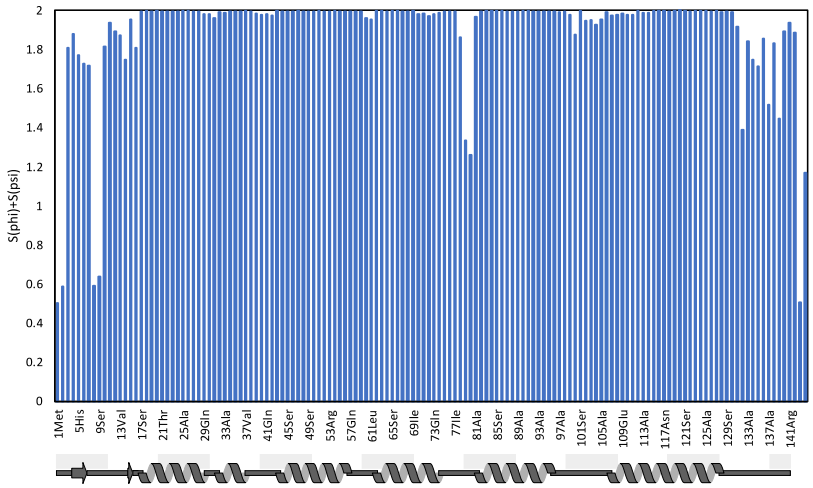


Figure 5.8. Sum of S(phi) and S(psi) for each residue in the structure ensemble, with secondary structure shown at each region. The closer the values are to 2 indicated more convergence of the phi and psi angles at a given residue.

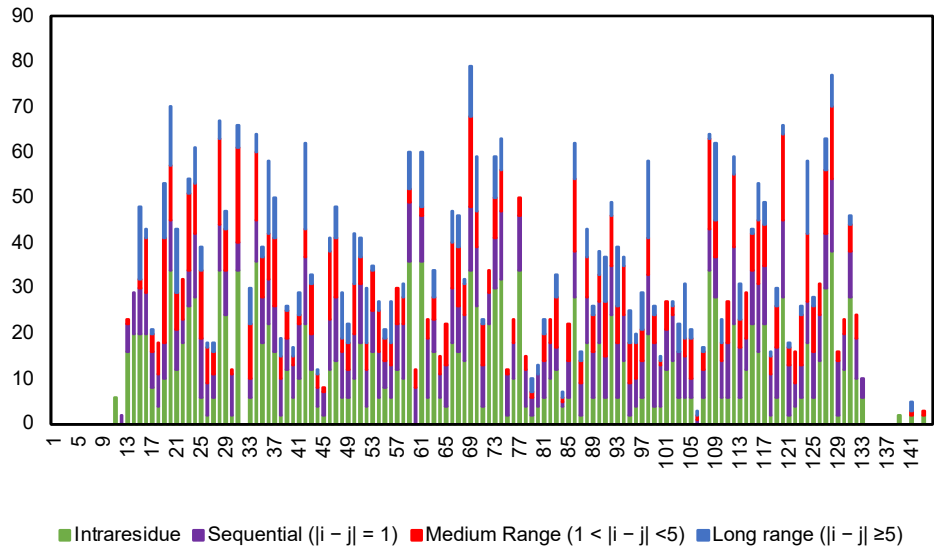


Figure 5.9. Summation of NOE contacts per residue.

From statistics of the generated and water refined structures, it is shown that reasonable numbers of unique long- and medium-range NOEs were used in the iterative structure

calculation (Tables 5.3 and A5). Importantly, these restraints are broadly dispersed throughout the protein (Figure 5.9) and i_{+3}/i_{+4} contacts are seen frequently throughout regions predicted to be α-helical (Figure 5.10), giving further confidence in the structural ensemble generated. The RMSD over the structural ensemble for the well-ordered residues (17-133; Figure 5.4) was 0.5 Å for backbone atoms and 0.8 Å for all heavy atoms, quantitatively indicating that structure is well converged (Table 5.2). Ramachandran plot statistics calculated by PROCHECK[128] only indicate a small percentage of disallowed regions which are likely linked to poor convergence in some more flexible regions (Figure 5.6). Other issues noted such as restraint and dihedral violations, and the MolProbity Clashscore[129] (overlap of non-donor–acceptor atoms by more than 0.4 Å per 1000 atoms), will be minimized with further refinement of the structural ensemble.

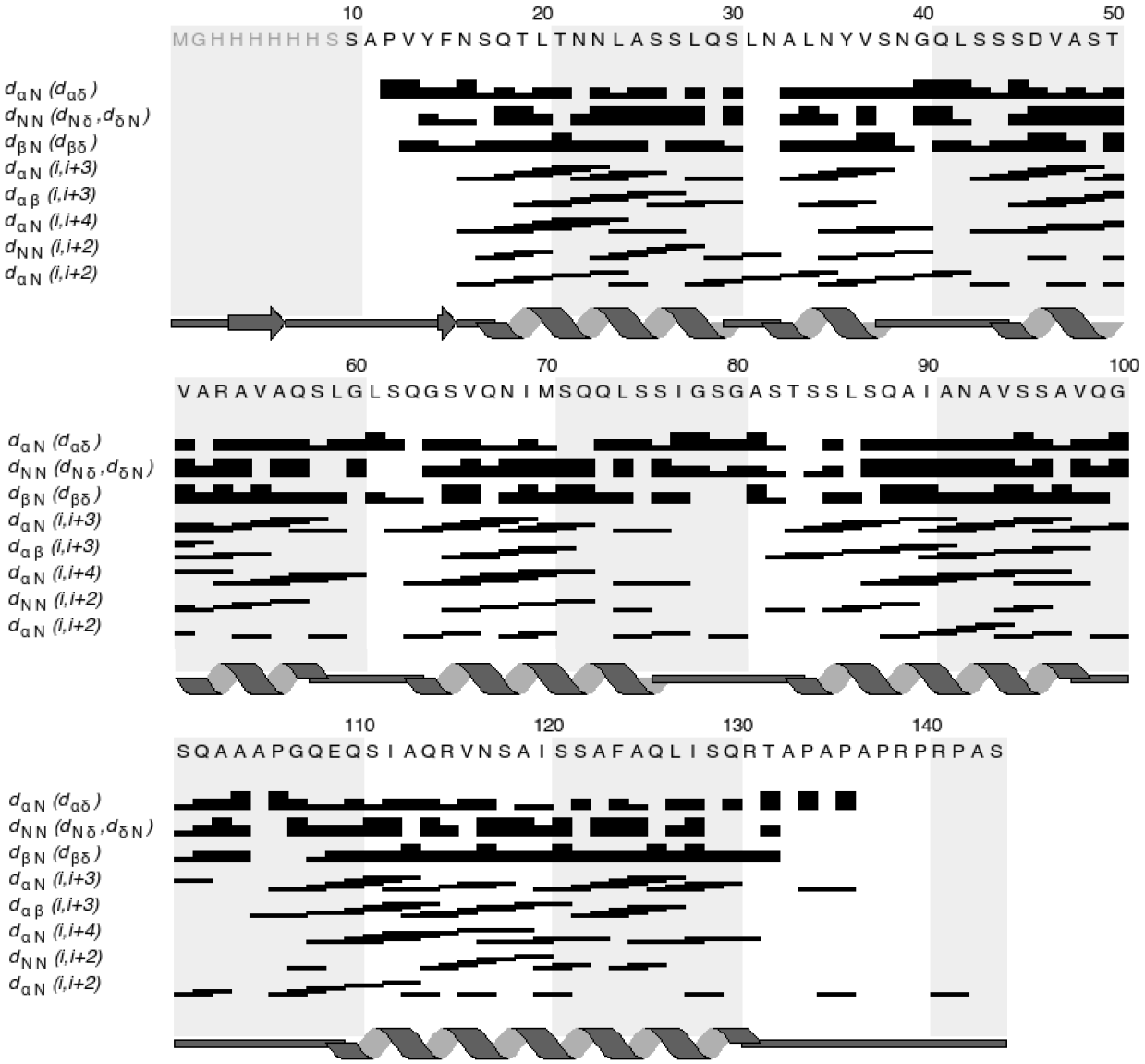


Figure 5.10. Sequential graphical plot of NOE contacts in Δ9-70+Δ206-242-HPy$_1$.

While the structural convergence has been outstanding over the last few iterations of ARIA (e.g., Figure 5.11), I am currently working toward minimizing dihedral violations (Table 5.2), which appear to be the result of restraints that DANGLE has predicted for highly dynamic regions at the N- and C-termini of the protein (Table A6). NOE violation statistics are excellent (Table 5.2), but ARIA iterations to this point have employed the

“do not trust assignments” setting which allows for a more rigorous/unbiased approach to distance constraint assignment. With this setting, however, ARIA has converged to a point where it only uses ~95% of the NOE contacts, with a similar proportion of NOEs being removed across all contact ranges (i.e., intra-residue, sequential, medium-range and long-range). For a fully robust structure, these discarded restraints are currently being manually assessed to determine why they were omitted in structure calculations and subsequent round(s) of final structure refinement will employ all assigned NOE restraints through application of ARIA’s “trust assignments” setting. Through my analysis of the NOE violations that are currently being observed and examination of NOEs that have been removed by ARIA, this appears to be likely predominantly due to CcpNmr Analysis having placed overly tight upper-bound distance constraints on some NOE contacts. Lastly, PyRPF analysis will be used to determine how well the generated structures fit with the assigned data[130].

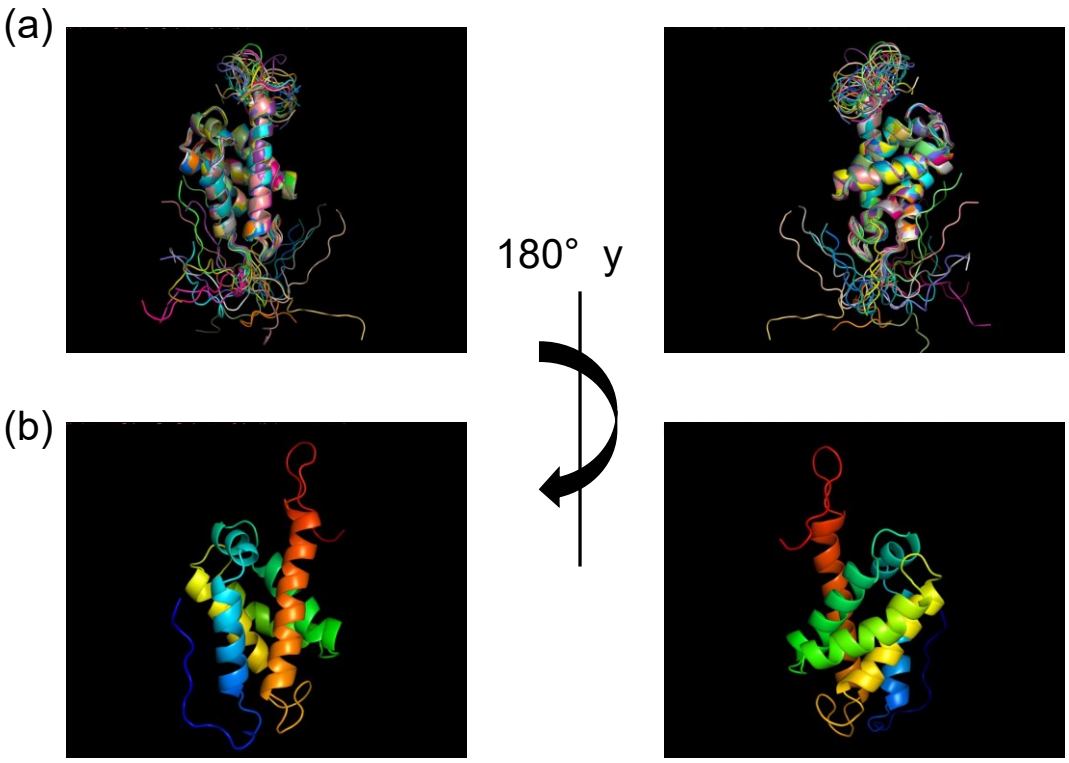


Figure 5.11. Current ensemble of 20 lowest energy, water refined structures for Δ9-70+Δ206-242-HPy$_1$ (a) and colour coded lowest energy structure going from blue at the N-terminus to red at the C-terminus (b).

From the ensemble of structures and lowest energy structure for Δ9-70+Δ206-242-HPy$_1$ (Figure 5.11), a 6-helix bundle was clearly observed as indicated by secondary structure predictions generated in CcpNmr Analysis with delineations of α-helical regions relative to random coil agreeing well with these predictions and predictions from PSVS. To test whether structure folding predication approaches support the NMR-derived structural ensemble, AlphaFold was used to predict protein structures based upon the protein sequence for HPy$_1$ and Δ9-70+Δ206-242-HPy$_1$, and both the protein sequence and backbone chemical shift assignments for HPy$_1$ focusing on the structured core (Δ9-

70+Δ197-242-HPy$_1$) were used for CS-ROSETTA structural calculations, noting that this was carried out at an earlier stage in refinement. For AlphaFold predictions, 3/5 structures for HPy$_1$ and 4/5 structures for Δ9-70+Δ206-242-HPy$_1$ were well-overlayed with the lowest energy ARIA structure for HPy Δ9-70+Δ206-242-HPy$_1$ [Figure 5.12(a)]. However, when comparing the lowest 10 energy CS-ROSETTA structures relative to the lowest energy ARIA structure, 6/10 structures showed the same pattern of helical bundles as ARIA and helices were similarly oriented but slightly offset from each other [Figure 5.12(b)]. Details of non-overlaying ARIA and CS-ROSETTA structures are described below. It should be noted that CS-ROSETTA failed to converge for Δ1-70+Δ197-243-HPy$_1$ (Figure A4), as evidenced by CA RMSD's of 4.849 ± 0.742 and 2.504 ± 0.357 for the whole structure and select residues (32-120, 122-123), respectively.

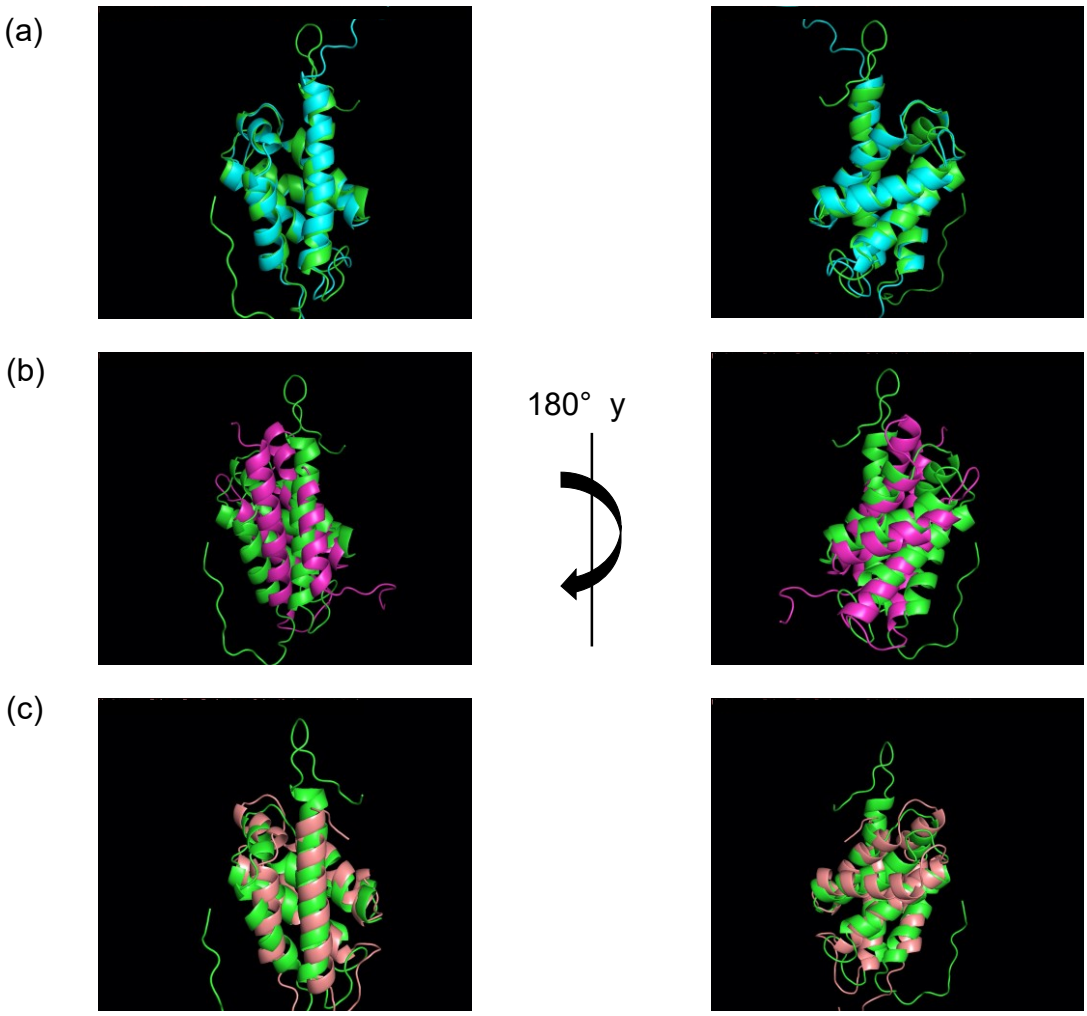


Figure 5.12. Structure of the current lowest energy ARIA ensemble structure (green) compared to the lowest energy AlphaFold (a) and CS-Rosetta (b) structures, and the third-lowest energy CS-Rosetta structure (c).

All AlphaFold predictions that didn't overlay with my NMR-derived ARIA-based structure and 3/4 non-converged, lowest energy CS-ROSETTA structures differed by the swapping of N-terminal helices 1 and 2 with the C-terminal helix 6 (Figure 5.13). While both placements of the terminal helices give viable-looking structures, numerous through-space NOE contacts assigned during the course of structural refinement allow

unambiguous defining of the topology. This highlights the need for rigorous structural determination for a previously uncharacterized class of protein such as pyriform silk, rather than reliance upon structural prediction algorithms. Furthermore, my NMR-derived structure serves to explain the anomalous chemical shifts observed for 109E, which is shown to be near 15F [Figure 5.13(a)] and supported by observed NOE contacts. In the alternative placement of the terminal helices predicted by AlphaFold and CS-ROSETTA, 109E is next to 99Q [Figure 5.13(b)]. Comparing the two placements, the observed 109E chemical shifts indicating a highly perturbed local environment that is better explained by the proximity to the phenyl ring of 15F than the side chain of 99Q. electron-withdrawing than electron-donating functionality. Therefore, the NMR-derived structural ensemble is supported by observed chemical shifts and contacts in the NMR spectra, independent predictions from AlphaFold and CS-ROSETTA, and allows the ruling out of an alternatively predicted structure in AlphaFold and CS-ROSETTA.

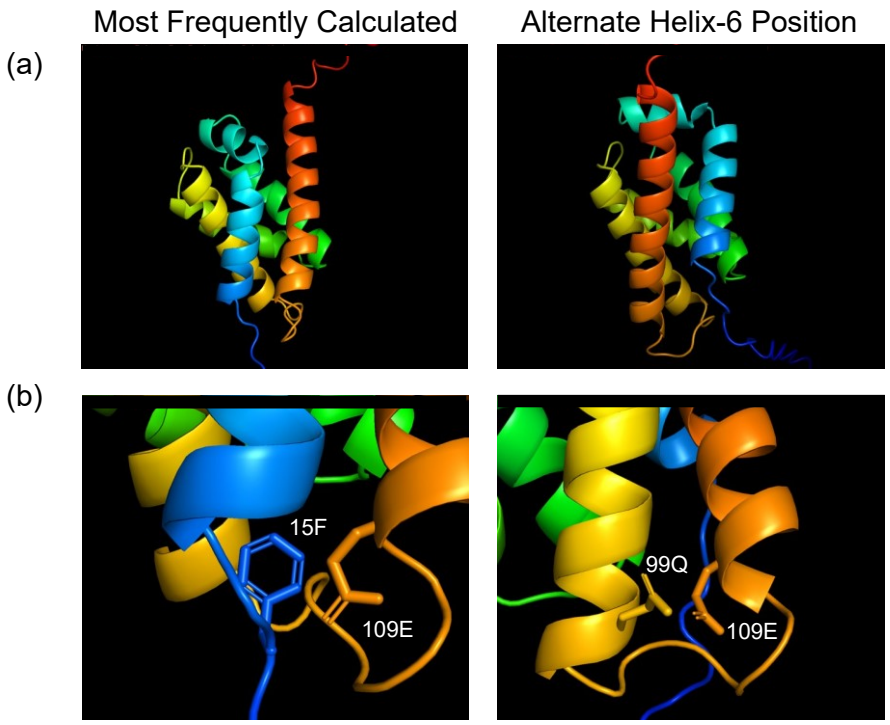


Figure 5.13. AlphaFold predicted alternate helix-6 position. Shown are representative AlphaFold predicted structures for Δ9-70+Δ206-242-HPy$_1$ showing the two orientations observed for helices 1, 2, and 6 through AlphaFold and CS-ROSETTA calculations (a). The nearest side chains capable of electron density interactions with 109E are shown for each predicted conformation (b). The structures are colour coded going from blue at the N-terminus to red at the C-terminus.

5.3.2. Investigating the C-Terminal PX-rich motif

Having investigated the core structured region of pyriform silk and finding no clear information regarding the role of the conserved C-terminal PX-rich motif in Chapter 4, I removed the PX-rich motif from HPy$_1$ (Δ196-230 HPy$_1$) to determine the effect of this truncation on Py unit structure. When this protein was examined by NMR, a noticeable change in the ^{1}H-^{15}N chemical shift profile was observed in the first examination [Figure

5.14(a)]. Precipitation was then observed to occur, which led to restoration of the original HPy_1 chemical shift pattern [Figure 5.13(b)].

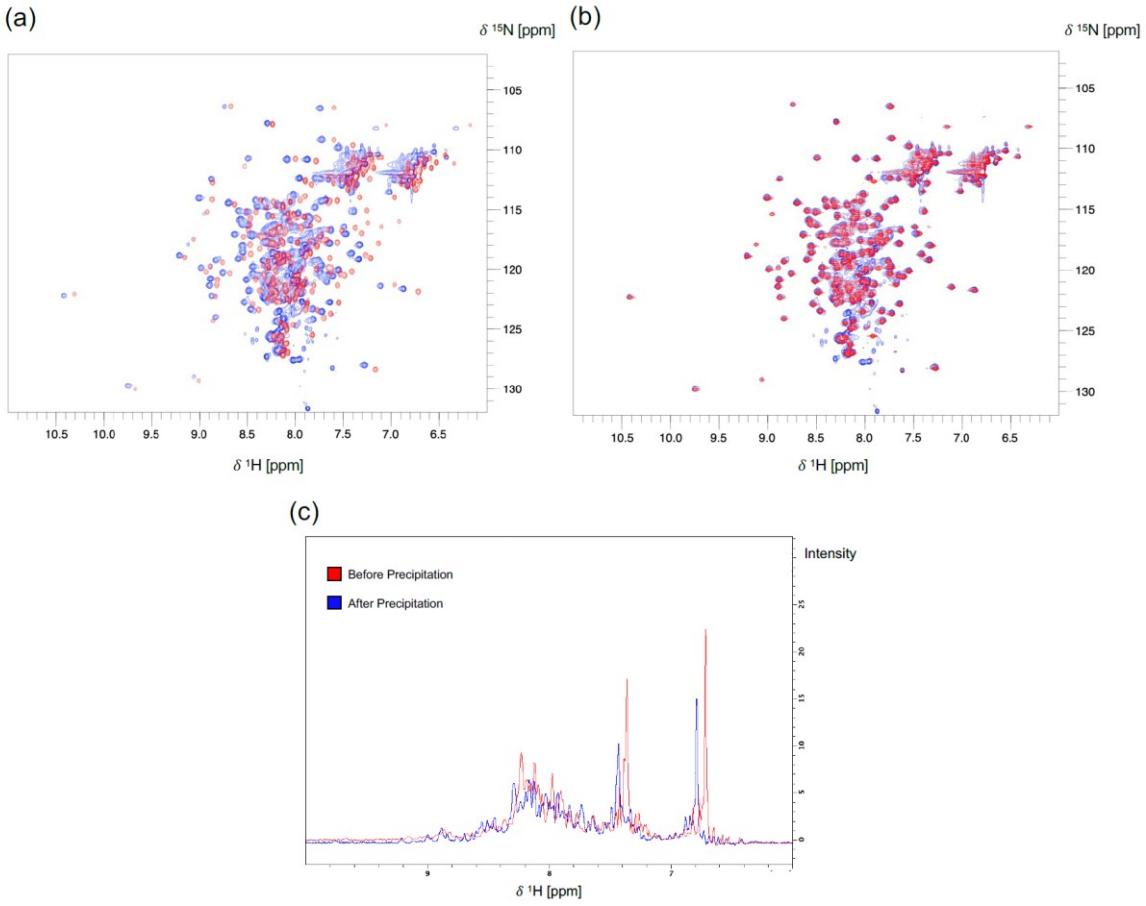


Figure 5.14. Overlay of ^{1}H-^{15}N HSQC spectra for Δ196-230 HPy_1 before (a; ~325 μM) and after (b) slight sample precipitation (red) and compared to full length HPy_1 (blue). An overlay of the first increments of ^{1}H-^{15}N HSQC experiments is shown before and after precipitation (c). The spectrum before precipitation was collected with DS=32, NS=4, ^{1}H TD=2048, ^{15}N TD=128, while the spectrum after precipitation was collected with DS=16, NS=4, ^{1}H TD=2048, ^{15}N TD=512.

To understand this changing behaviour, the amino acid sequences of Δ196-230 HPy with one and two truncated repeat units were used to generate structural predictions with

AlphaFold. Without this proline-rich region, it was predicted that the 6th C-terminal helix extended beyond the length observed in HPy_1 proteins, with the protein packing being slightly changed as a result [Figure 5.15(a)]. Structures predicted with a second Δ196-230 Py unit showed different orientations of one Py unit to the other. Still, these consistently predicted the elongated helices to be on the same face of the helical bundle [Figure 5.15(b)]. While the elongated helix alone would likely be unstable based upon how far it extends past the structured core, this leads me to hypothesize that this helix could potentially be stabilized by interaction with another of elongated helix either in an inter-protein or inter-Py unit manner. If this does lead to slight changes in the structural packing as predicted by AlphaFold, this would likely lead to a similar but non-uniformly perturbed peak pattern in the ^{1}H-^{15}N HSQC, accounting for what was observed experimentally (Figure 5.14).

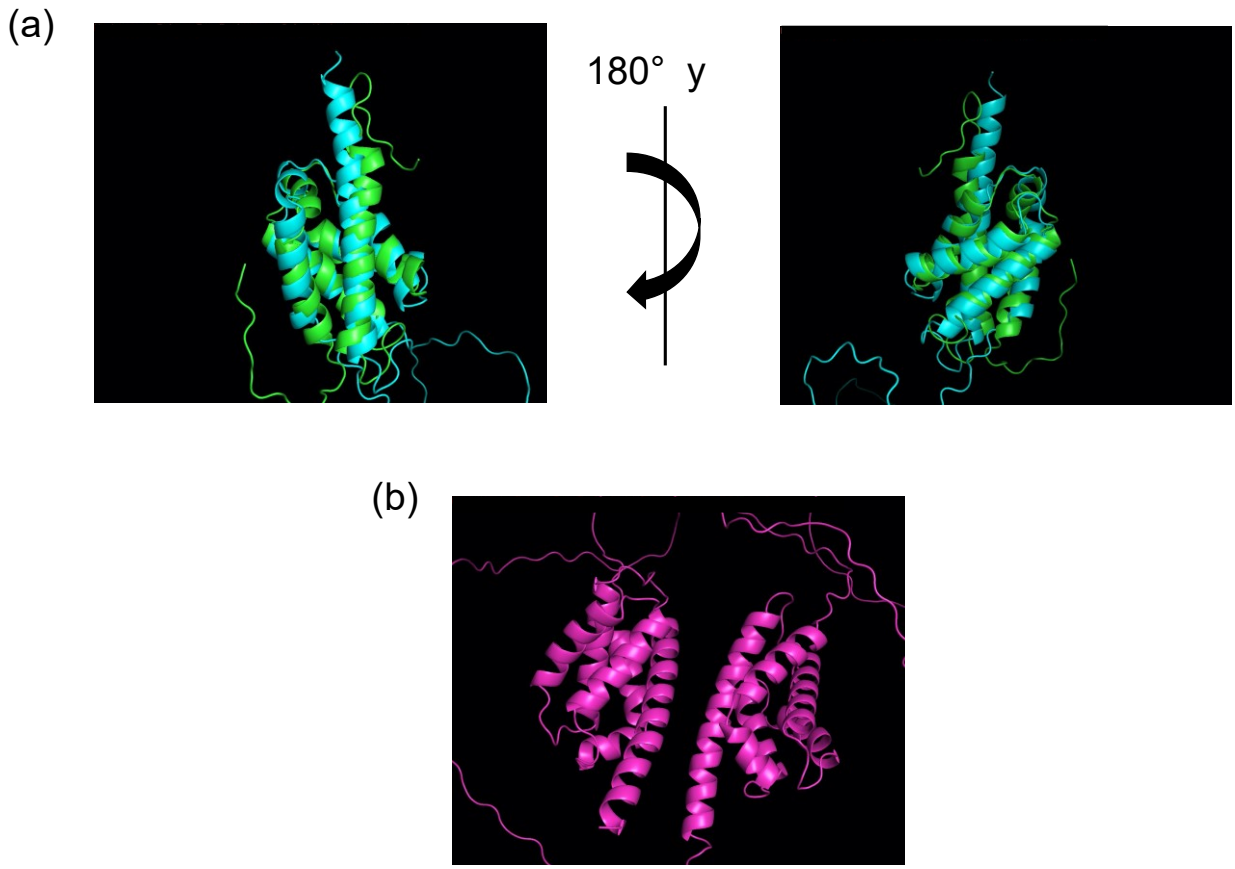


Figure 5.15. Lowest energy AlphaFold prediction of Δ196-230 HPy_1 relative to Δ9-70+Δ206-242-HPy_1 (a) and lowest energy AlphaFold structure of Δ196-230 HPy-Δ196-230 (b).

5.3.3. Temperature-Based Chemical Shift Perturbation and Local Structural Environment

To evaluate the atomic-level temperature stability for Δ9-70+Δ206-242-HPy_1, ^{1}H-^{15}N spectra were collected at the temperature used for all other NMR experiments presented so far, 40 °C, and two elevated temperatures, 45 and 50 °C, and $\Delta\delta$ values were calculated (Figure 5.16). Of particular interest are residues 108-110, which showed a higher chemical shift displacement at 50 °C relative to the remainder of the protein (Figure

5.16, Figure A5). Based on the proximity of these residues to F15, this would be consistent with a weakening of interactions with the phenyl ring, causing these residues to be more shielded and giving a larger upfield shift than observed in other residues. Given the correlation with $\Delta\delta^{120}$, it can potentially be predicted that the high change in chemical shift here is the result of a change in secondary structure at the beginning of helix-6. Individual Δppm values for ^{15}N and ^{1}H are shown in Figure A5. It should be noted that precipitation was observed to have occurred upon sample inspection following experimental data collection at an unknown point in the temperature ramp, meaning that the protein concentration would be lower than initially measured.

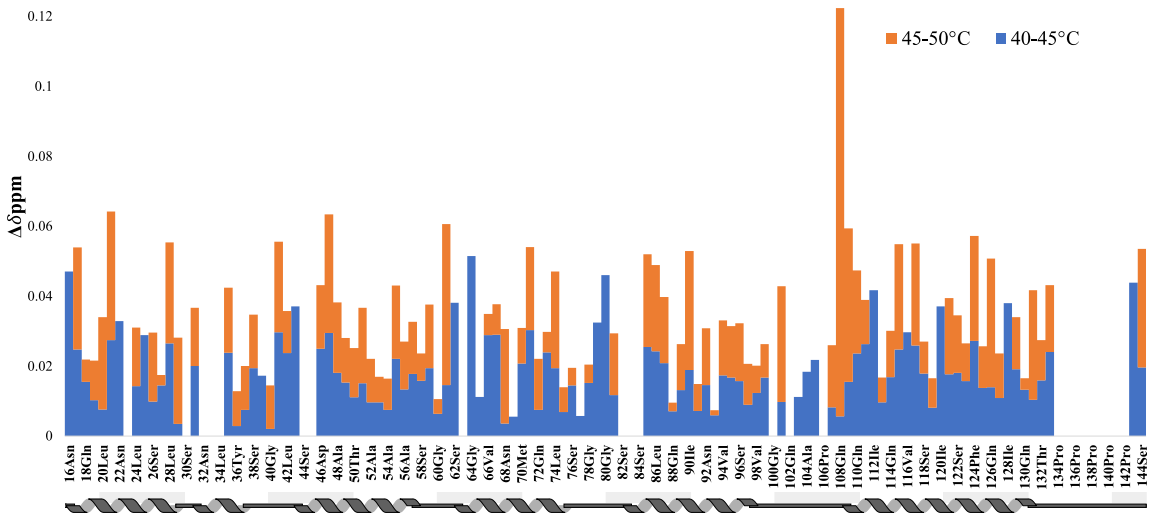


Figure 5.16. Combined Δδ of temperature-dependent chemical perturbations for Δ9-70+Δ206-242-HPy$_1$. Spectra were processed by me and analyzed by Skylah McLeod van Wagoner.

5.3.4. Solid-State NMR and Insight Into the PySp1 α→β Transition

To model the best combination of amino acids for selective unlabelling in a simpler protein than HPy_2, HPy_1 was used as a model system to characterize efficiency of amino acid unlabelling. Two supplementation schemes were used (Table 5.1, cocktails B1 and B2) based on previous aciniform silk studies carried out in the Rainey lab. The first of these (cocktail B1) showed low signal-to-noise in the ^{1}H-^{15}N HSQC indicative of high degrees of non-specific scrambling taking place. To compensate for this presumed non-specific scrambling, His, Glu, and Met were excluded from the supplementation mixture in the second expression scheme. As many of the residues of interest were observed in the ^{1}H-^{15}N HSQC for HPy_1 produced under these conditions (Table 5.4, cocktail B2), the concentration of each amino acid in M9 was doubled for the expression of HPy_2 (cocktail C1). HPy_2 expressed with supplementation scheme C1 showed a lower signal to noise relative to HPy_1 expressed with condition B2 and fewer of the targeted residues were apparent (Table 5.4), indicating issues with non-specific scrambling using these concentrations, which can be addressed in future by slightly lowering concentrations of supplemented amino acids (particularly ones at higher concentration like Gln, Pro, and Ser). Having expressed (cocktail C1), purified, dialyzed and freeze dried the HPy_2 protein, it was spun into fibres and used for solid-state NMR data collection due to instrumental constraints and time constraints around COVID-induced laboratory lockdowns at both Dalhousie and at the National High Magnetic Field Laboratory where the NMR experiments were carried out.

Table 5.4. Residues identified in HPy_1 and HPy_2 selectively unlabelled using schemes B1, B2, and C1. Targeted residues are highlighted in red.

HPy1	Fraction Identified		
	B1	**B2**	**C1**
Arginine (R): 8	2/8	4/8	4/8
Asparagine (N): 11	3/11	7/11	4/11
Aspartic Acid (D): 1	0/1	1/1	1/1
Cystine (C): 0	0	0	0
Glycine (G): 9	0/9	1/9	0/9
Glutamine (Q): 34	1/34	2/34	0/34
Glutamic Acid (E): 2	1/2	1/2	1/2
Histidine (H): 6	0/6	0/6	0/6
Isoleucine (I): 7	0/7	1/7	0/7
Leucine (L): 12	0/12	0/12	0/12
Lysine (K): 0	0	0	0
Methionine (M): 2	0/2	1/2	0/2
Phenylalanine (F): 4	2/4	3/4	2/4
Proline (P): 19	N/A	N/A	N/A
Serine (S): 55	4/55	14/55	0/55
Threonine (T): 8	2/8	7/8	4/8
Tyrosine (Y): 4	2/4	3/4	4/4
Tryptophan (W): 0	0	0	0
Valine (V): 17	0/17	2/17	1/17

While a lot of fibre spinning conditions have been presented in this thesis for HPy_2 proteins expressed in both LB and M9 media, I targeted 2× stretching in air for solid-state NMR sample preparation as this condition have previously yielded the most mechanically consistent fibres. This condition includes a post-spin stretching step that should (at least based on precedents for other more extensively studied silks[131] and results shown in Figures 2.10 and 2.11) induce structural transition to β-sheet and increase alignment of structured regions of the protein in the fibres, improving sample mechanics and homogeneity. Before sending samples to Florida for data acquisition, a random sampling of fibre segments was mechanically tested to confirm that the observed overall

mechanical behaviour fell within expected ranges (Figure 5.17; stress at failure: 62.7 ± 9.9, strain: 128 ± 75, toughness: 69.2 ± 42.2, Young's Modulus: 1.79 ± 0.23, diameter: 17.9 ± 1.8)

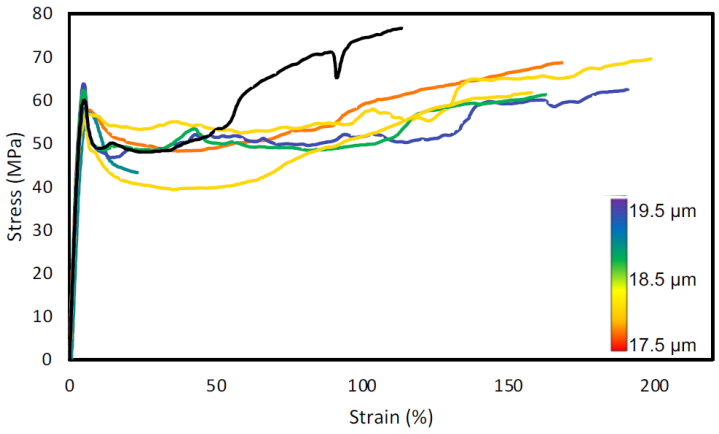


Figure 5.17. Mechanical results of fibres spun from selectively unlabelled HPy_2 and stretched 2× in air for solid-state NMR analysis (n=8). The black line indicates a fiber 13.6 µm in diameter. This is a reproduction of figure 3.4(b)

Having confirmed that the fibres' mechanical properties were similar to those of other fibres produced under these spinning conditions, as detailed in Chapter 3 (Table 3.5), solid-state NMR experiments were carried out. ^{13}C and ^{15}N CP-MAS data were collected to evaluate the NMR spectroscopic properties of the fibres as well as the linewidths and level of labelling. I have annotated both the ^{13}C and ^{15}N CP-MAS spectra on the basis of expected random coil chemical shifts[84, 132] to evaluate the types of residues being observed in the HPy_2 fibres (Figures 5.18, Figure 5.19). From this annotation, all expected residue types were observed in the spectra. There was also a clear presence of Ala in the chemical shift data, indicating some Ala residues were still labelled despite the lack of observation of these residues in the solution-state spectra.

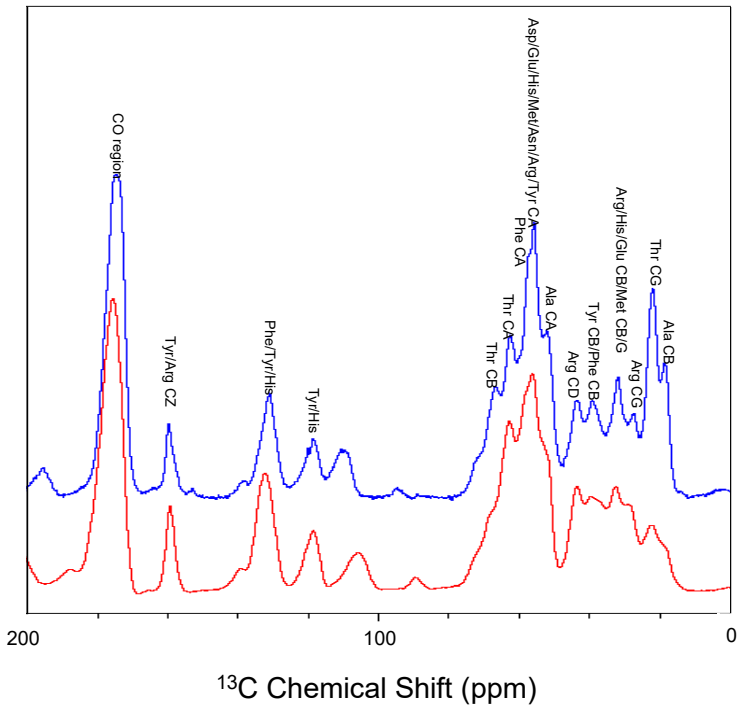


Figure 5.18. ^{13}C CP-MAS spectra with (red) and without (blue) DNP enhancement annotated based on random coil chemical shifts[84].

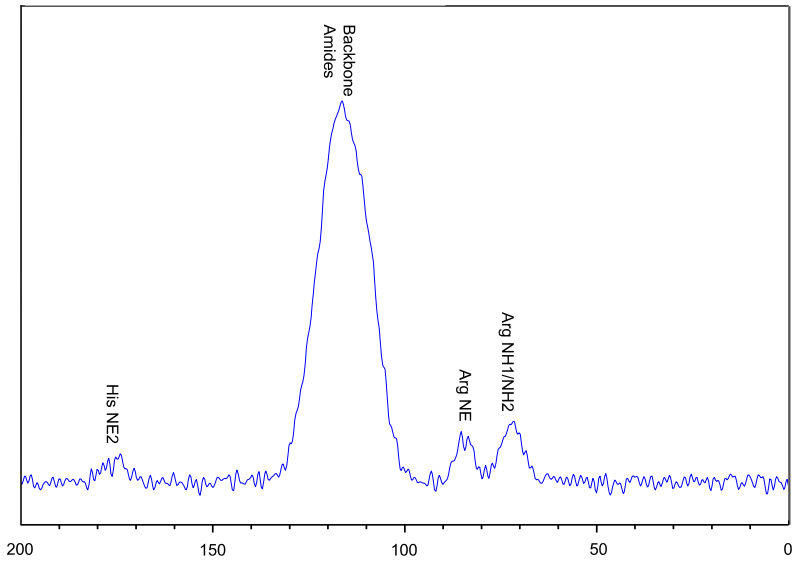


Figure 5.19. ^{15}N CP-MAS spectra without DNP enhancement annotated based on random coil ^{15}N chemical shifts[132].

As signal-to-noise was relatively modest, experiments using DNP enhancement were also acquired. A substantial enhancement in signal was observed (~50-60 fold) for all ^{13}C resonances other than the methyl groups of Thr, Ala and Met (Figure 5.20). The enhanced signal intensity seemed promising, with further experiments then pursued.

2D ^{13}C-^{13}C DARR spectra were acquired both under standard MAS conditions and using DNP enhancement. As with the 1D ^{13}C CP-MAS experiment, a significant intensity increase was observed for the sample with DNP enhancement. However, the resolution was not increased in the DNP-enhanced ^{13}C-^{13}C DARR spectrum (Figure 5.20), which is likely the result of the low temperature needed for DNP (~100 K) leading to signal broadening. As such, the ^{13}C-^{13}C DARR data acquired at 18.8 T without DNP modulation were used for the final data analysis.

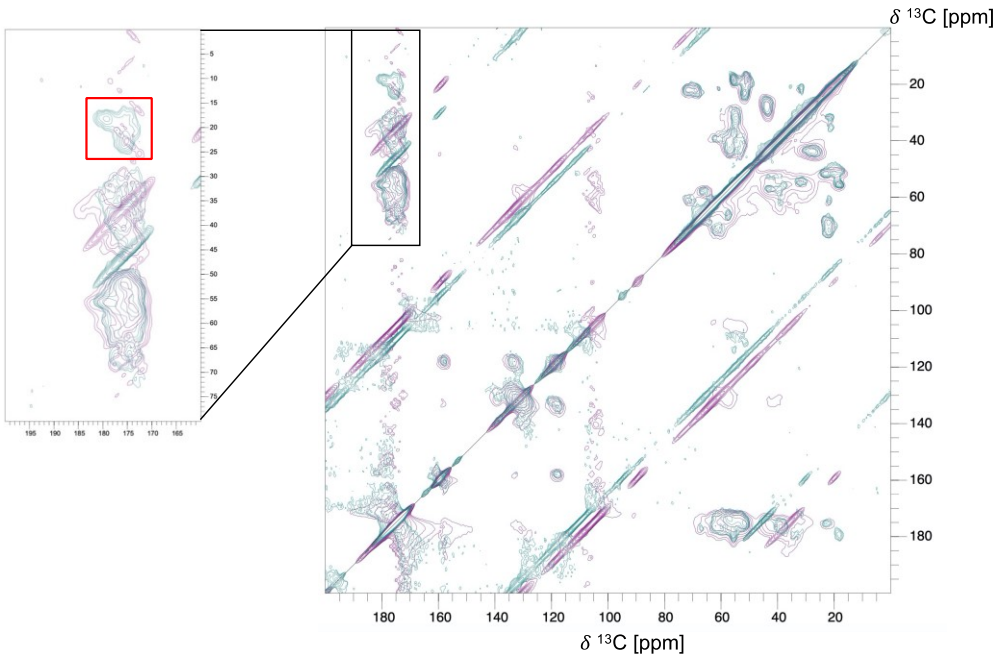


Figure 5.20. Overlay of ^{13}C-^{13}C DARR spectra collected with (purple) and without (green) DNP enhancement. The region not resolvable with DNP is shown in the red box.

Following peak picking, manual evaluation of each set of directly-correlated ^{13}C resonances in the DARR spectrum (50 ms mixing time, with other mixing times analyzed in the future) by PLUQin[124] was consistent with a mix of α-helix, β-sheet, and coil structures (Table 5.5), indicating partial α→β conversion in the spun fibres. While the ability to distinguish between individual peaks in some highly overlapped regions left some parts of the spectrum too ambiguous to assign robustly, PLUQin analysis indicated an α→β conversion in helix-3 and helix-4. Correlations observed between corresponding to Asp, Arg, and Met respectively were calculated by PLUQin as having β-sheet character and while Arg is in multiple locations in the protein, Asp and Met (other than the first Met residue) are unique to helix-3 and helix-4 respectively. As well, the correlation found for Glu was calculated to be in a random coil environment. Given the locations in the soluble protein, this could either indicate the only correlation that could be distinguished is for the residue in the disordered region, or helix-6 is converting to a disordered structure during the fibre forming process. Further, two non-targeted residues were clearly observed in the ^{13}C-^{13}C DARR spectra, Ala and Val (Figure 5.21). Ala residues were observed in sheet, coil, and helix structures, and Val was observed with a strong helix correlation and a weak β-sheet correlation. PLUQin calculated secondary structures of targeted residues observed in the ^{13}C-^{13}C DARR spectra are tabulated below compared to secondary structure distribution in the current 20-member solution-state ensemble (Table 5.5), with the ^{13}C-^{13}C DARR spectra annotated in Figure 5.20 showing key peak assignments.

Table 5.5. Tabulated results for residues per repeat unit and calculated secondary structures in the solution and fibre states.

Targeted Amino Acids and number present in HPy_1	Solution-State Structure	Solid-State Structure
Arg (8) R	Disordered/Pro Rich (5) Helix 3 (1) Helix 6 (2)	Sheet
Asn (11) N	Disordered (4) Helix 1 (2) Helix 2/Disordered (1) Helix 2 (2) Helix 4 (1) Helix 5 (1) Helix 6 (1)	Unable to Identify in Data
Asp (1) D	Helix 3 (1)	Sheet
Glu (2)E	Disordered (1) Helix 6 (1)	Coil
His (6) H	Disordered His-6 tag (6)	Not Diagnostic
Met (2) M	First Residue (1) Helix 4 (1)	Sheet
Phe (4) F	Disordered (3) Helix 6 (1)	Helix
Thr (8) T	Disordered (3) Helix 1 (1) Helix 2 (1) Helix 4 (1) Helix 5 (1) Helix 6/Disordered (1)	Helix/Coil
Tyr (4) Y	Disordered (3) Helix 2 (1)	Inconclusive due to overlap

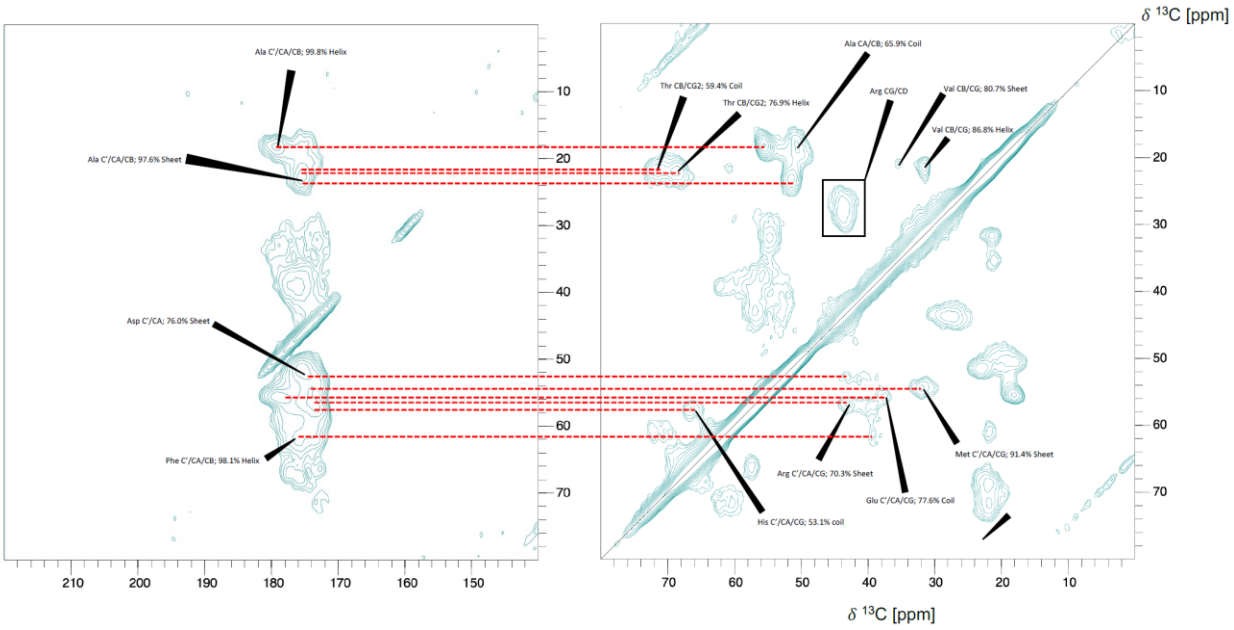


Figure 5.21. ^{13}C-^{13}C DARR spectra annotated based on PLUQin[124] assignment. Percentages shown are the highest percent probability PLUQin assigns to a given amino acid giving rise to that peak would have the indicated chemical shift. Red lines indicate if multiple peaks believed to belong to the same residue type were used for the PLUQin analysis.

5.3.5. Supramolecular Characterization of the Assembled Fibre

To understand the supramolecular assembly of recombinant pyriform fibres, I started by testing if there were any optically observable differences in fibre features beyond non-uniformities in cross-sectional shape observed by SEM. Based on literature precedents[111, 112] and the utility of the technique once optimized, I approached this through non-linear optical techniques, specifically SHG microscopy. Knowing precedents exist for SHG of recombinant spider silks, I wanted to determine if that phenomenon could be used to give more insight into higher-ordered assembly within HPy_2 fibres than was observed by SEM.

THG or SHG alone were no more diagnostic of structure than SEM shown in Chapter 2 if collected directly for the fibre (Figure A6). However, when SHG images were taken at different angles and averaged (SHG-PIPO), thin (1-2 μm), elongated regions giving consistent SHG signal were evident throughout the fibre (Figure 5.22). Features of this nature were not evident by SEM (Figure 2.10). Due to SHG's reliance on asymmetric structure in a material, the inclusion of PIPO likely allows visualization of microfibrillar units within the assembled silk fibres. A variety of ranges of microfibril diameter were observed based on two images of the same fibre, with some as small as 1.0 μm and others as large as 5.5 μm. With further investigation, the density and diameter of microfibrils could be linked with structural information to determine how fibre processing and atomic-level structural changes affect microfibril diameter and, therefore, mechanical properties.

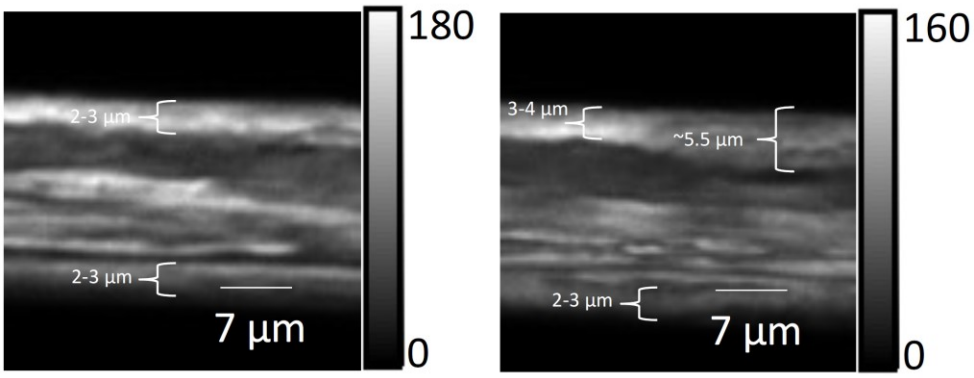


Figure 5.22. Summed images from SHG-PIPO analysis of fibres stretched 2x in EtOH collected at two locations on the same fibre.

5.4. Discussion

From folding studies based on comprehensive chemical shift and correlation assignments of Δ9-70+Δ206-242-HPy$_1$, the 6 helices predicted in Chapter 4 were found to be tightly

packed. This is similar to structures reported for repetitive domains of AcSp1, MaSp1, TuSp1, and TuSp2 from *N. antipodiana,* (Figure 5.23). Of these structures, Δ9-70+Δ206-242-HPy_1 was found to overlay well with the *N. antipodiana* TuSp1-RP2[43] and MiSp1 structures [Figure 5.23(b, c)] with only slight offsets between helices and similar lengths. Of all the reported structures, the only one that appears to be completely different from Δ9-70+Δ206-242-HPy_1 is the AcSp1 repeat unit structure from *A. trifasciata*[29], which is far less compact and is missing the helical segment bridging over the structure. This indicates the general structure observed in the other 5 structures (pyriform repeat unit included) might be a more common repeat unit structure in the solution state, with repeat units from some species and silk classes having the potential for drastic deviations.

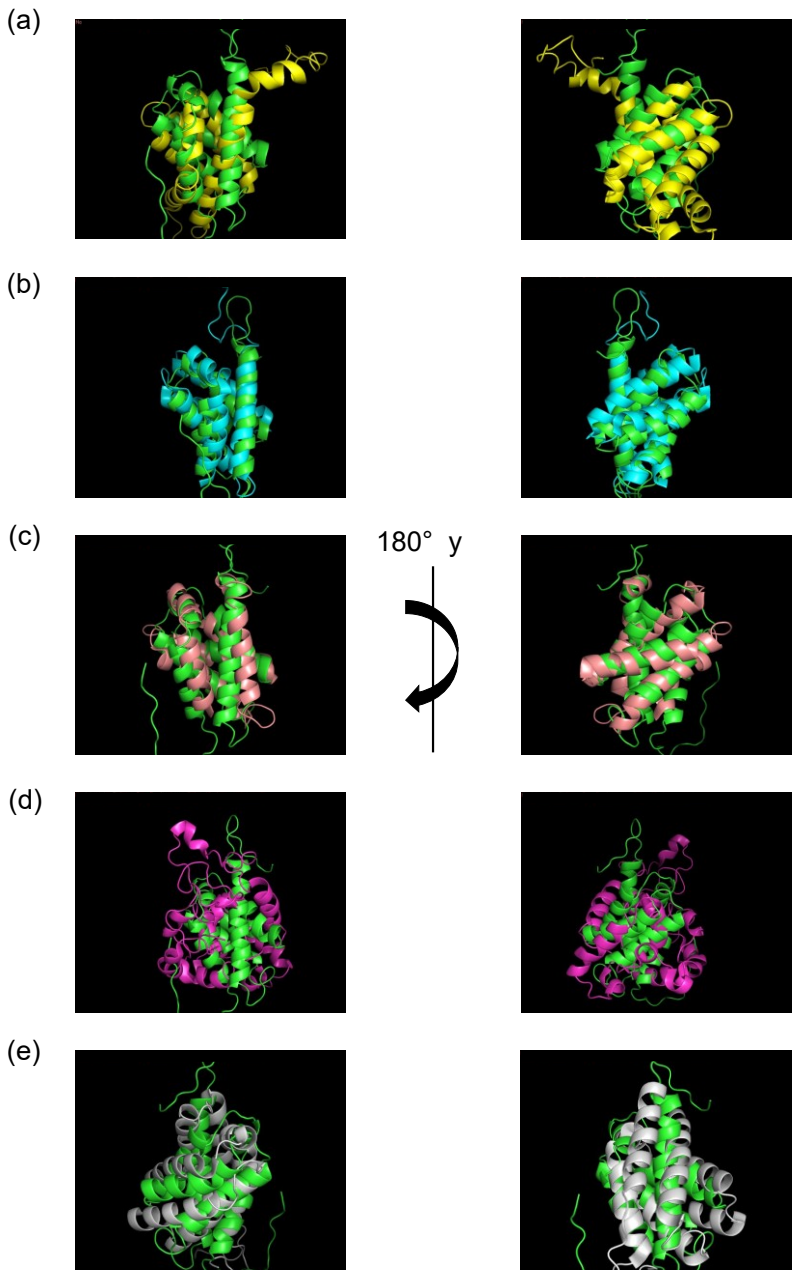


Figure 5.23. Overlay of the lowest energy Δ9-70+Δ206-242-HPy$_1$ ARIA structure after water refinement (green) with reported repetitive domain structures. Shown as well are structure for TuSp1-RP1[43] (a; PDB DOI: 10.2210/pdb2K3N/pdb), TuSp1-RP2 (b; PDB DOI:10.2210/pdb2K3O/pdb), MiSp1 (c; PDB DOI: 10.2210/pdb2MQA/pdb), and AcSp1 from *A. trifasciata*[29] (d; PDB DOI: 10.2210/pdb2MU3/pdb) from and *N. antipodiana*[32] (e; PDB DOI: 10.2210/pdb2LYI/pdb).

Structures generated from the removal of the PX-rich motif indicated extension of helix 6 into a semi-stable conformation and slight changes to the core structure as a result, which is supported by non-uniform chemical shift disturbances in ^{1}H-^{15}N HSQC data at a higher concentration, then restoration upon precipitation. This suggests a role of the C-terminal motif in limiting the length of helix-6, creating a stable core structure. Changes in structural stability as a function of temperature are also particularly evident around 109E, the first residue of helix-6, which is flanked by two glutamine residues. The potential interaction of these three residues with 15F, as indicated both by anomalous chemical shifts and through NOE-derived distance restraints, would provide a strong energetic driving force to keep this helix close to the core of the protein. These data might suggest a mechanism of fibre formation which is triggered by decompaction of helix-6, analogous to the C-terminal decompaction proposed for aciniform silk fibrillogenesis[41].

Examination of the atomic-level and microscopic fibre structural features showed that the mixed α-helical/coil pyriform silk protein soluble state (Figure 5.21) undergoes an $\alpha \rightarrow \beta$ conversion upon fibrillogenesis, resulting in a supramolecular assembly of aligned microfibrils as previously reported in other silks[133]. Based on solid-state NMR analysis, helix-3 and helix-4 are believed to undergo this structural conversion, while helix-6 might be converting to a disordered structure during fibrillogenesis. If this is found to be the case, it could indicate a decompaction during fibre spinning similar to the process during aciniform silk fibre formation[41]. In the next chapter, I will relate these findings to the rest of my thesis, to highlight what key insights have been developed from this work, and future experiments relating to these developments.

Chapter 6

Conclusions and Future Directions

6.1. Introduction

The previous chapters detailed the work I carried out to understand the mechanical and structural properties of recombinant pyriform silks. In this Chapter, I summarize the key mechanical (Section 6.2) and structural (Section 6.3) conclusions and discuss future work for each. In Section 6.4, I explain the broader significance of my work to highlight the impact of these key results.

6.2. Mechanical Characterization

6.2.1. Pyriform Silk Mechanical Properties

Through the work presented in this thesis, I developed methods of expressing and purifying recombinant HPy_2 proteins based on the repetitive domain of *A. argentata* PySp1. This allowed for the first in-depth characterization of pyriform silk without confounding effects from glue components and allowed me to specifically develop an understanding of the central repetitive domain of *A. argentata* PySp1. Fibres wet-spun from HPy_2 proteins were strong relative to the molecular weight of HPy_2, with moderate extensibility. These fibres exhibited increased strength upon post-spin stretching, with subsequent increases in extensibility when exposed to 40% EtOH during the stretching process.

Expressing the protein in M9 medium with a tailored amino acid supplementation scheme and subsequent IMAC purification led to significant increases in protein yield and fibres with similar strength to HPy_2 expressed in LB medium. However, these fibres appeared mechanically insensitive to differences in the applied stretching conditions relative to LB-expressed HPy_2. Based on SDS-PAGE and mechanical testing results, a qualitative difference was observed in the appearance of lower proportions of truncations of the protein in the M9-expressed HPy_2 relative to that expressed in LB-medium. Thus, expressing the protein in M9 medium with amino acid supplementation could create a protein that can entangle more uniformly with other proteins, giving more extensible fibres. This method of high yield expression, with optimization, could likely be applied to other silks to increase expression yield and potentially improve mechanical properties without the need for post-spin processing.

6.2.2. Mechanical Characterization Future Directions

6.2.3.1. Mechanical Characterization of HPy_1, HPy_3, and Proteins without the PX-rich motif

During this thesis work, fibre spinning and mechanical testing focused on HPy_2 as a model protein. However, precedents exist for spinning fibres from the recombinant pyriform silk proteins with 1-3 repeat units[39, 40], with fibres formed using proteins containing more repeat units giving higher tensile strength as observed in other silks[40, 44, 61]. HPy_1 will likely be weaker than HPy_2 based on previous findings from Zhu et al.[40]; however, changes in tensile strength of fibres spun from recombinant proteins with differing numbers of repeat units based on *A. argentata* pyriform silk are currently uncharacterized. Wet-spinning

higher MW recombinant proteins could allow for post-spin processing at higher stretching ratios, as previously performed in our laboratory for a aciniform silk[47].

During syringe filtration of soluble Δ196-230 HPy_1 proteins in phosphate buffer, I observed the formation of fibres from the end of the filter (Figure 6.1), implying that HPy_1 and other one repeat proteins are amenable to fibre formation. What makes the Δ196-230 HPy_1 truncation particularly interesting for follow-up studies is the appearance of the PX-rich motif in all pyriform silk protein sequences solved to date[1, 14], and the apparent solution-state structural changes observed from the removal of the PX-rich motif (Figure 5.14). Knowing that this might lead to different mechanical properties in silks spun from Δ196-230 HPy proteins, wet-spinning and mechanically testing one and two repeat HPy recombinant proteins with this deletion will elucidate whether the PX-rich motif has a mechanical role in the wet-spun fibres. From results of these experiments, we can then infer whether the presence of the PX-rich motif plays a key functional role in the fibres, or in stipulated interactions with pyriform glue components.

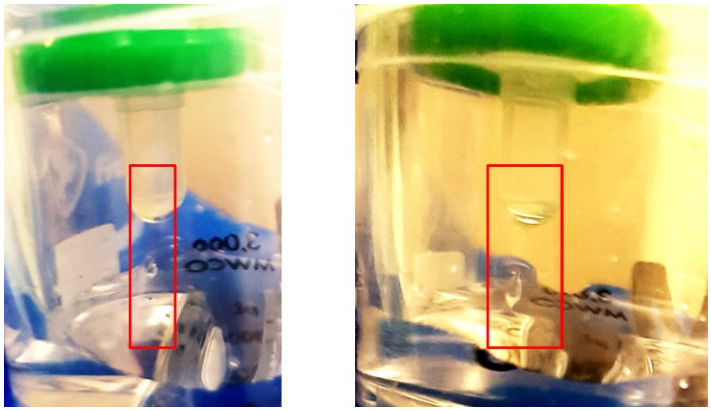


Figure 6.1. Fibre-like structures observed when syringe filtering Δ196-230 HPy_1. The image on the left shows a buffer droplet on the fibre, while the image on the right shows a fibre-like structure applying surface tension to the top and bottom of a buffer droplet.

6.3. Structural Characterization

6.3.1. Solution State Structure of Pyriform Silk and Structural Transition

Through rigorous NMR analysis of proteins representative of the repetitive domain of *A. argentata* pyriform silk, I found that the repetitive domain contains a modular "beads-on-a-string" architecture with a tight bundle of 6 α-helices in the core, and long, disordered terminal regions linking the "beads". This general structure was similar to all central repetitive domain structures that have been solved and reported in the literature, with the exception of one solved aciniform silk[29], implying potential for similarity in structural packing in this region across silks despite the differences in amino acid composition and dispersion, repeat unit lengths, and sequence motifs. Furthermore, with AlphaFold[119], predictions appeared highly accurate for the well-folded pyriform proteins (Figure 5.12). What remains challenging with interpreting these AlphaFold data, however, is the fact that some of the proposed structures had key differences in helical bundle topology – without experimental data, it would be hard to rigorously pick between the two predicted topologies. Keeping this caveat in mind, I would propose that AlphaFold[119] can be used as a screening tool to see if other pyriform members are predicted to share similar central repetitive domain tertiary structures or if newly discovered sequences are forming a tight α-helix bundle in the core or a more loosely clustered core.

Using solid-state NMR spectroscopy, we were able to observe correlated sets of ^{13}C resonances consistent with a mixture of α-helix, β-sheet, and random coil structuring, indicating that partial but not complete α→β transition occurs in HPy_2 fibres when stretched 2× in air. Of the signals observed, Asp and Met were found to be β-sheet, while

Glu was found to be disordered. This indicates an α→β transition for helix-3 and helix-4, with potential loss of α-helical structure in helix-6. While further rigorous analysis and data collection is still needed to further elucidate what is happening with the rest of the protein once spun into fibers, this provides the first atomic-level evidence of the α→β transition in pyriform silks.

6.3.2. Structural Characterization Future Directions

6.3.3.1. Final Refinement of Δ9-70+Δ206-242-HPy$_1$ Structure

Having obtained a highly converged ensemble of structures for Δ9-70+Δ206-242-HPy$_1$, the last steps before finalizing the structure are to minimize the number of dihedral angle and NOE violations, and to set ARIA to trust peaks now that I am confident in the assigned correlations. As discussed in Section 5.3.1, ARIA is currently using about 95% of the peaks, relatively evenly removing peaks from all types of observable contacts from data analysis. After an iteration of ARIA with the "trust peaks" setting enabled, more inter- and intra-residue NOE contact violations were observed, indicating that one of the data exclusion sources was ARIA trying to minimize NOE errors. With some errors observed being the result of apparently overly tight constraints despite very clear and unambiguous resonance assignments (e.g., a pair of atoms observed to be 4 Å apart in the structural ensemble, with CcpNmr's NOE calibration algorithm setting a distance restraint maximum of 3.5 Å and ARIA being used in such a way that it will not extend NOE restraint upper bounds), this requires some manual editing of restraints to be more representative of the data. This is frequently required, as CcpNmr relies on a very rigid restraint length calibration routine in comparison to the broader NOE "bin" classification approach that

has long been the norm for NMR structure calculations[134]. Also, dihedral angle restraints that are flagged as frequently violating by > 0.5 Å will be evaluated to determine if they result from software limitations in the DANGLE interface in CcpNMR Analysis (i.e., with dihedrals close to 180°, where value near -180° appear excluded) or whether these are strictly falling in disordered regions in the termini where structural convergence is less likely. As of now, no concerning dihedral or NOE restraints were observed in the analysis. Once this final manual refinement of NOE and dihedral angle restraints is completed with the remaining violations minimized, I will have finalized the first atomic-level structure of a PySp1 central repetitive domain.

6.3.3.1. Mutagenesis, Temperature-Based Denaturation and Disulfide Locking

To develop a more robust insight into the fibre forming process of these pyriform silk fibres, atomic-level temperature denaturation studies can be continued by collecting ^{1}H-^{15}N HSQC spectra over a wider range and with smaller step sizes (i.e., 2.5 °C). Residue specific $\Delta\delta$ values can then be compared to determine which regions are most impacted by the increase in temperature. This would show us which helices are the most impacted with increased temperature, and this lower stability would allow us to infer which are the first to unfold during fibre spinning. Using that information, an analogous strategy to previous studies in our lab could be performed[41], where helices found to be the most perturbed could be reengineered to allow for disulfide locking and checked via ^{1}H-^{15}N HSQC spectra to ensure the solution state structure is preserved. Fibres could then be

spun from the disulfide locked proteins to determine if fibre formation is prevented, indicating which helices are important for initiating the fibrillogenesis process.

6.3.3.2. Atomic Level Elucidation of Fibre Assembly Process

To gain more insight into exactly which changes are occurring during the fibrillogenesis process, further optimization needs to be performed on the ratios of supplemented amino acids used in HPy_2 expression through NMR of selectively labelled HPy_2. While it appears that non-targeted residues visualized by solution-state NMR of HPy_1 were unlabelled by doubling the amino acid concentrations for HPy_2, this came at the cost of signal-to-noise by increased non-specific scrambling, and loss of labelling efficiency for targeted residues (Table 5.4). In addition, Ala and Val, which were not visible in ^{1}H-^{15}N HSQC spectra, were evident in the DARR spectrum. This indicates a need to further optimize the concentration of amino acid supplementation in the M9 medium and ensure that the level of nonspecific scrambling that takes place does not lead to loss of labelling efficiency that, in turn, is detrimental for the signal-to-noise of solid-state experiments. Proteins would, therefore, likely need to be expressed with a slightly lower concentration of most supplemented amino acids, with the exception being an increase in the Ala concentration. Optimizing this ratio and obtaining higher signal-to-noise without relying on DNP experiments can allow for observation of cross-peaks with less broadening and, therefore, better delineation of resonances in general. Once this is better optimized, atomic-level structural differences between fibres that are prepared using different labelling schemes could more readily be explored via this analysis method.

6.3.3.3. Micro-Scale Structure of Recombinant Pyriform Fibres

As discussed in Section 5.3.4, SHG experiments provide a promising method of analyzing fibre structural features in a non-destructive manner, and the results shown just scratch the surface of what is possible with this technique[111, 112]. Initially, a worthwhile target would be analysis of the consistency and diameter of silk microfibrils. This could provide insight into changes in mechanical properties that are not clearly linked to surface morphology, or secondary structure features between fibres spun from different protein preparations or processed via different wet-spinning conditions. For example, if inconsistencies were found between microfibril sizes and lengths, this could provide a mechanistic reason as to why some fibres have differences in extensibility or strength than others.

One expansion on this technique that would provide a lot of insight is stretching the fibres to determine how the microfibre structures change when a force is applied to the fibres. This was something I had the chance to investigate briefly with MacAulay Harvey, a graduate student in the Tokarz lab at Saint Mary's University. Using a home-built device (Figure 6.2) comprising a frame and two ORIC® 20 mm Linear Translation Stages with Piezoelectric Inertia Drives (Thorlabs; NJ, USA), we could stretch recombinant aciniform fibres by gluing them to a thin sheet of PDMS and applying lines by marker to visualize where the fibre was. This allowed us to easily mount the fibres, find them under the SHG microscope, and collect data on those fibres. Having set these methods up on another recombinant silk, this will allow for the same analysis to be performed on pyriform silk fibres in the future.

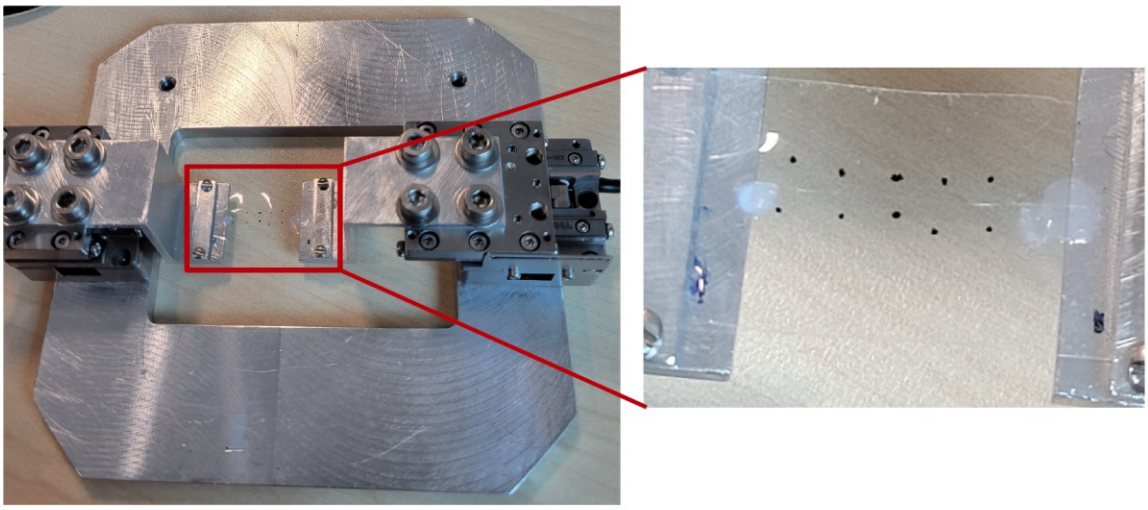


Figure 6.2. Device used to optimize the protocol for fibre stretching with SHG measurements (left) and a zoom-in of the mounted fibre (right, between the blue marks).

To analyze the assembly of pyriform silks, I also carried out TEM analysis (full methods described in Methods A1). Studies on native silk collected following spinning by spiders have indicated the formation of a skin-core structuring in these silks[135]. Examining my recombinant silks via TEM indicated that there could be the presence of skin and core regions as two independent TEM sample preparations carried out on samples from the same sets of fibres both showed an outer area with a differing degree of electron transmission than the fibre core (Figure 6.3). However, as is evident from these images, the observed skin layer was inconsistent between identical preparations, with a smaller layer observed for the second preparation than the first for fibres stretched 2× in 40% EtOH, and a reversal of the relative degree of observed electron transmission between the core and skin sections between the 2× in air preparations. This indicates that, while there is the potential of a skin-core morphology in these recombinant silks, it is also

possible that the observed differences are the result of a difference in sample preparation not yet fully accounted for, such as the effect of osmium tetroxide fixation on the fibres or an artifact coming from the staining for visualization of the sample. Hence, although these preliminary results are compelling, further studies must be conducted with rigorous control of sample preparation and treatment to determine if there is a skin-core morphology in these recombinant silks. If this is found to be a true skin-core morphology, this could mean that mechanical results presented for recombinant fibres may be lower than the actual values of these fibres, particularly if both layers are not able to handle the same amount of tensile force.

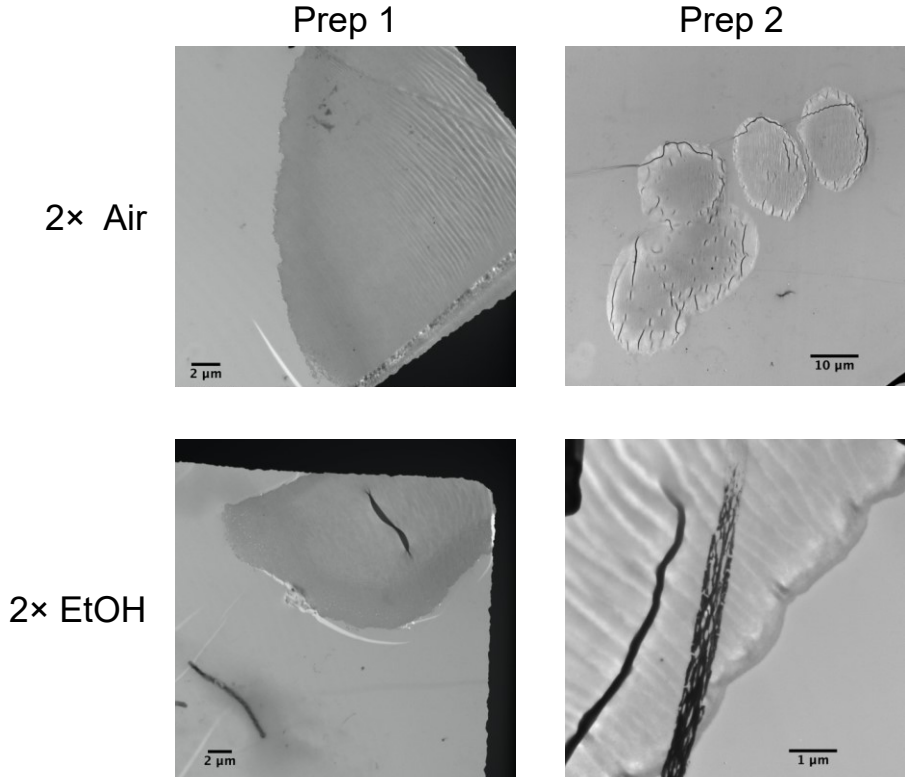


Figure 6.3. TEM images of fibres PS stretched 2× in air (top) and 2× in 40% EtOH (bottom).

6.4. Significance

My work presented in this thesis provided the first mechanical characterization of recombinant pyriform silk fibres representative of the central repetitive domain of *A. argentata* PySp1. In previous studies, the only mechanical behaviour probed was for attachment discs which are composite materials that include pyriform silk among other constituents. My study was, thus, the first to give insight into pyriform silk mechanical properties in general, showing that HPy_2 fibres were strong relative to the size of protein used, with moderate extensibility. Fibre extensibility was significantly enhanced when HPy_2 protein used for wet-spinning was expressed in M9 medium with amino acid supplementation in comparison to protein expressed in LB medium. Expression in M9 with amino acid supplementation also showed significant increases in protein yield coupled with decreases in protein truncation products relative to what could be achieved in LB medium. This provides a simple and cost-effective method of increasing protein production at the laboratory scale, and this method can potentially be applied to other silk proteins.

Examining the structure of these recombinant proteins in the soluble form, I determined the first solution-state structure of a pyriform silk central repetitive domain, showing this to be composed of a 6-helix core with long disordered termini. I also showed for the first time that each central domain repeat unit is modular. In the fibre form, I presented the first atomic level evidence of the $\alpha \rightarrow \beta$ transition in pyriform silk and was able to observe microfibrils in wet-spun fibres. Examining wet-spun pyriform silk fibres by TEM implied the potential for a skin-core morphology which, if confirmed, would be the first reporting

of this type of morphology in a recombinant silk fibre and indicate supramolecular organization beyond microfibril formation.

Bibliography

(1) Chaw, R. C.; Saski, C. A.; Hayashi, C. Y. Complete gene sequence of spider attachment silk protein (PySp1) reveals novel linker regions and extreme repeat homogenization. *Insect Biochem. Mol. Biol.* **2017**, *81*, 80-90. DOI: 10.1016/j.ibmb.2017.01.002.

(2) Eisoldt, L.; Smith, A.; Scheibel, T. Decoding the secrets of spider silk. *Mater. Today* **2011**, *14* (3), 80-86. DOI: 10.1016/s1369-7021(11)70057-8.

(3) Piorkowski, D.; Blackledge, T. A. Punctuated evolution of viscid silk in spider orb webs supported by mechanical behavior of wet cribellate silk. *Sci. Nat.* **2017**, *104* (7-8), 67.

(4) Diaz, C.; Maksuta, D.; Amarpuri, G.; Tanikawa, A.; Miyashita, T.; Dhinojwala, A.; Blackledge, T. A. The moth specialist spider *Cyrtarachne akirai* uses prey scales to increase adhesion. *J. R. So. Interface* **2020**, *17* (162), 20190792.

(5) Hayashi, C. Y., Nichola H. Shipley, N. H., Lewis, R. V. Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *Int. J. Biol. Macromol.* **1999**, *24*, 271–275. DOI: 10.1016/S0141-8130(98)00089-0.

(6) Blackledge, T. A.; Hayashi, C. Y. Silken toolkits: biomechanics of silk fibers spun by the orb web spider Argiope argentata (Fabricius 1775). *J. Expt Biol.* **2006**, *209* (13), 2452-2461.

(7) Wolff, J. O.; Řezáč, M.; Krejčí, T.; Gorb, S. N. Hunting with sticky tape: functional shift in silk glands of araneophagous ground spiders (Gnaphosidae). *J. Expt. Biol.* **2017**, *220* (12), 2250-2259.

(8) Gosline, J. M.; Guerette, P. A.; Ortlepp, C. S.; Savage, K. N. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J, Exp. Biol.* **1999**, *202* (23), 3295-3303.

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

(10) Wallace, J. A.; Shen, J. K. Unraveling a trap-and-trigger mechanism in the pH-sensitive self-assembly of spider silk proteins. *J. Phys. Chem. Lett.* **2012**, *3* (5), 658-662.

(11) Andersson, M.; Chen, G.; Otikovs, M.; Landreh, M.; Nordling, K.; Kronqvist, N.; Westermark, P.; Jörnvall, H.; Knight, S.; Ridderstråle, Y. Carbonic anhydrase generates CO2 and H+ that drive spider silk formation via opposite effects on the terminal domains. *PLoS Biol.* **2014**, *12* (8), e1001921.

(12) Landreh, M.; Askarieh, G.; Nordling, K.; Hedhammar, M.; Rising, A.; Casals, C.; Astorga-Wells, J.; Alvelius, G.; Knight, S. D.; Johansson, J. A pH-dependent dimer lock in spider silk protein. *J. Mol. Biol.* **2010**, *404* (2), 328-336.
(13) Collin, M. A.; Clarke III, T. H.; Ayoub, N. A.; Hayashi, C. Y. Genomic perspectives of spider silk genes through target capture sequencing: Conservation of stabilization mechanisms and homology-based structural models of spidroin terminal regions. *Int. J. Biol. Macromol.* **2018**, *113*, 829-840. Eisoldt, L.; Thamm, C.; Scheibel, T. The role of terminal domains during storage and assembly of spider silk proteins. *Biopolymers* **2012**, *97* (6), 355-361.
(14) Wang, K.; Wen, R.; Wang, S.; Tian, L.; Xiao, J.; Meng, Q. The molecular structure of novel pyriform spidroin (PySp2) reveals extremely complex central repetitive region. *Int. J. Biol. Macromol.* **2020**, *145*, 437-444.
(15) Liu, F. Y.; Liu, J. Y.; Yao, X.; Wang, B. Hybrid sequencing reveals the full-length *Nephila pilipes* pyriform spidroin 1 (PySp1). *Int. J. Biol. Macromol.* **2022**, *200*, 362-369.
(16) Vollrath, F.; Knight, D. P. Liquid crystalline spinning of spider silk. *Nature* **2001**, *410* (6828), 541.
(17) Shao, Z.; Young, R. J.; Vollrath, F. The effect of solvents on spider silk studied by mechanical testing and single-fibre Raman spectroscopy. *Int. J. Biol. Macromol.* **1999**, *24* (2-3), 295-300.
(18) Lefevre, T.; Boudreault, S.; Cloutier, C.; Pezolet, M. Diversity of molecular transformations involved in the formation of spider silks. *J. Mol. Biol.* **2011**, *405* (1), 238-253. DOI: 10.1016/j.jmb.2010.10.052.
(19) Andersson, M.; Holm, L.; Ridderstråle, Y.; Johansson, J.; Rising, A. Morphology and composition of the spider major ampullate gland and dragline silk. *Biomacromolecules* **2013**, *14* (8), 2945-2952.
(20) Heim, M.; Keerl, D.; Scheibel, T. Spider silk: from soluble protein to extraordinary fiber. *Angew. Chem. Int. Ed.* **2009**, *48* (20), 3584-3596.
(21) Jin, H.-J.; Kaplan, D. L. Mechanism of silk processing in insects and spiders. *Nature* **2003**, *424* (6952), 1057-1061.
(22) Parent, L. R.; Onofrei, D.; Xu, D.; Stengel, D.; Roehling, J. D.; Addison, J. B.; Forman, C.; Amin, S. A.; Cherry, B. R.; Yarger, J. L.; et al. Hierarchical spidroin micellar nanoparticles as the fundamental precursors of spider silks. *Proc. Natl. Acad. Sci.* **2018**, *115* (45), 11507-11512.

(23) Hayashi, C. Y.; Lewis, R. V. Spider flagelliform silk: lessons in protein design, gene structure, and molecular evolution. *Bioessays* **2001**, *23* (8), 750-756.
(24) Salles, H. C.; Volsi, E. C. F.; Marques, M. R.; Souza, B. M.; Dos Santos, L. D.; Tormena, C. F.; Mendes, M. A.; Palma, M. S. The Venomous Secrets of the Web Droplets from the Viscid Spiral of the Orb-Weaver Spider *Nephila clavipes* (Araneae, Tetragnatidae). *Chem. Biodivers* **2006**, *3* (7), 727-741.
(25) Townley, M. A.; Pu, Q.; Zercher, C. K.; Neefus, C. D.; Tillinghast, E. K. Small organic solutes in sticky droplets from orb webs of the spider *Zygiella atrica* (Araneae; Araneidae): β-Alaninamide is a novel and abundant component. *Chem. Biodivers.* **2012**, *9* (10), 2159-2174.
(26) Vollrath, F.; Fairbrother, W. J.; Williams, R. J.; Tillinghast, E. K.; Bernstein, D. T.; Gallagher, K. S.; Townley, M. A. Compounds in the droplets of the orb spider's viscid spiral. *Nature* **1990**, *345* (6275), 526-528.
(27) Vollrath, F.; Edmonds, D. T. Modulation of the mechanical properties of spider silk by coating with water. *Nature* **1989**, *340* (6231), 305.
(28) Hayashi, C. Y.; Blackledge, T. A.; Lewis, R. V. Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. *Mol. Biol. Evol.* **2004**, *21* (10), 1950-1959. DOI: 10.1093/molbev/msh204.
(29) Tremblay, M. L.; Xu, L.; Lefevre, T.; Sarker, M.; Orrell, K. E.; Leclerc, J.; Meng, Q.; Pezolet, M.; Auger, M.; Liu, X. Q.; et al. Spider wrapping silk fibre architecture arising from its modular soluble protein precursor. *Sci. Rep.* **2015**, *5*, 11502. DOI: 10.1038/srep11502.
(30) Sarker, M.; Orrell, K. E.; Xu, L.; Tremblay, M.-L.; Bak, J. J.; Liu, X.-Q.; Rainey, J. K. Tracking Transitions in Spider Wrapping Silk Conformation and Dynamics by ^{19}F Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **2016**, *55* (21), 3048-3059. DOI: 10.1021/acs.biochem.6b00429.
(31) Tremblay, M.-L.; Xu, L.; Sarker, M.; Liu, X.-Q.; Rainey, J. Characterizing aciniform silk repetitive domain backbone dynamics and hydrodynamic modularity. *Int. J. Mol. Sci.* **2016**, *17* (8), 1305.
(32) Wang, S.; Huang, W.; Yang, D. NMR structure note: repetitive domain of aciniform spidroin 1 from *Nephila antipodiana*. *J. Biomol. NMR* **2012**, *54* (4), 415-420.

(33) Blasingame, E.; Tuton-Blasingame, T.; Larkin, L.; Falick, A. M.; Zhao, L.; Fong, J.; Vaidyanathan, V.; Visperas, A.; Geurts, P.; Hu, X. Pyriform spidroin 1, a novel member of the silk gene family that anchors dragline silk fibers in attachment discs of the black widow spider, Latrodectus hesperus. *J. Biol. Chem.* **2009**, *284* (42), 29097-29108.
(34) Sahni, V.; Harris, J.; Blackledge, T. A.; Dhinojwala, A. Cobweb-weaving spiders produce different attachment discs for locomotion and prey capture. *Nat. Commun.* **2012**, *3*, 1106. DOI: 10.1038/ncomms2099.
(35) Kovoor, J., Zyleberberg, L. Fine Structural Aspects of Silk Secretion In a Spider (*Araneus diadematus*). I. Elaboration in the Pyriform Glands. *Tissue Cell* **1980**, *12* (3), 547-556.
(36) Kovoor, J.; Zylberberg, L. Fine structural aspects of silk secretion in a spider. II. Conduction in the pyriform glands. *Tissue Cell* **1982**, *14* (3), 519-530.
(37) Wolff, J. O.; Grawe, I.; Wirth, M.; Karstedt, A.; Gorb, S. N. Spider's super-glue: thread anchors are composite adhesives with synergistic hierarchical organization. *Soft Matter* **2015**, *11* (12), 2394-2403. DOI: 10.1039/c4sm02130d.
(38) Hesselberg, T.; Vollrath, F. The mechanical properties of the non-sticky spiral in Nephila orb webs (Araneae, Nephilidae). *J. Exp. Biol.* **2012**, *215* (19), 3362-3369.
(39) Geurts, P., Zhao, L., Hsia, Y., Gnesa, E., La Mattina, C., Franz, A., Larkin, L., ; Tang, S., Jeffery, F., Vierra, C. Synthetic spider silk fibers spun from pyriform spidroin 2, a glue silk protein discovered in orb-weaving spider attachment discs. *Biomacromolecules* **2010**, *11*, 3495–3503. DOI: 10.1021/bm101002w.
(40) Zhu, H.; Rising, A.; Johansson, J.; Zhang, X.; Lin, Y.; Zhang, L.; Yi, T.; Mi, J.; Meng, Q. Tensile properties of synthetic pyriform spider silk fibers depend on the number of repetitive units as well as the presence of N-and C-terminal domains. *Int. J. Biol. Macromol.* **2020**.
(41) Sulekha, A. *DISULPHIDE LOCKING* : CONTRASTING EFFECTS ON DISPARATE PROTEINS Dalhousie Department of Biochemistry & Molecular Biology PhD Thesis, 2021.
(42) Arndt, T.; Greco, G.; Schmuck, B.; Bunz, J.; Shilkova, O.; Francis, J.; Pugno, N. M.; Jaudzems, K.; Barth, A.; Johansson, J. Engineered Spider Silk Proteins for Biomimetic Spinning of Fibers with Toughness Equal to Dragline Silks. *Adv. Funct. Mater.* **2022**, 2200986.
(43) Lin, Z.; Huang, W.; Zhang, J.; Fan, J.-S.; Yang, D. Solution structure of eggcase silk protein and its implications for silk fiber formation. *Proc. Nat. Acad. Sci.* **2009**, *106* (22), 8906-8911.

(44) Xu, L.; Rainey, J. K.; Meng, Q.; Liu, X. Q. Recombinant minimalist spider wrapping silk proteins capable of native-like fiber formation. *PLoS One* **2012**, *7* (11), e50227. DOI: 10.1371/journal.pone.0050227.
(45) Garnier, J.; Osguthorpe, D. J.; Robson, B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* **1978**, *120* (1), 97-120. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet.* **2000**, *16* (6), 276-277.
(46) Xu, L.; Lefevre, T.; Orrell, K. E.; Meng, Q.; Auger, M.; Liu, X. Q.; Rainey, J. K. Structural and mechanical roles for the C-terminal nonrepetitive domain become apparent in recombinant spider aciniform silk. *Biomacromolecules* **2017**, *18* (11), 3678-3686. DOI: 10.1021/acs.biomac.7b01057.
(47) Xu, L.; Weatherbee-Martin, N.; Liu, X. Q.; Rainey, J. K. Recombinant silk fiber properties correlate to prefibrillar self-assembly. *Small* **2019**, 1805294. DOI: 10.1002/smll.201805294.
(48) Simmons, J. R.; Xu, L.; Rainey, J. K. Recombinant Pyriform Silk Fiber Mechanics Are Modulated by Wet-spinning Conditions. *ACS Biomater. Sci. Eng. & Engineering* **2019**.
(49) Shevchenko, A.; Tomas, H.; Havli, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1* (6), 2856-2860.
(50) Koeppel, A.; Holland, C. Progress and trends in artificial silk spinning: a systematic review. *ACS Biomater. Sci. Eng.* **2017**, *3* (3), 226-237. DOI: 10.1021/acsbiomaterials.6b00669.
(51) Weatherbee-Martin, N.; Xu, L.; Hupe, A.; Kreplak, L.; Fudge, D. S.; Liu, X. Q.; Rainey, J. K. Identification of wet-spinning and post-spin stretching methods amenable to recombinant spider aciniform silk. *Biomacromolecules* **2016**, *17* (8), 2737-2746. DOI: 10.1021/acs.biomac.6b00857.
(52) Kuipers, B. J. H., Gruppen, H. Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography−mass spectrometry analysis. *J. Agric. Food Chem.* **2007**, *55*, 5445−5451. DOI: 10.1021/jf070337l.
(53) Zhou, G.; Shao, Z.; Knight, D. P.; Yan, J.; Chen, X. Silk fibers extruded artificially from aqueous solutions of regenerated Bombyx mori silk fibroin are tougher than their natural counterparts. *Adv. Mater.* **2009**, *21* (3), 366-370. DOI: 10.1002/adma.200800582.

(54) Gullekson, C.; Lucas, L.; Hewitt, K.; Kreplak, L. Surface-sensitive Raman spectroscopy of collagen I fibrils. *Biophys. J.* **2011**, *100* (7), 1837-1845. DOI: 10.1016/j.bpj.2011.02.026.
(55) Schnaitman, C. A. Solubilization of the Cytoplasmic Membrane of Escherichia coli by Triton X-100. *J. Bacteriol* **1971**, *108* (1), 545-552. Bordier, C. Phase Separation of Integral Membrane Proteins in Triton X-114 Solution. *J. Biol. Chem.* **1981**, *256* (4), 1604-1607. Stumpe, M. C.; Grubmüller, H. Interaction of urea with amino acids: implications for urea-induced protein denaturation. *J. Am. Chem. Soc.* **2007**, *129* (51), 16126-16131.
(56) Hames, B. D. Gel Electrophoresis of Proteins A Practical Approach, Third Edition. *Oxford University Press Inc., New York* **1998**, 33.
(57) Greenfield, N. J. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc.* **2006**, *1* (6), 2527.
(58) Greenfield, N. J. Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal. Biochem.* **1996**, *235* (1), 1-10.
(59) Boutry, C.; Blackledge, T. A. Evolution of supercontraction in spider silk: structure-function relationship from tarantulas to orb-weavers. *J. Exp. Biol.* **2010**, *213* (Pt 20), 3505-3514. DOI: 10.1242/jeb.046110. Blackledge, T. A.; Boutry, C.; Wong, S. C.; Baji, A.; Dhinojwala, A.; Sahni, V.; Agnarsson, I. How super is supercontraction? Persistent versus cyclic responses to humidity in spider dragline silk. *J. Exp. Biol.* **2009**, *212* (Pt 13), 1981-1989. DOI: 10.1242/jeb.028944.
(60) Gnesa, E.; Hsia, Y.; Yarger, J. L.; Weber, W.; Lin-Cereghino, J.; Lin-Cereghino, G.; Tang, S.; Agari, K.; Vierra, C. Conserved C-terminal domain of spider tubuliform spidroin 1 contributes to extensibility in synthetic fibers. *Biomacromolecules* **2012**, *13* (2), 304-312. DOI: 10.1021/bm201262n.
(61) Xia, X.-X.; Qian, Z.-G.; Ki, C. S.; Park, Y. H.; Kaplan, D. L.; Lee, S. Y. Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. *Proc. Natl. Acad. Sci.* **2010**, *107* (32), 14059-14063.
(62) Kronqvist, N.; Sarr, M.; Lindqvist, A.; Nordling, K.; Otikovs, M.; Venturi, L.; Pioselli, B.; Purhonen, P.; Landreh, M.; Biverstål, H. Efficient protein production inspired by how spiders make silk. *Nat. Commun.* **2017**, *8* (1), 1-15.

(63) Krishnarjuna, B.; Jaipuria, G.; Thakur, A.; D'Silva, P.; Atreya, H. S. Amino acid selective unlabeling for sequence specific resonance assignments in proteins. *J. Biomol. NMR* **2011**, *49* (1), 39-51.
(64) Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **2005**, *41* (1), 207-234.
(65) Kumar, J.; Chauhan, A. S.; Shah, R. L.; Gupta, J. A.; Rathore, A. S. Amino acid supplementation for enhancing recombinant protein production in E. coli. *Biotechnol. Bioeng.* **2020**.
(66) Chow, D. C.; Dreher, M. R.; Trabbic-Carlson, K.; Chilkoti, A. Ultra-high expression of a thermally responsive recombinant fusion protein in E. coli. *Biotechnol. Prog.* **2006**, *22* (3), 638-646.
(67) Langelaan, D. N.; Reddy, T.; Banks, A. W.; Dellaire, G.; Dupré, D. J.; Rainey, J. K. Structural features of the apelin receptor N-terminal tail and first transmembrane segment implicated in ligand binding and receptor trafficking. *Biochim. Biophys. Acta. Biomembr.* **2013**, *1828* (6), 1471-1483.
(68) Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **1995**, *6* (2), 135-140.
(69) Azatian, S. B.; Kaur, N.; Latham, M. P. Increasing the buffering capacity of minimal media leads to higher protein yield. *J. Biomol. NMR* **2019**, *73* (1), 11-17.
(70) Palmer III, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Mag. Reson. (1969)* **1991**, *93* (1), 151-170. Kay, L.; Keifer, P.; Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **1992**, *114* (26), 10663-10665. Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sørensen, O. W.; Griesinger, C. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J. Biomol. NMR* **1994**, *4* (2), 301-306.
(71) Xu, L.; Tremblay, M.-L.; Meng, Q.; Liu, X.-Q.; Rainey, J. K. ^{1}H, ^{13}C and ^{15}N NMR assignments of the aciniform spidroin (AcSp1) repetitive domain of Argiope trifasciata wrapping silk. *Biomol. NMR Assign.* **2012**, *6* (2), 147-151

(72) Kay, L. E.; Xu, G.-Y.; Singer, A. U.; Muhandiram, D. R.; Formankay, J. D. A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlations in H_2O samples of proteins. *J. Mag. Reson.* **1993**, *101* (3), 333-337.
(73) Davis, A. L.; Keeler, J.; Laue, E. D.; Moskau, D. Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients. *J. Mag. Reson. (1969)* **1992**, *98* (1), 207-216.
(74) Cheung, M.-S.; Maguire, M. L.; Stevens, T. J.; Broadhurst, R. W. DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *J. Mag. Reson.* **2010**, *202* (2), 223-233.
(75) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **1995**, *6* (3), 277-293.
(76) Maciejewski, M. W.; Schuyler, A. D.; Gryk, M. R.; Moraru, I. I.; Romero, P. R.; Ulrich, E. L.; Eghbalnia, H. R.; Livny, M.; Delaglio, F.; Hoch, J. C. NMRbox: a resource for biomolecular NMR computation. *Biophys. J.* **2017**, *112* (8), 1529-1534.
(77) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins:Struct. Funct. Genet.* **2005**, *59* (4), 687-696.
(78) Hoffman, R. E. Standardization of chemical shifts of TMS and solvent signals in NMR solvents. *Mag. Reson. Chem.* **2006**, *44* (6), 606-616.
(79) Stern, A. S.; Donoho, D. L.; Hoch, J. C. NMR data processing using iterative thresholding and minimum l1-norm reconstruction. *J. Mag. Reson.* **2007**, *188* (2), 295-300. Hyberts, S. G.; Milbradt, A. G.; Wagner, A. B.; Arthanari, H.; Wagner, G. Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. *J. Biomol. NMR* **2012**, *52* (4), 315-327.
(80) Ying, J.; Delaglio, F.; Torchia, D. A.; Bax, A. Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J. Biomol. NMR* **2017**, *68* (2), 101-118.
(81) Wishart, D. S.; Sykes, B. D.; Richards, F. M. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992**, *31* (6), 1647-1651.

(82) Wishart, D. S.; Sykes, B. D. The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR* **1994**, *4* (2), 171-180.
(83) Wang, B.; Wang, Y.; Wishart, D. S. A probabilistic approach for validating protein NMR chemical shift assignments. *J. Biomol. NMR* **2010**, *47* (2), 85-99.
(84) Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J. Biomol. NMR* **1995**, *5* (1), 67-81.
(85) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **1994**, *33* (19), 5984-6003.
(86) Lam, S. L.; Hsu, V. L. NMR identification of left-handed polyproline type II helices. *Biopolymers: Original Research on Biomolecules* **2003**, *69* (2), 270-281.
(87) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **1998**, *54* (5), 905-921.
(88) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **2003**, *160* (1), 65-73.
(89) Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M. ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **2007**, *23* (3), 381-382.
(90) Güntert, P. Automated NMR structure calculation with CYANA. In *Protein NMR Techniques*, Springer, 2004; pp 353-378.
(91) Nilges, M.; Macias, M. J.; O'Donoghue, S. I.; Oschkinat, H. Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. Mol. Biol.* **1997**, *269* (3), 408-422.
(92) Honegger, P.; Di Pietro, M. E.; Castiglione, F.; Vaccarini, C.; Quant, A.; Steinhauser, O.; Schröder, C.; Mele, A. The Intermolecular NOE Depends on Isotope Selection: Short Range vs Long Range Behavior. *J. Phys. Chem.* **2021**, *12* (35), 8658-8663.

(93) Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M. Refinement of protein structures in explicit solvent. *Proteins: Struct. Funct. Genet.* **2003**, *50* (3), 496-506.
(94) Nederveen, A. J.; Doreleijers, J. F.; Vranken, W.; Miller, Z.; Spronk, C. A.; Nabuurs, S. B.; Güntert, P.; Livny, M.; Markley, J. L.; Nilges, M. RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins: Struct. Funct. Genet.* **2005**, *59* (4), 662-672.
(95) Lefèvre, T.; Paquet-Mercier, F.; Rioux-Dubé, J. F.; Pézolet, M. Structure of silk by Raman spectromicroscopy: From the spinning glands to the fibers. *Biopolymers* **2012**, *97* (6).
(96) Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Rance, M.; Skelton, N. J. *Protein NMR spectroscopy: principles and practice. 2nd ed*; Academic Press, New York, 2007.
(97) Andrew, E. R. Magic angle spinning in solid state nmr spectroscopy. *Philos. Trans. Royal Soc. A PHILOS T R SOC A* **1981**, *299* (1452), 505-520.
(98) Kolodziejski, W.; Klinowski, J. Kinetics of cross-polarization in solid-state NMR: a guide for chemists. *Chem. Rev.* **2002**, *102* (3), 613-628.
(99) Thankamony, A. S. L.; Wittmann, J. J.; Kaushik, M.; Corzilius, B. Dynamic nuclear polarization for sensitivity enhancement in modern solid-state NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *102*, 120-195.
(100) Hartmann, S.; Hahn, E. Nuclear double resonance in the rotating frame. *Phys. Rev.* **1962**, *128* (5), 2042.
(101) Asakura, T.; Yao, J.; Yamane, T.; Umemura, K.; Ulrich, A. S. Heterogeneous structure of silk fibers from Bombyx M Ori Resolved by ^{13}C Solid-State NMR spectroscopy. *J. Am. Chem. Soc.* **2002**, *124* (30), 8794-8795.
(102) Takegoshi, K.; Nakamura, S.; Terao, T. ^{13}C–^{1}H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem Phys. Lett.* **2001**, *344* (5-6), 631-637.
(103) Hooper, R. W.; Klein, B. A.; Michaelis, V. K. Dynamic Nuclear Polarization (DNP) 101: A New Era for Materials. *Chem. Mater.* **2020**, *32* (11), 4425-4430.
(104) Jähnig, F.; Kwiatkowski, G.; Ernst, M. Conceptual and instrumental progress in dissolution DNP. *J. Mag. Reson.* **2016**, *264*, 22-29.
(105) Sauvée, C.; Rosay, M.; Casano, G.; Aussenac, F.; Weber, R. T.; Ouari, O.; Tordo, P. Highly efficient, water-soluble polarizing agents for dynamic nuclear polarization at high frequency. *Angew. Chem.* **2013**, *125* (41), 11058-11061.

(106) Mentink-Vigier, F.; Marin-Montesinos, I.; Jagtap, A. P.; Halbritter, T.; van Tol, J.; Hediger, S.; Lee, D.; Sigurdsson, S. T.; De Paëpe, G. l. Computationally assisted design of polarizing agents for dynamic nuclear polarization enhanced NMR: the AsymPol family. *J. Am. Chem. Soc.* **2018**, *140* (35), 11013-11019.
(107) Zhao, L.; Pinon, A. C.; Emsley, L.; Rossini, A. J. DNP-enhanced solid-state NMR spectroscopy of active pharmaceutical ingredients. *Mag. Reson. Chem.* **2018**, *56* (7), 583-609.
(108) Yan, P.; Millard, A. C.; Wei, M.; Loew, L. M. Unique contrast patterns from resonance-enhanced chiral SHG of cell membranes. *J. Am. Chem. Soc.* **2006**, *128* (34), 11030-11031.
(109) Van Gulick, L.; Saby, C.; Morjani, H.; Beljebbar, A. Age-related changes in molecular organization of type I collagen in tendon as probed by polarized SHG and Raman microspectroscopy. *Sci. Rep.* **2019**, *9* (1), 1-12.
(110) Iachina, I.; Brewer, J. R. Strain-Dependent Structural Changes in Major and Minor Ampullate Spider Silk Revealed by Two-Photon Excitation Polarization. *Biomacromolecules* **2019**, *20* (6), 2384-2391.
(111) Zhao, Y.; Hien, K. T. T.; Mizutani, G.; Rutt, H. N. Second-order nonlinear optical microscopy of spider silk. *Appl; Phys. B* **2017**, *123* (6), 1-8.
(112) Zhao, Y.; Li, Y.; Hien, K. T. T.; Mizutani, G.; Rutt, H. N. Observation of spider silk by femtosecond pulse laser second harmonic generation microscopy. *Surf. Interface Anal.* **2019**, *51* (1), 56-60.
(113) Rice, W. L.; Firdous, S.; Gupta, S.; Hunter, M.; Foo, C. W.; Wang, Y.; Kim, H. J.; Kaplan, D. L.; Georgakoudi, I. Non-invasive characterization of structure and morphology of silk fibroin biomaterials using non-linear microscopy. *Biomaterials* **2008**, *29* (13), 2015-2024.
(114) Cheng, J.-X.; Xie, X. S. Green's function formulation for third-harmonic generation microscopy. *JOSA B* **2002**, *19* (7), 1604-1610.
(115) Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M. ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **2007**, *23* (3), 381-382.
(116) Bhattacharya, A.; Tejero, R.; Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins: Struct. Funct. Genet.* **2007**, *66* (4), 778-795.

(117) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules* **1983**, *22* (12), 2577-2637.
(118) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G.; Eletsky, A.; Wu, Y.; Singarapu, K. K.; Lemak, A. Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci.* **2008**, *105* (12), 4685-4690.
(119) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583-589.
(120) Doyle, C. M.; Rumfeldt, J. A.; Broom, H. R.; Sekhar, A.; Kay, L. E.; Meiering, E. M. Concurrent increases and decreases in local stability and conformational heterogeneity in Cu, Zn superoxide dismutase variants revealed by temperature-dependence of amide chemical shifts. *Biochemistry* **2016**, *55* (9), 1346-1361.
(121) Trainor, K.; Palumbo, J. A.; MacKenzie, D. W.; Meiering, E. M. Temperature dependence of NMR chemical shifts: Tracking and statistical analysis. *Protein Sci.* **2020**, *29* (1), 306-314.
(122) Schumann, F. H.; Riepl, H.; Maurer, T.; Gronwald, W.; Neidig, K.-P.; Kalbitzer, H. R. Combined chemical shift changes and amino acid specific chemical shift mapping of protein–protein interactions. *J. Biomol. NMR* **2007**, *39* (4), 275-289.
(123) Gor'kov, P. L.; Brey, W. W.; Long, J. R. Probe development for biosolids NMR spectroscopy. *Solid-State NMR Studies of Biopolymers* **2010**, 141.
(124) Fritzsching, K. J.; Hong, M.; Schmidt-Rohr, K. Conformationally selective multidimensional chemical shift ranges in proteins from a PACSY database purged using intrinsic quality criteria. *J. Biomol. NMR* **2016**, *64* (2), 115-130.
(125) Brittain, K.; Harvey, M.; Cisek, R.; Pillai, S.; Christie, S. D.; Tokarz, D. Second harmonic generation microscopy of otoconia. *Biomed. Opt. Express* **2022**, *13* (6), 3593-3600. Purvis, K.; Brittain, K.; Joseph, A.; Cisek, R.; Tokarz, D. Third-order nonlinear optical properties of phycobiliproteins from cyanobacteria and red algae. *Chem. Phys. Lett.* **2019**, *731*, 136599.
(126) Tejero, R.; Snyder, D.; Mao, B.; Aramini, J. M.; Montelione, G. T. PDBStat: a universal restraint converter and restraint analysis software package for protein NMR. *J. Biomol. NMR* **2013**, *56* (4), 337-351.

(127) Lüthy, R.; Bowie, J. U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* **1992**, *356* (6364), 83-85.
(128) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **1993**, *26* (2), 283-291.
(129) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallog. D* **2010**, *66* (1), 12-21.
(130) Huang, Y. J.; Rosato, A.; Singh, G.; Montelione, G. T. RPF: a quality assessment tool for protein NMR structures. *Nucleic Acids Res.* **2012**, *40* (W1), W542-W546.
(131) An, B.; Hinman, M. B.; Holland, G. P.; Yarger, J. L.; Lewis, R. V. Inducing β-sheets formation in synthetic spider silk fibers by aqueous post-spin stretching. *Biomacromolecules* **2011**, *12* (6), 2375-2381.
(132) Platzer, G.; Okon, M.; McIntosh, L. P. pH-dependent random coil ^{1}H, ^{13}C, and ^{15}N chemical shifts of the ionizable amino acids: a guide for protein pKa measurements. *J. Biomol. NMR* **2014**, *60* (2), 109-129.
133) Vollrath, F.; Holtet, T.; Thøgersen, H. C.; Frische, S. Structural organization of spider silk. *Proc. R. Soc. B: Biol. Sci.* **1996**, *263* (1367), 147-151. Nguyen, A. T.; Huang, Q. L.; Yang, Z.; Lin, N.; Xu, G.; Liu, X. Y. Crystal networks in silk fibrous materials: from hierarchical structure to ultra performance. *Small* **2015**, *11* (9-10), 1039-1054.
(134) Wüthrich, K. NMR of Proteins and Nucleic Acids. John Wiley & Sons, Inc.: 1991.
(135) Riekel, C.; Burghammer, M.; Rosenthal, M. Skin-core morphology in spider flagelliform silk. *Appl. Phys. Lett.* **2019**, *115* (12), 123702. Frische, S.; Maunsbach, A.; Vollrath, F. Elongate cavities and skin-core structure in Nephila spider silk observed by electron microscopy. *J. Microsc.* **1998**, *189* (1), 64-70.
(136) Grzesiek, S.; Bax, A. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. Mag. Reson. (1969)* **1992**, *96* (2), 432-440.
(137) Schleucher, J.; Sattler, M.; Griesinger, C. Coherence selection by gradients without signal attenuation: application to the three-dimensional HNCO experiment. *Angew. Chem. Int. Ed. Engl.* **1993**, *32* (10), 1489-1491.

(138) Kay, L. E.; Xu, G. Y.; Yamazaki, T. Enhanced-sensitivity triple-resonance spectroscopy with minimal H2O saturation. *J. Mag. Reson., Series A* **1994**, *109* (1), 129-133.
(139) Clubb, R. T.; Thanabal, V.; Wagner, G. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue ^{1}HN, ^{15}N, and ^{13}C'chemical shifts in ^{15}N-^{13}C-labelled proteins. **1992**.
(140) Muhandiram, D.; Kay, L. E. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Mag. Reson., Series B* **1994**, *103* (3), 203-216.
(141) Wittekind, M.; Mueller, L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-and beta-carbon resonances in proteins. *J. Mag. Reson.. Series B (Print)* **1993**, *101* (2), 201-205.
(142) Bermel, W.; Bertini, I.; Duma, L.; Felli, I. C.; Emsley, L.; Pierattelli, R.; Vasos, P. R. Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. *Angew. Chem. Int. Ed.* **2005**, *44* (20), 3089-3092.
(143) Duma, L.; Hediger, S.; Lesage, A.; Emsley, L. Spin-state selection in solid-state NMR. *J. Mag. Reson.* **2003**, *164* (1), 187-195.
(144) Bermel, W.; Bertini, I.; Felli, I. C.; Pierattelli, R. Speeding up ^{13}C direct detection biomolecular NMR spectroscopy. *J. Am. Chem. Soc.* **2009**, *131* (42), 15339-15345.
(145) Bermel, W.; Bertini, I.; Csizmok, V.; Felli, I. C.; Pierattelli, R.; Tompa, P. H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. *J. Mag. Reson.* **2009**, *198* (2), 275-281.
(146) Bertini, I.; Duma, L.; Felli, I. C.; Fey, M.; Luchinat, C.; Pierattelli, R.; Vasos, P. R. A heteronuclear direct-detection NMR spectroscopy experiment for protein-backbone assignment. *Angew. Chem. Int. Ed.* **2004**, *43* (17), 2257-2259.
(147) Bermel, W.; Bertini, I.; Felli, I. C.; Kümmerle, R.; Pierattelli, R. Novel 13C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. *J. Mag. Reson.* **2006**, *178* (1), 56-64.
(148) Mori, S.; Abeygunawardana, C.; Johnson, M. O. n.; Vanzijl, P. C. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J. Mag. Reson., Series B* **1995**, *108* (1), 94-98.

Appendices

Methods A1. Sample preparation for TEM

Fibres spun as previously described were brought to Mary Ann Travers for sample fixation and preparation. Samples were fixed for a minimum of 2 h with 2.5% Glutaraldehyde diluted with 0.1 M sodium cacodylate buffer followed by three rinses of 0.1 M sodium cacodylate buffer (10 min minimum each). This was followed by 1% osmium tetroxide fixation for 2h, immediately rinsing with distilled water afterward, and placed in 0.25% uranyl acetate overnight at 4°C. After overnight uranyl acetate exposure, samples were gradually dehydrated with acetone (50%, 10 min; 70%, 10 min x 2; 95%, 10 min x 2; 100%, 10 min x 2; dried 100%, 10 min). Samples were infiltrated with Epon Araldite Resin (3 parts dried 100% acetone:1 part resin, 3h; 1 part dried 100% acetone:3 parts resin overnight; 100% resin for 3 h x2) then embedded 100% Epon Araldite Resin and cured at 60 °C for 48 h. Thin sections were cut using a Reichert – Jung Ultracut E Ultramicrotome with a diamond knife (approximately 100 nm thick) and placed on 300 mesh copper grids. Samples were stained (2% aqueous uranyl acetate, 10 minutes; 5 minutes distilled water rinse x2; lead citrate, 4 minutes; rinse with distilled water; air dry), viewed using a JEOL JEM 1230 Transmission Electron Microscope at 80k, and images were captured using a Hamamatsu ORCA-HR digital camera.

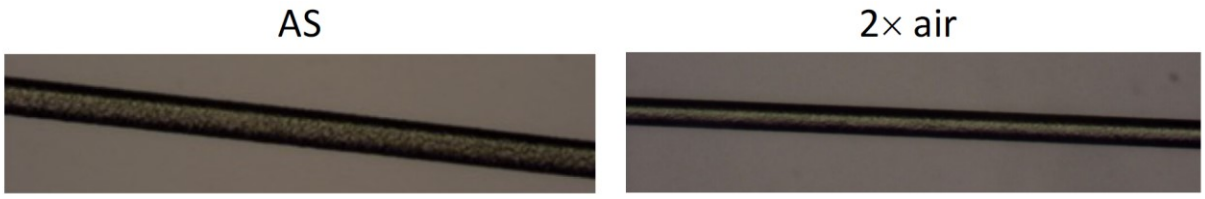


Figure A1. Microscopy images of AS and 2× air stretched fibres.

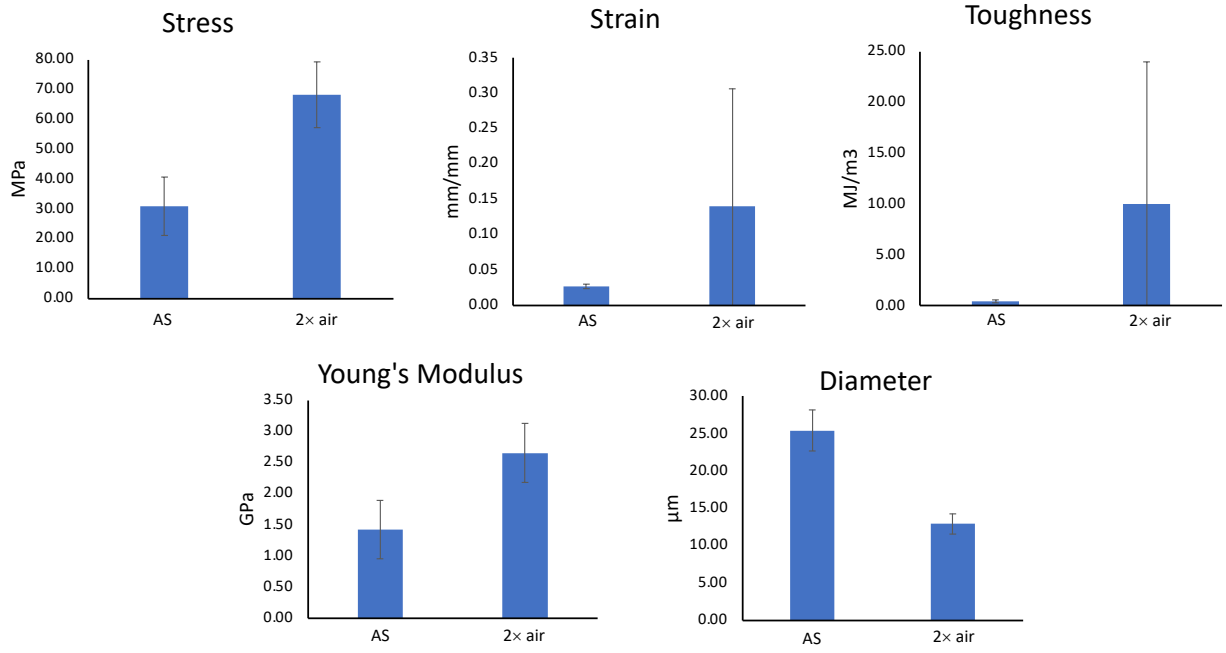


Figure A2. Mechanical testing results of Ni^{2+}-IMAC purified HPy_2 spun into fibres at 8% w/v.

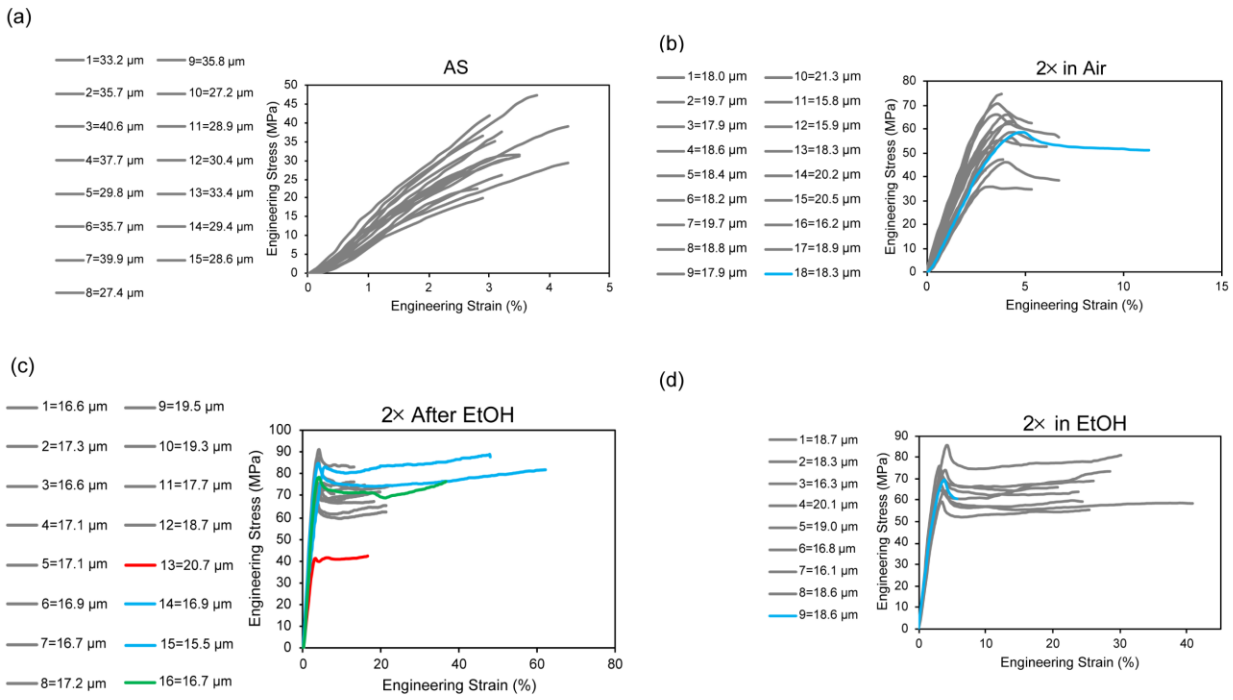


Figure A3. Overlay of data used for averages vs. removed as outliers. Overlay of data used for averages (grey) with data not averaged for strength (red), extensibility (blue), and toughness (green). Reprinted by permission from ACS Biomaterials Science & Engineering[48] .

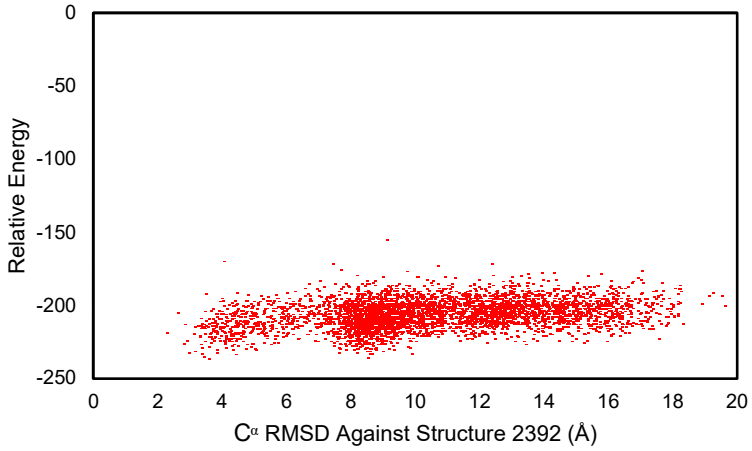


Figure A4. CA RMSD values relative to structure 2393 plotted against relative energies as calculated by CS-ROSETTA for all 5000 structures.

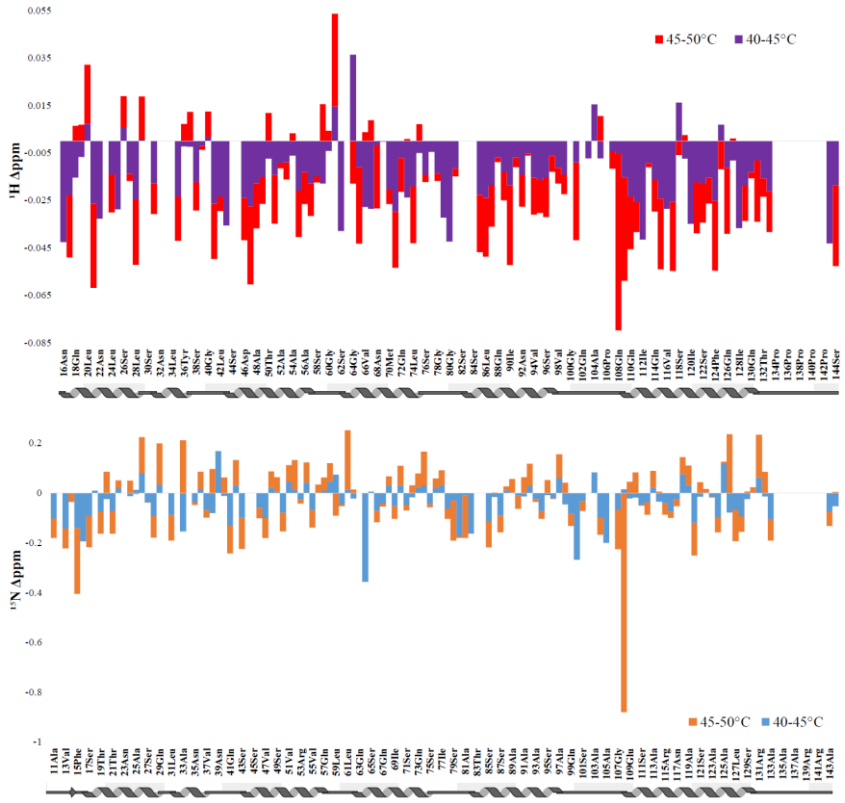


Figure A5. ^{1}H (top) and ^{15}N (bottom) Δppm values showing residue-specific perturbation based on temperature.

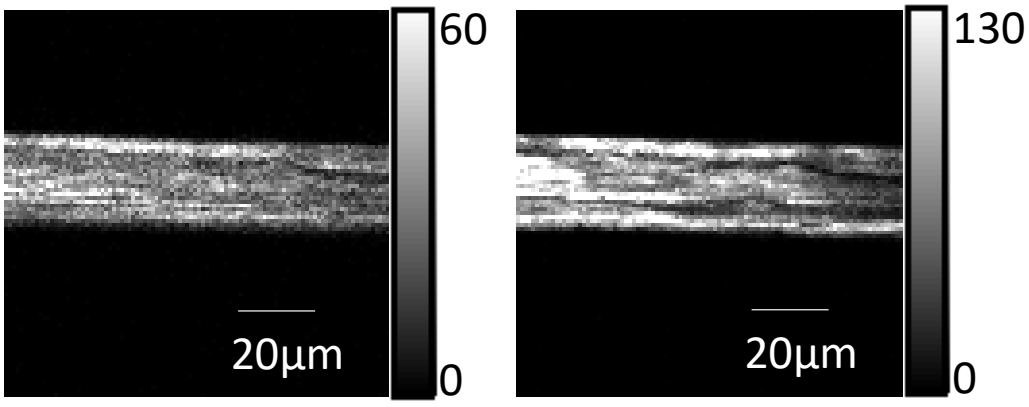


Figure A6. SHG (left) and THG (right) of HPy_2 fibres post-spin stretched 2x in EtOH.

Table A1. Mechanical properties of fibres spun from 15% and 25% w/v M9-HPy$_2$ dopes, and initially collected sets of fibres with consistent higher diameter

% HPy2	% EtOH Coagulation Bath	Post-spinning condition	Temp. (°C)	% Humidity	Diameter (µm)	Stress at failure (MPa)	Ultimate Tensile Strength (MPa)	% Strain	Toughness (MJ/m3)	Young's Modulus (GPa)	Number of Fibres (N)
15%	95%	2× in air	27.8	25.8	18.9 ± 2.6	25.6 ± 11.1	29.2 ± 11.7	81 ± 29	20.8 ± 12.9	1.04 ± 0.34	8
15%	95%	as spun	27.8	25.7	28.7 ±4.6	12.5 ± 5.2	17.5 ± 5.9	79 ± 101	7.8 ± 9.8	0.56 ± 0.23	10
20%	95%	2× in air	27.9	27.3	29.7 ± 1.9	22.7 ± 3.0	23.3 ± 3.4	129 ± 31	24.8 ± 7.5	0.84 ± 0.15	16
20%	95%	2× in 40% EtOH	~26	~25	30.4 ± 3.5	27.7 ± 6.8	34.1 ± 8.1	88 ± 84	22.5 ± 22.6	1.18 ± 0.12	8
20%	100%	2× in air	26	25.2	25.4 ± 3.1	29.6 ± 9.2	29.8 ± 9.0	206 ± 57	43.2 ± 11.7	0.82 ± 0.19	17
25%	90%	2× in air	Room Temp (25-30)	25-30%	30.0 ± 8.2	27.7 ±11.9	29.03 ± 11.9	143 ± 43	33.8 ± 18.4	0.97 ± 0.42	20
25%	95%	2× in air	Room Temp (25-30)	25-30%	26.9 ± 5.6	25.7 ±9.9	30.8 ± 11.0	119 ± 71	28.3 ± 19.5	1.14 ± 0.49	19

Table A2. NMR spectra acquisition conditions for HPy_1. Spectra used for backbone assignment are shaded in grey, while ^{13}C-rooted spectra are shaded in pink.

Experiment	Protein Concentration (μM)	Pulse Program (Bruker)	Delay (s)	# of Scans	Fid Size	Sweep Width (ppm)	NUS	^{1}H Frequency (MHz)
^{1}H-^{15}N HSQC	Prepared at ~950 μM, but exact concentration unknown due to precipitation	hsqcetf3gpsi2	1	2	^{1}H: 14.0 ^{15}N: 26.0	^{1}H: 16.1 ^{15}N: 32.0	N/A	700
HNCO[136-138]	~325	hncogp3d	2	8	^{1}H: 2048 ^{15}N: 80 ^{13}C: 128	^{1}H: 14.0 ^{15}N: 26.0 ^{13}C: 12.0	10%; 256 points	700
HN(CA)CO[138, 139]	~325	hncacogp3d	2	16	^{1}H: 2048 ^{15}N: 80 ^{13}C: 128	^{1}H:14.0 ^{15}N: 26.0 ^{13}C: 12.0	18.5%; 473 points	700
HNCA[136-138]	~325	hncagp3d	2	16	^{1}H: 2048 ^{15}N: 80 ^{13}C: 128	^{1}H: 16.1 ^{15}N: 26.0 ^{13}C: 25.0	18.5%; 473 points	700

Table A2 Continued

Experiment	Protein Concentration (μM)	Pulse Program (Bruker)	Delay (s)	# of Scans	Fid Size	Sweep Width (ppm)	NUS	^{1}H Frequency (MHz)
CBCA(CO)NH[136, 140]	~325	cbcaconhgp3d	2	16	^{1}H: 2048 ^{15}N: 64 ^{13}C: 128	^{1}H: 14.0 ^{15}N: 26.0 ^{13}C: 75.0	23%; 471 points	700
HNCACB[140, 141]	~325	hncacbgp3d	1.95	16	^{1}H: 2048 ^{15}N: 64 ^{13}C: 128	^{1}H: 14.0 ^{15}N: 26.0 ^{13}C: 75.0	45%; 921 points	700
^{1}H-^{15}N HSQC NOESY (95 ms mixing time)	Prepared at ~950 μM, but exact concentration unknown due to precipitation	noesyhsqcetf3gp3d	1.5	8	^{1}H: 2048 ^{15}N: 40 ^{1}H: 96	^{1}H: 16.1 ^{15}N: 25.0 ^{1}H: 12.0	N/A	700
CON[142, 143]	~500	c_con_iasq	1	8	^{13}C: 1024 ^{15}N: 128	^{13}C: 20.2 ^{15}N: 40.0	N/A	800
(H)CACO[143-145]	~500	c_hcaco_ctiare	1	16	^{13}C: 1024 ^{13}C: 512	^{13}C: 20.2 ^{13}C: 50.0	N/A	800

Table A2 Continued

Experiment	Protein Concentration (μM)	Pulse Program (Bruker)	Delay (s)	# of Scans	Fid Size	Sweep Width (ppm)	NUS	^{1}H Frequency (MHz)
(H)CANCO[142, 143, 146]	~500	c_hcanco_ia3d	1	8	^{13}C: 1024 ^{15}N: 64 ^{13}C: 64	^{13}C: 40.7 ^{15}N: 47.0 ^{13}C: 40.7	N/A	800
(H)CBCANCO[145, 147]	~500	c_hcbcanco_ia3d	1	8	^{13}C: 1024 ^{15}N: 64 ^{13}C: 64	^{13}C: 40.7 ^{15}N: 47.0 ^{13}C: 80.0	N/A	800
(H)CBCACON[142, 143, 145, 147]	~500	c_hcbcacon_ia3d	1	8	^{13}C: 1024 ^{15}N: 48 ^{13}C: 102	^{13}C: 40.7 ^{15}N: 47.0 ^{13}C: 80.0	N/A	800
HNCO[142, 143]	~500	c_hnco_ia3d	1	8	^{13}C: 1024 ^{15}N: 64 ^{1}H: 90	^{13}C: 40.7 ^{15}N: 47.0 ^{1}H: 10.0	N/A	800
(H)CACON[143, 145]	~500	c_hcacon_ia3d	1	8	^{13}C: 1024 ^{15}N: 64 ^{13}C: 64	^{13}C: 40.7 ^{15}N: 47.0 ^{13}C: 40.7	N/A	800

Table A3. NMR spectra acquisition conditions for Δ9-70, Δ206-242-HPy$_1$. Spectra used for backbone assignment are shaded in grey, while spectra used for the side-chain assignment are shaded in orange.

Experiment	Protein Concentration (μM)	Pulse Program (Bruker)	Delay (s)	Number of Scans	Number of Data Points	Sweep Width (ppm)	NUS	^{1}H Frequency (MHz)
^{1}H-^{15}N HSQC[148]	~450	fhsqcf3gpph	1	2	^{1}H: 2048 ^{15}N: 200	^{1}H: 16.0 ^{15}N: 26.0	N/A	800
^{1}H-^{13}C HSQC	~450	hsqcetgpsp.2	1	4	^{1}H: 2048 ^{13}C: 128	^{1}H: 16.0 ^{13}C: 31.4	N/A	800
HNCO	~450	hncogpwg3d	1	2	^{1}H: 2048 ^{15}N: 64 ^{13}C: 32	^{1}H: 16.0 ^{15}N: 26.0 ^{13}C: 10.0	34% 174 Points	800
HN(CA)CO	~450	hncacogpwg3d	1	8	^{1}H: 2048 ^{15}N: 64 ^{13}C: 40	^{1}H: 16.0 ^{15}N: 26.0 ^{13}C: 10.0	50% 320 Points	800
CBCA(CO)NH	~450	cbcaconhgpwg3d	1	4	^{1}H: 2048 ^{15}N: 64 ^{13}C: 190	^{1}H: 16.0 ^{15}N: 26.0 ^{13}C: 70.0	25% 760 Points	800

Table A3. Continued

Experiment	Protein Concentration (μM)	Pulse Program (Bruker)	Delay (s)	Number of Scans	Number of Data Points	Sweep Width (ppm)	NUS	^{1}H Frequency (MHz)
HNCACB	~450	hncacbgpwg3d	1	4	^{1}H: 2048 ^{15}N: 64 ^{13}C: 190	^{1}H: 16.0 ^{15}N: 26.0 ^{13}C: 70.0	25%; 760 points	800
^{1}H-^{15}N HSQC NOESY (85 ms mixing time)	~450	noesyhsqcf3gpwg3d	1	8	^{1}H: 2048 ^{15}N: 64 ^{1}H: 256	^{1}H: 16.0 ^{15}N: 26.0 ^{1}H: 12.0	N/A	800
^{1}H-^{13}C HSQC NOESY (85 ms mixing time)	~450	noesyhsqcetgp3d	1	8	^{1}H: 2048 ^{13}C: 84 ^{1}H: 256	^{1}H: 16.0 ^{13}C: 31.4 ^{1}H: 12.0	N/A	800
HCCH-COSY	~450	hcchcogp3d	1	4	^{1}H: 2048 ^{13}C: 80 ^{1}H: 140	^{1}H: 16.0 ^{13}C: 31.4 ^{1}H: 6.5	N/A	800
HCCH-TOCSY	~450	hcchdigp3d	1	4	^{1}H: 2048 ^{13}C: 84 ^{1}H: 150	^{1}H: 16.0 ^{13}C: 31.4 ^{1}H: 6.5	N/A	800

Table A4. Tabulated backbone chemical shift assignments for Δ9-70+Δ206-242-HPy$_1$.

Resonance	Chemical Shift (ppm)	SD	Peaks
10SerC	173.65976	0	1
10SerCa	58.30652	0	1
10SerCb	64.02723	0.03692	2
10SerH	8.3468	0.01179	7
10SerN	118.12045	0.06121	5
11AlaC	175.31839	0.01219	2
11AlaCa	50.63541	0.00536	6
11AlaCb	18.38472	0.05034	11
11AlaH	8.14043	0.00885	16
11AlaHa	4.5975	0.01454	8
11AlaHb*	1.35722	0.00924	13
11AlaN	126.90135	0.01995	14
12ProC	176.22278	0	1
12ProCa	63.08977	0	2
12ProCb	32.02722	0	1
12ProHa	4.38661	0.00324	4
13ValC	175.13467	0.00864	2
13ValCa	61.89564	0.02834	12
13ValCb	33.17615	0.06961	7
13ValCga	21.45805	0.07181	5
13ValCgb	20.3687	0.03882	16
13ValH	7.92073	0.00303	19
13ValHa	4.11577	0.00645	20
13ValHb	1.95306	0.00654	17
13ValHga*	0.84733	0.00624	12
13ValHgb*	0.8653	0.00631	25
13ValN	118.72365	0.01902	15
14TyrC	175.16236	0.01165	2
14TyrCa	56.53679	0.04498	14
14TyrCb	40.27145	0.04651	20
14TyrCd*	133.52942	0	1
14TyrCe*	118.58742	0	1
14TyrH	8.00165	0.00569	25
14TyrHa	4.83773	0.00875	22

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
14TyrHba	2.85835	0.00616	18
14TyrHbb	3.15572	0.00507	13
14TyrHd*	7.13467	0.00486	15
14TyrHe*	6.80072	1.19E-07	11
14TyrN	121.75519	3.94E-02	17
15PheC	174.17406	0.01414	2
15PheCa	58.62842	0.02231	16
15PheCb	40.26433	0.0445	31
15PheCd*	132.93773	0	1
15PheCe*	130.82975	0	1
15PheCz	129.5759	0	1
15PheH	8.83621	0.00636	34
15PheHa	4.27837	0.00769	28
15PheHba	2.83494	0.00518	18
15PheHbb	3.02073	0.00451	33
15PheHd*	7.04969	0.00506	34
15PheHe*	7.05188	0.00615	22
15PheHz	7.20932	0.0145	16
15PheN	124.10907	0.02295	23
16AsnC	174.51999	0.00152	2
16AsnCa	51.94412	0.05913	18
16AsnCb	38.24095	0.04731	21
16AsnCg	177.78583	0.00643	2
16AsnH	9.74537	0.00976	32
16AsnHa	4.7424	0.01009	25
16AsnHba	2.46159	0.00629	19
16AsnHbb	2.94127	0.00836	33
16AsnHd2a	6.79618	0.00248	7
16AsnHd2b	7.46729	0.00181	22
16AsnN	129.85531	0.03694	22
16AsnNd2	111.74476	0.06903	15
17SerC	176.14474	0.02523	2
17SerCa	61.75541	0.00906	18
17SerCb	62.86692	0.04978	18

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
17SerH	8.37748	0.01009	27
17SerHa	4.27919	0.00565	23
17SerHba	3.99108	0.01194	11
17SerHbb	4.60959	0.00961	29
17SerN	120.63536	0.03259	20
18GlnC	177.7993	8.09E-04	2
18GlnCa	59.19322	0.02069	9
18GlnCb	28.07845	0.01374	14
18GlnCd	180.50199	1.80E-02	2
18GlnCg	34.18	0.01318	7
18GlnH	8.28269	0.00441	31
18GlnHa	4.04948	0.00795	13
18GlnHba	2.14094	0.00487	16
18GlnHbb	2.14094	0.00488	2
18GlnHe2a	6.86979	0.00313	13
18GlnHe2b	7.43339	0.00471	23
18GlnHga	2.36686	0.00317	6
18GlnHgb	2.36686	0.00317	12
18GlnN	122.8241	0.04458	23
18GlnNe2	112.67269	0.02596	17
19ThrC	176.65447	0.01419	2
19ThrCa	66.25524	0.03765	16
19ThrCb	68.4802	0.04182	15
19ThrCg2	22.39716	0.03908	22
19ThrH	7.61375	0.00479	43
19ThrHa	3.89613	0.01438	23
19ThrHb	3.92069	0.00571	27
19ThrHg2*	1.09568	0.0047	40
19ThrN	116.75898	0.02216	23
20LeuC	177.13095	0.02289	2
20LeuCa	59.15925	0.04048	16
20LeuCb	41.4711	0.02264	25
20LeuCda	24.40554	0.03929	26
20LeuCdb	26.58718	0.0211	10

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
20LeuCg	27.56395	0.04818	11
20LeuH	7.08807	0.00778	44
20LeuHa	3.94626	0.00688	31
20LeuHba	1.30964	0.00728	18
20LeuHbb	2.16456	0.00616	32
20LeuHda*	1.0669	0.00736	38
20LeuHdb*	1.07275	0.0089	24
20LeuHg	1.67756	0.00552	24
20LeuN	121.43457	0.02663	27
21ThrC	175.85843	0.01659	2
21ThrCa	67.24573	0.02932	17
21ThrCb	68.60621	0.09265	10
21ThrCg2	22.32512	0.01983	23
21ThrH	8.5886	0.00592	40
21ThrHa	3.59586	0.0056	28
21ThrHb	4.19504	0.00584	18
21ThrHg2*	1.17239	0.00888	42
21ThrN	115.16333	0.01516	25
22AsnC	178.32441	0.00964	2
22AsnCa	56.21793	0.01696	11
22AsnCb	37.9659	0.06232	17
22AsnCg	176.02261	0	1
22AsnH	8.58744	0.005	36
22AsnHa	4.40129	0.00773	15
22AsnHba	2.79745	0.006	20
22AsnHbb	2.88134	0.01655	12
22AsnHd2a	6.86904	0.00141	5
22AsnHd2b	7.49176	6.52E-05	17
22AsnN	119.1949	0.03215	26
22AsnNd2	111.14313	0.02406	15
23AsnC	177.736	0.00421	2
23AsnCa	56.696	0.02042	10
23AsnCb	39.01726	4.96E-02	27
23AsnCg	175.75626	0.01262	2

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
23AsnH	8.0146	0.00887	42
23AsnHa	4.41014	0.00695	22
23AsnHba	2.64612	0.00775	20
23AsnHbb	2.78211	0.00591	29
23AsnHd2a	7.12163	0.00926	22
23AsnHd2b	7.20759	0.01336	44
23AsnN	118.63703	0.03133	25
23AsnNd2	110.45938	0.03514	29
24LeuC	177.67843	0.02197	3
24LeuCa	57.6282	0.0226	22
24LeuCb	42.67447	0.0563	25
24LeuCda	25.81319	0.01483	8
24LeuCdb	24.854	0.02087	34
24LeuCg	27.41002	0	4
24LeuH	8.42896	0.00783	48
24LeuHa	4.28725	0.00742	35
24LeuHba	1.66484	0.00789	33
24LeuHbb	1.69375	0.01422	13
24LeuHda*	0.91304	0.0117	28
24LeuHdb*	0.99354	0.00499	47
24LeuHg	1.66932	0.00723	20
24LeuN	120.42302	0.03535	27
25AlaC	179.7618	0.02127	3
25AlaCa	55.85445	0.01861	10
25AlaCb	17.8129	0.07833	17
25AlaH	8.88899	0.00755	47
25AlaHa	3.82983	0.00885	23
25AlaHb*	1.35998	0.00568	37
25AlaN	121.41534	0.03849	28
26SerC	177.21809	0.03008	3
26SerCa	61.37453	0.03634	9
26SerCb	63.00027	0.02839	11
26SerH	7.53472	0.00811	34
26SerHa	4.19903	0.0045	12

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
26SerHba	3.95917	0.00771	8
26SerHbb	3.96065	0.00825	15
26SerN	109.87252	0.01953	22
27SerC	177.08063	0.02006	3
27SerCa	61.8471	0.02709	9
27SerCb	62.94484	0.04074	14
27SerH	7.82383	0.00851	40
27SerHa	4.244	1.42E-05	10
27SerHba	3.88152	0.01253	6
27SerHbb	3.91814	4.01E-05	16
27SerN	117.57231	0.03161	26
28LeuC	177.90898	0.00662	2
28LeuCa	57.24701	0.04224	19
28LeuCb	42.3177	0.05138	25
28LeuCda	27.316	4.00E-02	26
28LeuCdb	23.89176	0.00878	9
28LeuCg	27.00241	8.73E-03	10
28LeuH	8.5241	0.00713	52
28LeuHa	4.16283	0.00789	33
28LeuHba	1.33916	0.00618	24
28LeuHbb	1.96273	0.01234	34
28LeuHda*	0.8412	0.01	42
28LeuHdb*	0.90431	0.00749	17
28LeuHg	1.99373	0.00594	22
28LeuN	122.82218	0.01534	29
29GlnC	176.80051	0.01575	2
29GlnCa	58.70616	0.04842	10
29GlnCb	28.5016	0.06983	14
29GlnCd	180.08206	0	1
29GlnCg	33.9247	0.00814	14
29GlnH	7.3767	0.00565	46
29GlnHa	3.89484	0.01549	22
29GlnHba	2.13197	0.00818	15
29GlnHbb	2.1556	0.01463	11

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
29GlnHe2a	6.73053	0.00321	9
29GlnHe2b	7.27168	0.0046	17
29GlnHga	2.40097	0.01428	12
29GlnHgb	2.53279	0.00955	20
29GlnN	115.19115	0.03315	30
29GlnNe2	110.58196	0.08333	14
30SerC	174.21901	0.0127	2
30SerCa	58.41421	0.02697	9
30SerCb	63.87399	0.00523	10
30SerH	7.35566	0.00763	32
30SerHa	4.47297	0.00264	10
30SerHba	4.00007	0.01006	11
30SerHbb	4.02485	0.01414	4
30SerN	111.37951	0.07862	22
31LeuC	177.83367	0.00924	2
31LeuCa	55.09637	0.03791	11
31LeuCb	41.39763	0.05466	28
31LeuCda	24.84736	0.01603	13
31LeuCdb	24.81237	0.0384	24
31LeuCg	27.28093	0.01122	8
31LeuH	7.70936	0.00445	42
31LeuHa	4.37311	0.00729	18
31LeuHba	1.75123	0.00758	28
31LeuHbb	1.87089	0.00967	17
31LeuHda*	0.74013	0.00516	25
31LeuHdb*	1.00025	0.00821	38
31LeuHg	1.57397	0.00678	21
31LeuN	123.62422	0.03634	28
32AsnC	177.52388	0	1
32AsnCa	56.23581	0	1
32AsnCb	38.1356	9.54E-07	6
32AsnCg	176.01323	0	1
32AsnHba	2.78938	0	1
32AsnHbb	2.78938	4.21E-08	10

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
32AsnHd2a	6.95698	1.86E-04	5
32AsnHd2b	7.64122	5.62E-05	15
32AsnNd2	112.54524	0.01144	11
33AlaC	178.66158	0	1
33AlaCa	54.39929	0.06892	15
33AlaCb	19.10885	4.36E-02	23
33AlaH	8.098	0.0069	38
33AlaHa	3.95319	0.00413	23
33AlaHb*	1.02131	7.60E-03	35
33AlaN	118.87118	4.12E-02	24
34LeuC	179.07422	2.64E-03	2
34LeuCa	57.52323	5.40E-02	16
34LeuCb	40.52628	0.03377	25
34LeuCda	22.24663	0.02422	26
34LeuCdb	26.01356	0.0424	9
34LeuCg	27.36067	0.04508	9
34LeuH	7.34859	0.00671	45
34LeuHa	3.80823	0.00779	29
34LeuHba	1.40757	0.00631	33
34LeuHbb	1.86446	0.00452	21
34LeuHda*	0.49944	0.00435	61
34LeuHdb*	0.86667	0.00668	25
34LeuHg	1.61533	0.00703	20
34LeuN	112.95999	0.05914	27
35AsnC	176.7758	0.01036	2
35AsnCa	56.27331	0.01807	13
35AsnCb	38.42229	0.05383	21
35AsnCg	175.80378	0	1
35AsnH	7.92924	0.00707	47
35AsnHa	4.33103	0.00698	22
35AsnHba	2.59759	0.01213	16
35AsnHbb	2.63837	0.00515	24
35AsnHd2a	7.0024	6.33E-04	18
35AsnHd2b	7.40724	0.01093	11

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
35AsnN	117.71127	0.04012	29
35AsnNd2	113.55683	0.02237	12
36TyrC	176.88796	0.02573	2
36TyrCa	59.45276	0.03486	19
36TyrCb	37.7366	0.04846	35
36TyrCd*	132.66005	0	1
36TyrCe*	118.04474	0	1
36TyrH	7.46267	0.00828	49
36TyrHa	4.53854	0.00762	32
36TyrHba	2.9724	8.91E-03	20
36TyrHbb	3.1166	0.00699	35
36TyrHd*	7.10687	0.00974	24
36TyrHe*	6.79919	0.01009	22
36TyrN	117.06614	0.04334	34
37ValC	179.31462	0.01362	2
37ValCa	64.43297	0.05931	21
37ValCb	31.46946	0.05822	19
37ValCga	22.82251	0.01238	7
37ValCgb	22.10454	0.00877	27
37ValH	7.40128	0.00764	42
37ValHa	4.17977	0.00668	36
37ValHb	2.22757	0.00707	34
37ValHga*	0.8976	0.01047	18
37ValHgb*	0.92473	0.00125	49
37ValN	118.72265	0.03682	26
38SerC	175.74209	0.00634	2
38SerCa	61.41198	0.02185	7
38SerCb	63.2432	0.05317	15
38SerH	7.84792	0.00586	41
38SerHa	4.24204	0.00142	9
38SerHba	3.91378	0.00384	11
38SerHbb	3.91423	0.00394	20
38SerN	115.62187	0.02963	24
39AsnC	176.57423	0.00407	2

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
39AsnCa	52.96333	0.03191	11
39AsnCb	38.38845	0.05368	24
39AsnH	7.78004	0.00508	43
39AsnHa	4.75126	0.00764	15
39AsnHba	2.83915	0.00494	14
39AsnHbb	3.04016	0.00518	30
39AsnN	114.87047	0.04483	27
40GlyC	174.78959	0.02529	2
40GlyCa	45.96676	0.01813	14
40GlyH	7.72537	0.00636	32
40GlyHaa	3.93507	0.00449	18
40GlyHab	4.14638	0.00511	11
40GlyN	106.59235	0.02966	19
41GlnC	176.04329	0.02912	2
41GlnCa	58.42342	0.03217	13
41GlnCb	29.68012	0.06761	18
41GlnCg	35.23414	0.02992	14
41GlnH	8.27381	0.00644	45
41GlnHa	4.18382	0.00656	19
41GlnHba	2.21134	0.01078	9
41GlnHbb	2.2271	0.00916	24
41GlnHe2a	6.61418	2.87E-04	8
41GlnHe2b	7.35596	2.33E-04	23
41GlnHga	2.32399	0.0064	22
41GlnHgb	2.32415	0.00726	10
41GlnN	119.87183	0.03019	25
41GlnNe2	110.95246	0.03009	18
42LeuC	175.58547	0.01321	2
42LeuCa	53.21707	0.04396	13
42LeuCb	45.18054	0.04418	32
42LeuCda	23.87905	0.01308	5
42LeuCdb	25.8442	0.02111	18
42LeuCg	27.28621	0.02465	12
42LeuH	7.63276	0.00813	45

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
42LeuHa	4.76838	8.81E-03	29
42LeuHba	1.37724	1.16E-02	21
42LeuHbb	1.43908	0.00963	30
42LeuHda*	0.86697	0.00681	17
42LeuHdb*	0.86962	0.00391	37
42LeuHg	1.57354	0.00765	24
42LeuN	117.87102	0.04199	26
43SerC	175.71066	0.01885	2
43SerCa	57.15446	0.04504	16
43SerCb	65.86634	0.04544	16
43SerH	9.20301	0.00712	43
43SerHa	4.61363	0.00418	22
43SerHba	4.00583	0.00273	12
43SerHbb	4.32226	0.0033	19
43SerN	118.8835	0.03283	27
44SerC	176.84396	0.01522	2
44SerCa	62.40432	6.91E-04	13
44SerCb	65.79556	0.05038	11
44SerH	9.11091	0.00538	29
44SerHa	4.32037	0.00484	17
44SerHba	4.00219	0.00352	7
44SerHbb	4.00274	0.00362	15
44SerN	118.01482	0.02586	16
45SerC	177.12569	0.02497	3
45SerCa	61.44193	0.04688	5
45SerCb	62.53128	0.08457	7
45SerH	8.16094	0.00522	22
45SerHa	4.23586	0.00373	8
45SerHba	3.86682	3.47E-04	2
45SerHbb	3.86697	3.49E-04	13
45SerN	116.75207	3.37E-02	16
46AspC	179.43647	0.00202	2
46AspCa	57.45354	0.02488	11
46AspCb	40.90815	0.06263	36

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
46AspH	7.80418	0.00512	44
46AspHa	4.45918	0.00553	16
46AspHba	2.47057	0.00566	30
46AspHbb	3.06714	0.00238	20
46AspN	123.38193	0.03453	29
47ValC	176.8234	0.00717	2
47ValCa	67.28414	0.04294	19
47ValCb	31.52236	0.08924	20
47ValCga	21.84001	0.05714	12
47ValCgb	24.3741	3.43E-02	25
47ValH	8.13705	6.33E-03	42
47ValHa	3.45556	0.00607	33
47ValHb	2.27904	0.00542	28
47ValHga*	0.91308	0.00737	21
47ValHgb*	1.04926	0.0047	49
47ValN	121.49047	0.03373	22
48AlaC	179.0604	0.00282	2
48AlaCa	56.18173	0.03029	15
48AlaCb	18.1786	0.03833	14
48AlaH	8.50288	0.00734	46
48AlaHa	3.74719	0.01085	28
48AlaHb*	1.4022	0.0079	35
48AlaN	120.92094	0.04777	28
49SerC	176.91411	0.01225	3
49SerCa	61.36388	0.04679	9
49SerCb	63.02543	0.09182	10
49SerH	8.08221	0.01014	38
49SerHa	4.23481	0.00256	19
49SerHba	3.90597	0.01352	14
49SerHbb	3.91705	0.00227	5
49SerN	110.88434	0.02383	24
50ThrC	176.30363	0.0109	2
50ThrCa	67.43631	0.08687	19
50ThrCb	68.49514	0.02958	15

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
50ThrCg2	20.97973	0.06201	22
50ThrH	7.73441	0.00823	38
50ThrHa	3.77441	0.00887	25
50ThrHb	3.85386	0.0065	28
50ThrHg2*	0.51033	0.00781	43
50ThrN	119.74993	0.04513	24
51ValC	177.10498	0.01445	2
51ValCa	67.2498	0.05143	14
51ValCb	31.6423	0.08531	15
51ValCga	22.36257	0.01538	9
51ValCgb	23.93547	0.05259	19
51ValH	8.43561	0.00952	45
51ValHa	3.35554	0.00747	28
51ValHb	2.04077	0.00559	27
51ValHga*	0.82331	0.01055	23
51ValHgb*	1.02067	0.0065	48
51ValN	119.84904	0.03008	28
52AlaC	178.89635	0.02475	2
52AlaCa	55.86932	0.03979	17
52AlaCb	18.43362	0.05102	22
52AlaH	8.15518	0.00641	49
52AlaHa	3.90723	0.01405	23
52AlaHb*	1.44085	0.00953	42
52AlaN	119.77307	0.03422	28
53ArgC	177.82244	0.00299	2
53ArgCa	59.88757	0.04567	18
53ArgCb	30.08531	0.05969	17
53ArgCd	43.63035	0.03509	15
53ArgCg	27.62131	0.06628	19
53ArgH	7.82432	0.00839	46
53ArgHa	3.90045	0.00532	30
53ArgHba	1.88025	0.01124	29
53ArgHbb	1.88157	0.01052	16
53ArgHda	3.18046	0.00528	12

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
53ArgHdb	3.1875	0.01126	24
53ArgHga	1.5346	0.00639	15
53ArgHgb	1.69644	0.0069	23
53ArgN	117.13018	0.04388	25
54AlaC	180.85193	0.01042	2
54AlaCa	54.77034	0.03528	13
54AlaCb	18.829	0.06328	30
54AlaH	7.64179	0.00658	45
54AlaHa	4.17974	0.00476	21
54AlaHb*	1.31142	0.00807	40
54AlaN	120.51958	0.05888	25
55ValC	176.72248	0.00936	2
55ValCa	66.77003	0.04428	14
55ValCb	31.94197	0.01194	4
55ValCga	24.78034	0.00912	22
55ValCgb	24.78034	0.00912	5
55ValH	8.54036	0.0086	44
55ValHa	3.35465	0.00593	24
55ValHb	2.04726	0.00437	16
55ValHga*	0.96401	0.00647	41
55ValHgb*	0.96495	0.00668	13
55ValN	118.60678	0.01968	26
56AlaC	180.28254	0.01134	2
56AlaCa	56.22262	0.02745	17
56AlaCb	18.16169	0.05294	18
56AlaH	8.47745	0.00654	42
56AlaHa	3.91082	0.00624	26
56AlaHb*	1.42792	0.00546	31
56AlaN	121.84917	0.02819	26
57GlnC	178.52318	0.03309	2
57GlnCa	58.67461	0.03253	9
57GlnCb	28.64198	0.06028	13
57GlnCd	180.22403	0	1
57GlnCg	34.11211	0.04003	10

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
57GlnH	8.30237	0.00803	45
57GlnHa	4.09569	0.006	13
57GlnHba	2.08376	0.00493	12
57GlnHbb	2.17965	0.00793	14
57GlnHga	2.3945	0.0018	9
57GlnHgb	2.55778	0.00735	17
57GlnN	114.77378	0.03815	25
58SerC	176.55841	0.05818	2
58SerCa	61.31987	0.02544	4
58SerCb	64.31902	0.03284	18
58SerH	7.73156	0.00728	40
58SerHa	4.24061	0.00327	7
58SerHba	3.73732	0.0038	19
58SerHbb	3.84238	0.00614	10
58SerN	112.47669	0.03826	30
59LeuC	176.99767	8.39E-04	2
59LeuCa	54.27069	0.02581	9
59LeuCb	43.60713	0.04182	18
59LeuCda	26.75053	0.04685	20
59LeuCdb	22.34532	0.0148	4
59LeuCg	26.69421	0.00997	12
59LeuH	7.72712	0.00964	34
59LeuHa	4.6026	0.00574	17
59LeuHba	1.43872	0.0055	21
59LeuHbb	1.62577	0.00468	16
59LeuHda*	0.65792	0.00736	43
59LeuHdb*	0.78875	0.01397	19
59LeuHg	1.79448	0.0042	21
59LeuN	116.83381	0.03532	22
60GlyC	174.80902	0.02071	2
60GlyCa	46.88244	0.06618	7
60GlyH	7.7127	0.00447	27
60GlyHaa	3.95156	0.0127	6
60GlyHab	3.95156	0.0127	8

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
60GlyN	109.24945	0.01929	19
61LeuC	175.83399	1.17E-02	2
61LeuCa	53.02356	0.03835	14
61LeuCb	44.41627	0.03928	28
61LeuCda	25.74455	0.04789	25
61LeuCdb	22.34623	0.03238	11
61LeuCg	27.08516	0.05995	7
61LeuH	7.31587	0.00835	37
61LeuHa	4.55354	0.00492	22
61LeuHba	1.18699	0.00433	20
61LeuHbb	1.57595	0.00653	30
61LeuHda*	0.63171	0.00651	40
61LeuHdb*	0.82029	0.00377	19
61LeuHg	1.43681	0.00722	25
61LeuN	118.02662	0.02937	21
62SerC	176.10275	0.01176	2
62SerCa	57.76689	0.03177	13
62SerCb	63.29934	0.03312	13
62SerH	8.40618	0.004	32
62SerHa	4.39545	0.01251	17
62SerHba	3.91427	0.00344	7
62SerHbb	4.00744	0.00825	18
62SerN	114.29112	0.03666	21
63GlnC	177.95506	0.00999	3
63GlnCa	60.16909	0.02062	18
63GlnCb	28.41284	0.01438	16
63GlnCd	180.1357	0	1
63GlnCg	33.89139	0.03107	23
63GlnH	9.05499	0.00364	31
63GlnHa	3.6358	0.00965	30
63GlnHba	1.96092	0.00628	24
63GlnHbb	2.08141	0.00519	12
63GlnHga	2.0918	0.00667	20
63GlnHgb	2.3007	0.01328	25

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
63GlnN	129.04205	0.03724	19
64GlyC	176.44792	0.01893	2
64GlyCa	47.25343	0.0541	18
64GlyH	8.73262	0.0036	33
64GlyHaa	3.73862	0.00494	10
64GlyHab	3.8647	0.01206	22
64GlyN	106.41588	0.02524	21
65SerC	176.91046	0.0179	3
65SerCa	61.18565	0.05649	12
65SerCb	62.86971	0.06333	16
65SerH	7.47572	0.00818	40
65SerHa	4.31978	0.01039	20
65SerHba	3.95436	0.00897	18
65SerHbb	3.95456	0.009	8
65SerN	116.49069	0.04328	24
66ValC	177.67735	0.01348	2
66ValCa	66.17996	0.05337	19
66ValCb	31.66152	0.06124	14
66ValCga	22.77415	0.06246	31
66ValCgb	23.4384	0.06563	12
66ValH	7.71549	0.00549	40
66ValHa	3.60635	0.00864	30
66ValHb	2.04486	0.00418	24
66ValHga*	0.86774	0.00993	49
66ValHgb*	0.93727	0.00826	21
66ValN	119.74994	0.04119	24
67GlnC	177.81699	0.01863	2
67GlnCa	60.13348	0.02842	17
67GlnCb	28.24201	0.05132	14
67GlnCd	178.8698	0	1
67GlnCg	33.86266	0.0656	18
67GlnH	8.87602	0.00673	51
67GlnHa	3.65533	0.01008	26
67GlnHba	2.1617	0.00323	6

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
67GlnHbb	2.16201	0.00454	19
67GlnHe2a	6.53608	0.00233	12
67GlnHe2b	7.29332	7.17E-04	19
67GlnHga	2.36508	0.00367	27
67GlnHgb	2.36513	0.00357	10
67GlnN	120.87141	0.03499	28
67GlnNe2	110.26057	0.03691	17
68AsnC	177.60288	0.04136	2
68AsnCa	56.69116	0.03757	13
68AsnCb	38.41604	0.06178	20
68AsnCg	175.78834	0.00118	2
68AsnH	7.80244	0.00655	44
68AsnHa	4.45019	0.0084	20
68AsnHba	2.88343	0.01038	22
68AsnHbb	2.91981	0.01395	12
68AsnHd2a	6.81932	5.34E-04	10
68AsnHd2b	7.56395	0.00118	20
68AsnN	117.77002	0.03283	24
68AsnNd2	112.2415	0.00967	15
69IleC	178.45854	0.01783	2
69IleCa	64.83942	0.03495	16
69IleCb	38.7297	0.09722	12
69IleCd1	13.36039	0.04907	19
69IleCg1	29.19033	0.02125	23
69IleCg2	17.70512	0.02597	24
69IleH	7.53018	0.00634	51
69IleHa	3.85942	6.02E-03	31
69IleHb	1.85378	0.00628	28
69IleHd1*	0.82419	0.00787	32
69IleHg1a	1.1832	0.00594	20
69IleHg1b	1.77618	0.00397	31
69IleHg2*	0.99419	0.00502	44
69IleN	120.09317	0.0341	31
70MetC	178.39796	0.02007	2

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
70MetCa	57.52022	0.01985	21
70MetCb	32.08836	0.08723	26
70MetCg	32.63665	0.05785	27
70MetH	8.54196	0.00786	52
70MetHa	4.21341	0.01085	32
70MetHba	1.90681	1.03E-02	19
70MetHbb	2.18616	0.00721	34
70MetHga	2.59855	0.01435	22
70MetHgb	2.62964	0.01998	35
70MetN	118.06854	0.04623	31
71SerC	176.8297	0.02556	3
71SerCa	62.14738	0.01932	9
71SerCb	62.67504	0.05012	16
71SerH	8.68897	0.00527	47
71SerHa	4.19497	0.00529	10
71SerHba	3.97536	0.00632	21
71SerHbb	3.99753	0.00597	6
71SerN	114.52044	0.01853	27
72GlnC	178.97555	0.00761	2
72GlnCa	58.99988	0.01913	8
72GlnCb	28.69346	0.0405	15
72GlnCd	180.15458	0	1
72GlnCg	33.98056	0.06375	10
72GlnH	7.72709	0.00693	48
72GlnHa	4.11707	0.00973	16
72GlnHba	2.19026	0.00675	5
72GlnHbb	2.22567	0.00648	19
72GlnHe2a	6.7174	0.00385	11
72GlnHe2b	7.30452	0.00124	21
72GlnHga	2.37603	0.01223	9
72GlnHgb	2.53383	0.00747	16
72GlnN	120.37452	0.02724	29
72GlnNe2	111.15488	0.08421	21
73GlnC	178.44227	0.01817	2

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
73GlnCa	58.40886	0.01933	8
73GlnCb	28.96705	0.04766	33
73GlnCd	178.71977	0	1
73GlnCg	34.14389	0.01711	22
73GlnH	8.18788	0.00472	40
73GlnHa	4.18632	0.01063	21
73GlnHba	1.99181	0.00983	26
73GlnHbb	2.28499	0.009	32
73GlnHe2a	6.78194	0.00244	18
73GlnHe2b	7.36502	0.00757	42
73GlnHga	2.4136	0.00898	18
73GlnHgb	2.57452	0.00789	32
73GlnN	117.03572	0.04255	26
73GlnNe2	111.25399	0.08204	29
74LeuC	178.24665	0.03349	2
74LeuCa	56.01651	0.02937	11
74LeuCb	41.07282	0.04976	24
74LeuCda	22.90606	0.06498	36
74LeuCdb	26.23364	0.0368	7
74LeuCg	27.27065	0.00958	12
74LeuH	8.23413	0.00716	41
74LeuHa	4.3654	0.00815	25
74LeuHba	1.59403	0.01121	35
74LeuHbb	1.8401	0.00688	18
74LeuHda*	0.7607	0.00647	54
74LeuHdb*	0.82514	0.0106	25
74LeuHg	1.75899	0.01447	22
74LeuN	116.96362	0.04134	25
75SerC	175.50849	0.01113	2
75SerCa	61.30369	0.02449	7
75SerCb	63.24055	0.04937	8
75SerH	7.67081	0.00345	34
75SerHa	4.23526	0.00285	9
75SerHba	4.03162	0.00192	7

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
75SerHbb	4.03162	0.00192	4
75SerN	114.27076	0.02058	22
76SerC	174.83391	0.02946	2
76SerCa	59.04736	0.05259	8
76SerCb	63.76901	0.06245	21
76SerH	7.7161	0.00472	30
76SerHa	4.4893	0.00928	13
76SerHba	3.91952	0.01322	22
76SerHbb	4.00592	0.00869	12
76SerN	114.37157	0.03235	20
77IleC	176.01117	0.01795	2
77IleCa	61.91219	0.0261	10
77IleCb	38.36253	0.03619	8
77IleCd1	13.69677	8.35E-04	13
77IleCg1	27.00626	0.02642	20
77IleCg2	17.72966	0.02447	12
77IleH	7.32001	0.00706	39
77IleHa	4.19271	0.00296	18
77IleHb	1.92861	0.00566	16
77IleHd1*	0.82552	0.01181	19
77IleHg1a	1.32541	0.00845	29
77IleHg1b	1.53664	0.01453	16
77IleHg2*	0.90937	0.00785	22
77IleN	119.2369	0.01944	23
78GlyC	173.48758	0.01361	2
78GlyCa	45.18373	0.02243	13
78GlyH	7.86298	0.0067	38
78GlyHaa	4.02142	0.0035	7
78GlyHab	4.02224	0.00315	12
78GlyN	110.87604	0.0392	22
79SerC	175.84047	0.02344	2
79SerCa	59.17268	0.02315	12
79SerCb	63.71693	0.06512	10
79SerH	8.273	0.00529	29

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
79SerHa	4.42892	0.00157	15
79SerHba	3.90161	0.00776	18
79SerHbb	3.90161	0.00776	5
79SerN	115.60014	0.03266	20
80GlyC	174.36464	0.00661	2
80GlyCa	45.73234	0.00866	11
80GlyH	8.87217	0.00394	30
80GlyHaa	3.98786	1.21E-03	17
80GlyHab	4.42846	1.73E-04	3
80GlyN	112.54487	0.02335	18
81AlaC	177.28797	0.02809	2
81AlaCa	52.37059	0.01473	9
81AlaCb	19.781	0.05755	16
81AlaH	7.62637	0.00582	37
81AlaHa	4.28439	0.01119	12
81AlaHb*	1.40089	0.0105	25
81AlaN	122.83788	0.02303	22
82SerC	175.57466	0.00188	2
82SerCa	57.18305	0.03521	11
82SerCb	65.05579	0.02647	14
82SerH	8.48701	0.00784	30
82SerHa	4.56823	0.00521	15
82SerHba	4.06372	0.00594	8
82SerHbb	4.34142	0.00542	20
82SerN	116.37972	0.02106	20
83ThrC	177.70978	0.02853	2
83ThrCa	66.47588	0.04318	13
83ThrCb	68.10118	0.04868	12
83ThrCg2	22.7742	0.05296	20
83ThrH	8.94356	0.00859	28
83ThrHa	3.89757	0.00601	21
83ThrHb	4.26417	0.00403	16
83ThrHg2*	1.23442	0.00772	33
83ThrN	115.42954	0.03784	15

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
84SerC	176.5715	0	1
84SerCa	61.4512	0	5
84SerCb	62.36467	0.0058	7
84SerH	8.32488	0.00908	25
84SerHa	4.23395	2.20E-03	9
84SerHba	3.87079	0.00783	7
84SerN	118.02651	0.0124	13
85SerC	177.41869	0.04089	2
85SerCa	61.50318	0.06459	8
85SerCb	62.74516	0.07393	12
85SerH	7.98218	0.00576	36
85SerHa	4.27035	0.00449	12
85SerHba	3.91454	0.00724	10
85SerHbb	4.00132	0.00219	15
85SerN	119.81552	0.04009	22
86LeuC	178.33525	0.01569	2
86LeuCa	58.22527	0.02293	17
86LeuCb	43.23983	0.04703	23
86LeuCda	26.04572	0.01051	3
86LeuCdb	25.06414	0.03695	26
86LeuCg	27.05025	0.01215	11
86LeuH	8.44763	0.01222	32
86LeuHa	4.02752	0.00863	26
86LeuHba	1.40655	0.00497	30
86LeuHbb	1.87583	0.01029	26
86LeuHda*	0.85531	0.00721	19
86LeuHdb*	0.86983	0.00738	48
86LeuHg	1.68301	0.00808	22
86LeuN	123.41151	0.01863	22
87SerC	175.94419	0.03261	2
87SerCa	62.63681	0.01001	15
87SerH	8.3873	0.00433	32
87SerHa	4.04657	0.00393	17
87SerHba	3.86155	0	1

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
87SerHbb	3.86155	0	1
87SerN	114.78468	0.0406	21
88GlnC	177.65256	0.01671	2
88GlnCa	58.97798	0.04429	18
88GlnCb	28.56735	0.01264	22
88GlnCd	179.72012	0.00598	2
88GlnCg	33.70246	0.03632	8
88GlnH	7.59352	0.0074	45
88GlnHa	3.8765	0.00722	25
88GlnHba	2.13876	0.00358	11
88GlnHbb	2.16746	7.10E-04	18
88GlnHe2a	6.63954	0.00411	11
88GlnHe2b	7.25252	0.0053	27
88GlnHga	2.37255	0.00452	10
88GlnHgb	2.39605	0.01054	12
88GlnN	120.59461	0.04526	29
88GlnNe2	111.22736	0.02664	17
89AlaC	180.56114	0.0054	2
89AlaCa	55.3638	0.02277	15
89AlaCb	18.54609	0.05619	21
89AlaH	8.07645	0.00927	38
89AlaHa	4.21009	0.00792	28
89AlaHb*	1.48678	0.00715	34
89AlaN	122.05086	0.02477	23
90IleC	177.12978	0.01355	2
90IleCa	65.89455	0.05675	16
90IleCb	38.27119	0.06868	12
90IleCd1	13.76758	0.00726	13
90IleCg1	29.734	0.0298	12
90IleCg2	18.45166	0.02313	13
90IleH	8.4363	0.00898	40
90IleHa	3.55283	0.00653	35
90IleHb	1.7887	0.00956	28
90IleHd1*	0.78606	0.00502	19

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
90IleHg1a	0.85985	0.00967	13
90IleHg1b	0.86021	0.00998	20
90IleHg2*	0.79809	0.0082	23
90IleN	117.16553	0.02173	23
91AlaC	178.91793	0.01899	2
91AlaCa	55.99719	0.01219	14
91AlaCb	19.66981	0.02367	22
91AlaH	8.16181	0.00633	44
91AlaHa	3.84357	0.01042	24
91AlaHb*	1.6032	7.46E-03	52
91AlaN	121.31561	0.0411	23
92AsnC	177.98998	0.01157	2
92AsnCa	56.19479	0.05278	16
92AsnCb	38.74	0.0571	21
92AsnCg	176.13846	0.01431	2
92AsnH	8.56633	0.00792	45
92AsnHa	4.37526	0.00607	24
92AsnHba	2.75115	0.0081	31
92AsnHbb	2.88745	0.0086	21
92AsnHd2a	6.90583	0.0051	32
92AsnHd2b	7.4394	0.00827	11
92AsnN	115.95302	0.0214	30
92AsnNd2	112.06622	0.02203	22
93AlaC	179.22833	0.00788	2
93AlaCa	55.26379	0.02446	17
93AlaCb	19.46862	0.06514	33
93AlaH	8.0165	0.00579	48
93AlaHa	4.04765	0.00543	28
93AlaHb*	1.50287	0.00685	48
93AlaN	123.46078	0.02251	30
94ValC	177.62674	0.01596	2
94ValCa	67.50511	0.03148	16
94ValCb	31.19118	0.02316	13
94ValCga	21.93635	0.01516	7

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
94ValCgb	23.55022	0.01451	20
94ValH	8.44533	0.00747	43
94ValHa	3.30265	0.00449	31
94ValHb	2.12256	0.00826	24
94ValHga*	0.78611	0.01106	16
94ValHgb*	0.90533	0.00885	33
94ValN	117.23795	0.05348	25
95SerC	177.23975	0.02132	2
95SerCa	62.25806	0.02255	9
95SerCb	63.34138	0.05038	12
95SerH	8.31871	0.00542	47
95SerHa	3.99419	0.01035	17
95SerHba	3.95825	0.00982	7
95SerHbb	3.9743	0.01092	17
95SerN	114.38216	0.04494	27
96SerC	175.79427	0.00184	2
96SerCa	62.12844	0.01206	7
96SerCb	63.16866	0.02998	10
96SerH	8.07577	0.00363	38
96SerHa	4.22309	0.00651	10
96SerHba	4.043	0.00355	12
96SerHbb	4.04748	0.00591	5
96SerN	115.66912	0.04201	25
97AlaC	178.18281	0.00475	2
97AlaCa	54.09223	0.01779	11
97AlaCb	18.76964	0.06475	20
97AlaH	7.93175	0.00479	48
97AlaHa	4.12908	0.01351	21
97AlaHb*	1.39379	0.00707	35
97AlaN	122.0403	0.04411	29
98ValC	177.78048	0.00152	2
98ValCa	63.81786	0.03209	18
98ValCb	32.48375	0.04877	14
98ValCga	21.86966	0.01873	33

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
98ValCgb	23.06237	0.06673	11
98ValH	8.0078	0.00888	40
98ValHa	3.65773	0.00496	34
98ValHb	1.95525	0.00587	30
98ValHga*	0.64559	0.00693	55
98ValHgb*	0.8209	0.00456	30
98ValN	113.79446	0.03975	22
99GlnC	176.25892	0.02372	2
99GlnCa	58.32332	0.0443	12
99GlnCb	28.4829	0.02053	10
99GlnCg	33.81634	0.04009	8
99GlnH	7.89083	0.00539	37
99GlnHa	3.78955	0.00581	22
99GlnHba	2.15852	0.0098	16
99GlnHbb	2.17286	0.0066	7
99GlnHga	2.47503	0.00643	7
99GlnHgb	2.47719	0.00656	15
99GlnN	121.20031	0.03237	26
100GlyC	173.83604	0.01228	2
100GlyCa	45.53751	0.02451	21
100GlyH	8.47715	0.00715	36
100GlyHaa	3.91339	0.00329	25
100GlyHab	3.99816	0.01158	10
100GlyN	110.78486	0.03808	23
101SerC	175.76347	7.24E-04	2
101SerCa	58.08238	0.04599	9
101SerCb	66.93433	0.04155	21
101SerH	7.83576	0.00587	35
101SerHa	4.69575	0.00464	15
101SerHba	3.89723	0.00748	17
101SerHbb	4.1089	0.00531	14
101SerN	115.82316	0.04879	23
102GlnC	176.10499	0.01224	2
102GlnCa	57.46058	0.0339	8

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
102GlnCb	28.22954	0.03792	17
102GlnCg	34.12545	0.05445	13
102GlnH	8.88858	0.00838	40
102GlnHa	4.11599	0.00479	15
102GlnHba	2.05965	0.00981	10
102GlnHbb	2.13333	0.01337	18
102GlnHga	2.39701	0.00826	18
102GlnN	119.86291	0.02862	23
103AlaC	177.29846	0.02444	2
103AlaCa	52.46314	0.0235	10
103AlaCb	19.40321	0.03733	12
103AlaH	7.75061	0.00523	39
103AlaHa	4.23243	0.00396	18
103AlaHb*	1.31081	0.01363	33
103AlaN	119.31605	0.04043	23
104AlaC	174.51509	1.61E-04	2
104AlaCa	51.3539	0.02965	16
104AlaCb	16.32598	0.05435	22
104AlaH	6.85309	0.0081	42
104AlaHa	3.35448	0.00419	21
104AlaHb*	0.49452	0.0047	43
104AlaN	121.68764	0.03048	24
105AlaC	174.52751	0.00355	2
105AlaCa	49.33792	0.02624	12
105AlaCb	19.14856	0.05142	13
105AlaH	7.26263	0.00641	35
105AlaHa	4.74336	0.00417	15
105AlaHb*	1.26346	8.00E-03	23
105AlaN	128.12535	0.0295	24
106ProC	177.49624	0	1
106ProCa	64.16107	0.02593	10
106ProCb	32.20661	0.0124	2
106ProHa	4.31793	0.00802	15
107GlyC	175.77375	0.00642	2

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
107GlyCa	45.8423	0.03443	24
107GlyH	8.99722	0.00664	43
107GlyHaa	3.9236	0.00571	18
107GlyHab	4.43292	0.00488	31
107GlyN	114.15925	0.02495	30
108GlnC	176.1238	0.02509	2
108GlnCa	54.83353	0.03883	17
108GlnCb	28.39237	0.05545	23
108GlnCd	180.18949	0	1
108GlnCg	33.90054	0.02161	20
108GlnH	7.87538	0.00706	40
108GlnHa	4.78984	0.00662	36
108GlnHba	2.17872	0.00672	24
108GlnHbb	2.72403	0.00487	18
108GlnHe2a	6.80322	1.38E-03	19
108GlnHe2b	7.49391	0.00402	11
108GlnHga	2.21886	0.00371	21
108GlnHgb	2.4198	0.01028	17
108GlnN	117.56748	0.02611	24
108GlnNe2	112.97015	0.04459	17
109GluC	178.21638	0.00243	2
109GluCa	63.36014	0.02291	20
109GluCb	29.00537	0.04957	31
109GluCg	36.3139	0.03431	30
109GluH	10.41853	0.00501	46
109GluHa	3.66619	0.00707	29
109GluHba	1.31558	0.00903	41
109GluHbb	1.98403	0.00965	22
109GluHga	1.34077	0.00712	17
109GluHgb	2.45012	0.00362	32
109GluN	122.25366	0.02352	29
110GlnC	178.12216	4.52E-04	2
110GlnCa	59.6794	0.05431	14
110GlnCb	28.08098	0.04904	9

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
110GlnCg	34.39988	0.03386	10
110GlnH	8.5467	0.00737	37
110GlnHa	3.94591	6.42E-03	27
110GlnHba	2.06118	0.00427	21
110GlnHbb	2.06478	0.00562	9
110GlnHe2a	6.29142	7.05E-04	14
110GlnHe2b	7.14005	0.00106	35
110GlnHga	2.28994	0.01083	19
110GlnHgb	2.29557	0.00454	7
110GlnN	116.26041	0.0434	23
110GlnNe2	108.1814	0.04378	26
111SerC	177.837	0.04799	2
111SerCa	61.2861	0.07279	8
111SerCb	62.83602	0.0194	16
111SerH	7.99076	0.00476	40
111SerHa	4.269	0.00555	12
111SerHba	4.00179	0.00743	24
111SerHbb	4.04014	0.00777	4
111SerN	116.50849	0.04706	27
112IleC	177.51454	0.02956	2
112IleCa	65.66077	0.06048	16
112IleCb	37.96631	0.01832	11
112IleCd1	14.47546	0.03397	23
112IleCg1	29.49593	0.01438	17
112IleCg2	18.33449	0.07347	17
112IleH	8.85991	0.0104	39
112IleHa	3.66771	0.00568	33
112IleHb	1.98244	0.00711	28
112IleHd1*	1.00311	0.00564	42
112IleHg1a	1.15585	0.00633	28
112IleHg2*	0.91233	0.01093	35
112IleN	122.27562	0.0215	22
113AlaC	178.99604	0.00622	2
113AlaCa	56.00529	0.04758	17

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
113AlaCb	18.42166	0.02709	25
113AlaH	7.99688	0.00456	45
113AlaHa	3.80073	0.00674	25
113AlaHb*	1.54136	0.00825	45
113AlaN	120.96789	3.61E-02	27
114GlnC	179.09969	0.01578	2
114GlnCa	59.20694	0.02436	11
114GlnCb	28.48172	0.03245	12
114GlnCg	33.91568	0.01016	16
114GlnH	8.03095	0.00601	37
114GlnHa	4.0738	0.00875	19
114GlnHba	2.19347	4.58E-04	11
114GlnHbb	2.21806	3.41E-03	9
114GlnHga	2.47355	0.01064	11
114GlnHgb	2.56393	0.00268	22
114GlnN	116.53786	0.03132	26
115ArgC	179.84458	0.02563	2
115ArgCa	59.61885	0.05173	20
115ArgCb	30.95145	0.02356	29
115ArgCd	44.19972	0.00975	8
115ArgCg	27.54554	0.03094	12
115ArgH	8.24547	0.00866	45
115ArgHa	4.11399	0.00661	29
115ArgHba	1.88677	0.00597	28
115ArgHbb	1.99821	0.0118	20
115ArgHda	3.18842	0.01366	10
115ArgHdb	3.19129	0.01226	22
115ArgHga	1.71271	0.00231	23
115ArgHgb	1.87033	0.00343	8
115ArgN	120.96855	0.05004	26
116ValC	177.57128	0.02465	2
116ValCa	66.23804	0.04123	19
116ValCb	31.38382	0.07214	11
116ValCga	23.57704	0.02581	13

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
116ValCgb	24.17929	0.0707	22
116ValH	8.82535	0.00655	45
116ValHa	3.78273	0.00612	38
116ValHb	2.13803	0.00895	24
116ValHga*	0.97665	0.01151	26
116ValHgb*	1.06098	0.00871	42
116ValN	119.29759	0.04211	27
117AsnC	177.98111	0.01481	2
117AsnCa	58.34834	0.06073	16
117AsnCb	39.86364	0.04248	26
117AsnCg	174.10597	0	1
117AsnH	8.75111	0.00596	41
117AsnHa	4.17652	0.00805	25
117AsnHba	2.71318	0.00424	35
117AsnHbb	2.8738	0.00351	23
117AsnHd2a	6.40318	0.00446	15
117AsnHd2b	7.40998	0.00709	28
117AsnN	120.34187	0.03455	25
117AsnNd2	110.68778	0.02095	22
118SerC	176.53653	0.01224	2
118SerCa	61.95014	5.64E-02	11
118SerCb	62.84363	0.02224	9
118SerH	8.03274	0.00873	37
118SerHa	4.21766	0.01277	14
118SerHba	4.04011	0.00924	6
118SerHbb	4.04298	0.00728	12
118SerN	114.35923	0.03216	24
119AlaC	180.68465	5.54E-04	2
119AlaCa	55.18037	0.04508	10
119AlaCb	18.89568	0.04004	21
119AlaH	7.80287	0.00387	45
119AlaHa	4.25027	0.00675	13
119AlaHb*	1.57117	0.0044	39
119AlaN	124.26345	0.05281	27

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
120IleC	177.34358	0.02435	2
120IleCa	66.51233	0.03985	16
120IleCb	38.07446	0.01988	11
120IleCd1	15.00228	0.01507	16
120IleCg1	26.99375	0.00208	15
120IleCg2	17.97805	0.0317	20
120IleH	8.97628	0.00572	44
120IleHa	3.38744	0.00599	27
120IleHb	1.9611	0.01252	23
120IleHd1*	0.76107	0.00892	30
120IleHg1a	1.32732	0.00836	26
120IleHg1b	1.56474	0.00845	9
120IleHg2*	0.87128	0.00728	27
120IleN	119.95373	0.02337	25
121SerC	177.67977	0	1
121SerCa	62.64914	0.0456	13
121SerH	8.50196	0.00831	39
121SerHa	4.15985	0.00925	18
121SerHba	4.06025	0.01153	7
121SerHbb	4.06276	0.01053	6
121SerN	115.26628	0.0467	22
122SerC	176.5817	0.00362	2
122SerCa	61.72768	0.08918	9
122SerCb	62.91043	0.0621	9
122SerH	8.10781	0.00574	27
122SerHa	4.29712	0.00668	14
122SerHba	3.98871	0.01195	12
122SerHbb	3.989	0.01036	4
122SerN	116.26017	0.04803	21
123AlaC	181.25063	0.03283	2
123AlaCa	55.1483	0.04855	9
123AlaCb	18.72027	0.0605	18
123AlaH	8.09334	0.00914	39
123AlaHa	4.19551	0.00615	13

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
123AlaHb*	1.3791	0.00427	33
123AlaN	124.69209	0.03458	23
124PheC	177.61757	0.01511	2
124PheCa	62.37225	0.0076	19
124PheCb	38.80355	4.81E-02	25
124PheCd*	131.65658	0	1
124PheCe*	130.225	0	1
124PheCz	128.37708	0	1
124PheH	8.62583	0.00697	45
124PheHa	4.20633	0.00715	24
124PheHba	2.92856	0.00458	15
124PheHbb	3.08951	0.00545	38
124PheHd*	7.1956	0.01064	32
124PheHe*	6.92316	0.01303	29
124PheHz	6.96934	0.01815	16
124PheN	117.12719	0.03108	28
125AlaC	180.70711	2.70E-04	2
125AlaCa	55.64846	0.05146	14
125AlaCb	17.83334	0.05449	18
125AlaH	8.0131	0.00629	42
125AlaHa	4.14125	0.00646	20
125AlaHb*	1.59151	0.00652	27
125AlaN	121.86286	0.04031	26
126GlnC	177.9033	0.00841	2
126GlnCa	58.62774	0.037	6
126GlnCb	28.56634	0.02472	15
126GlnCg	34.12577	0.01921	8
126GlnH	7.7218	0.01071	35
126GlnHa	4.0855	0.00648	13
126GlnHba	2.13051	0.0062	22
126GlnHbb	2.14825	0.01188	7
126GlnHga	2.36492	0.00585	6
126GlnHgb	2.49458	0.00625	13
126GlnN	117.2781	0.03733	21

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
127LeuC	178.72649	0.03148	2
127LeuCa	58.08197	0.06349	13
127LeuCb	42.64522	0.03692	27
127LeuCda	25.8618	0.08164	17
127LeuCdb	23.73268	0.07333	5
127LeuCg	27.11004	0.04054	8
127LeuH	7.65898	0.005	39
127LeuHa	4.11729	0.00972	22
127LeuHba	1.65	0.00705	19
127LeuHbb	1.84129	0.00516	31
127LeuHda*	0.82268	0.00688	40
127LeuHdb*	0.82718	0.00742	20
127LeuHg	1.50743	0.00646	21
127LeuN	121.06729	0.02356	26
128IleC	178.65569	0.03651	2
128IleCa	64.00773	0.04141	14
128IleCb	37.62844	0.06166	11
128IleCd1	12.85125	0.04632	28
128IleCg1	29.03487	0.02529	38
128IleCg2	17.55533	0.06628	20
128IleH	8.29244	0.00812	45
128IleHa	3.62279	0.00791	35
128IleHb	1.85133	0.0049	26
128IleHd1*	0.61215	0.00964	47
128IleHg1a	1.05076	0.00686	30
128IleHg1b	1.48187	0.00598	41
128IleHg2*	0.69373	0.00925	38
128IleN	116.77057	0.02307	27
129SerC	175.75486	0.00536	2
129SerCa	61.12605	4.64E-02	9
129SerCb	63.20336	0.06174	14
129SerH	7.86188	0.00823	38
129SerHa	4.23497	0.00742	13
129SerHba	3.99223	0.00216	18

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
129SerHbb	3.99272	0.00233	7
129SerN	116.32302	0.04631	23
130GlnC	177.26335	0.0257	2
130GlnCa	57.39526	0.02524	13
130GlnCb	28.96732	0.01771	20
130GlnCg	34.13375	0.02016	10
130GlnH	7.79349	0.00361	34
130GlnHa	4.25276	0.00777	21
130GlnHba	2.18024	0.00599	23
130GlnHbb	2.18185	0.00647	12
130GlnHga	2.41447	0.00855	6
130GlnHgb	2.52987	0.01419	17
130GlnN	119.38398	0.0449	23
131ArgC	176.65182	0.00352	2
131ArgCa	56.26646	0.04304	11
131ArgCb	30.72001	0.02503	21
131ArgCd	43.0154	0.03035	14
131ArgCg	27.28145	0.03722	13
131ArgH	7.85299	0.00655	40
131ArgHa	4.38309	0.00496	17
131ArgHba	1.81921	0.00477	10
131ArgHbb	1.95768	0.00415	24
131ArgHda	3.15594	0.00609	21
131ArgHga	1.6749	0.00603	10
131ArgHgb	1.68587	0.01043	17
131ArgN	118.20396	0.03008	23
132ThrC	173.79612	0.01794	2
132ThrCa	61.92252	0.02457	8
132ThrCb	70.17622	0.02665	10
132ThrCg2	21.59587	0.01373	20
132ThrH	7.78154	0.00434	38
132ThrHa	4.37397	0.00451	10
132ThrHb	4.25482	0.00718	17
132ThrHg2*	1.18729	0.00382	28

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
132ThrN	112.46429	0.03112	23
133AlaC	175.21345	0.01255	2
133AlaCa	50.69543	0.00623	8
133AlaCb	18.3728	0.09716	9
133AlaH	7.94628	0.00654	26
133AlaHa	4.61278	0.01011	13
133AlaHb*	1.37856	4.69E-04	14
133AlaN	127.61544	0.02921	16
134ProC	176.32168	0.02248	2
134ProCa	62.89742	9.54E-07	4
134ProCb	32.03003	0	1
134ProHa	4.38288	8.43E-08	4
135AlaC	175.50086	0	1
135AlaCa	50.36255	0.01816	4
135AlaCb	18.2331	2.38E-07	6
135AlaH	8.2298	0.01459	22
135AlaHa	4.57225	0.00362	4
135AlaHb*	1.34362	0.00598	8
135AlaN	125.56785	0.02727	15
136ProC	176.32168	0.02248	2
136ProCa	62.86628	0.03115	4
136ProCb	32.02651	0.00352	2
136ProHa	4.38288	8.43E-08	4
137AlaC	175.5257	0.02484	2
137AlaCa	50.35956	0.01656	5
137AlaCb	18.22885	0.00425	5
137AlaH	8.2298	0.01459	21
137AlaHa	4.57225	0.00362	4
137AlaHb*	1.34362	0.00598	9
137AlaN	125.56867	0.02378	21
138ProCa	62.99939	0.06998	2
138ProCb	32.08398	0.04379	2
139ArgC	174.29469	0	1
139ArgCa	53.93645	0.06767	3

Table A4. Continued

Resonance	Chemical Shift (ppm)	SD	Peaks
139ArgCb	30.33244	0.01715	3
139ArgH	8.28393	0.02053	16
139ArgHa	4.61177	0	2
139ArgHda	3.22239	6.87E-05	2
139ArgN	122.31174	0.06445	15
140ProCa	62.99939	0.06998	2
140ProCb	32.08398	0.04379	2
141ArgC	174.29423	0	1
141ArgCa	53.93645	0.06767	4
141ArgCb	30.33153	0.01222	4
141ArgH	8.26754	0.01929	25
141ArgHa	4.61517	0.00229	6
141ArgHda	3.22239	6.87E-05	2
141ArgN	122.26337	0.07047	18
142ProC	176.41684	0	1
142ProCa	63.18899	0	1
142ProCb	32.14227	0	1
143AlaC	176.80851	0.00E+00	1
143AlaCa	52.66991	9.54E-07	4
143AlaCb	19.36139	0	1
143AlaH	8.33558	5.91E-04	12
143AlaHa	4.41502	5.96E-08	5
143AlaN	124.84057	2.22E-02	9
144SerC	178.53994	0	1
144SerCa	59.87636	0	1
144SerCb	65.03404	0.00E+00	1
144SerH	7.78181	1.16E-04	9
144SerN	120.26894	0.00584	7

Table A5. Tabulated contacts assigned per residue.

Seq Element	Total	Intra Residue	Inter Residue	Sequential	Short Range	Short Non-seq	Long Range	Intra Chain	Contacted Residues
Chain A	4404	1840	2564	947	1821	874	743	4404	None
11Ala	5	5	0	0	0	0	0	5	
12Pro	2	0	2	2	2	0	0	2	13
13Val	26	17	9	8	9	1	0	26	12 14 15
14Tyr	33	17	16	14	14	0	2	33	13 15 22 23
15Phe	78	29	49	13	17	4	32	78	13 14 16 17 20 98 104 106 107 108 109 110 112
16Asn	42	13	29	13	27	14	2	42	15 17 18 19 20 109
17Ser	28	5	23	12	20	8	3	28	15 16 18 19 20 95 100
18Gln	45	21	24	13	24	11	0	45	16 17 19 20 21 22
19Thr	79	12	67	16	46	30	21	79	16 17 18 20 21 22 23 100 107 109 110 111 113
20Leu	124	36	88	20	42	22	46	124	15 16 17 18 19 21 22 23 24 26 91 94 95 98 99 100 104 107 109 110 112 113 116
21Thr	64	12	52	17	30	13	22	64	18 19 20 22 23 24 25 88 89 91 92 94 95
22Asn	36	13	23	9	22	13	1	36	14 18 19 20 21 23 24 25
23Asn	82	22	60	15	49	34	11	82	14 19 20 21 22 24 25 26 110 111 113
24Leu	112	34	78	23	50	27	28	112	20 21 22 23 25 26 27 28 90 91 94 112 113 116 117 118 120
25Ala	58	6	52	20	42	22	10	58	21 22 23 24 26 27 28 29 88 90 91 92 113
26Ser	53	12	41	10	29	19	12	53	20 23 24 25 27 28 29 91 110
27Ser	23	3	20	9	18	9	2	23	24 25 26 28 29 117
28Leu	106	28	78	18	57	39	21	106	24 25 26 27 29 30 31 34 86 87 88 90 91 92 94 120 121
29Gln	56	19	37	17	32	15	5	56	25 26 27 28 30 31 34 84 87 88
30Ser	21	5	16	14	16	2	0	21	28 29 31

Table A5. Continued

Seq Element	Total	Intra Residue	Inter Residue	Sequential	Short Range	Short Non-seq	Long Range	Intra Chain	Contacted Residues
31Leu	94	27	67	9	52	43	15	94	28 29 30 33 34 35 116 120 121 124 125 128
32Asn	3	3	0	0	0	0	0	3	
33Ala	47	6	41	8	24	16	17	47	31 34 35 36 47 50 51 124 125 128
34Leu	110	36	74	21	48	27	26	110	28 29 31 33 35 36 37 38 42 50 51 83 86 87 90 91 124 128
35Asn	55	15	40	20	38	18	2	55	31 33 34 36 37 38 39 83
36Tyr	74	17	57	16	30	14	27	74	33 34 35 37 38 39 41 42 83 86 128
37Val	83	15	68	18	53	35	15	83	34 35 36 38 39 40 41 42 43 47 79 86
38Ser	59	13	46	14	28	14	18	59	34 35 36 37 39 40 42 78 79 81 82 83 85 86
39Asn	34	12	22	9	20	11	2	34	35 36 37 38 40 41 80 86
40Gly	27	6	21	11	17	6	4	27	37 38 39 41 42 79 81
41Gln	82	32	50	21	39	18	11	82	36 37 39 40 42 43 79
42Leu	105	38	67	24	31	7	36	105	34 36 37 38 40 41 43 46 47 48 86 128
43Ser	48	11	37	17	35	18	2	48	37 41 42 44 45 46 47 79
44Ser	32	8	24	9	17	8	7	32	43 45 47 48 49 71 74
45Ser	20	7	13	6	12	6	1	20	43 44 46 48 49 67
46Asp	60	12	48	16	39	23	9	60	42 43 45 47 48 49 50 127
47Val	85	22	63	22	48	26	15	85	33 37 42 43 44 46 48 49 50 51 74 79 86 124 128
48Ala	59	6	53	21	31	10	22	59	42 44 45 46 47 49 51 52 66 67 70 71 74 81 86
49Ser	52	12	40	14	27	13	13	52	44 45 46 47 48 50 51 52 64 67 74 127
50Thr	73	10	63	23	41	18	22	73	33 34 46 47 49 51 52 53 54 120 123 124 127

Table A5. Continued

Seq Element	Total	Intra Residue	Inter Residue	Sequential	Short Range	Short Non-seq	Long Range	Intra Chain	Contacted Residues
51Val	104	17	87	31	48	17	39	104	33 34 47 48 49 50 52 53 54 55 66 69 70 74 90 116 120 124 127
52Ala	72	3	69	18	38	20	31	72	48 49 50 51 53 55 56 61 63 65 66 67 68 70 90 97
53Arg	94	53	41	21	38	17	3	94	50 51 52 54 55 56 57 63 68
54Ala	51	6	45	25	39	14	6	51	50 51 53 55 56 57 58 59 119 120 123
55Val	78	22	56	18	37	19	19	78	51 52 53 54 56 58 59 66 70 94 97 98 116 120
56Ala	81	5	76	17	37	20	39	81	52 53 54 55 57 58 59 60 61 62 63 65 66 94 97 116 120
57Gln	45	11	34	18	31	13	3	45	53 54 56 58 59 60 61 63 116
58Ser	41	7	34	19	30	11	4	41	54 55 56 57 59 60 66 119
59Leu	76	29	47	18	28	10	19	76	54 55 56 57 58 60 66 94 112 115 116 119 120
60Gly	33	9	24	16	23	7	1	33	56 57 58 59 61 66
61Leu	103	35	68	22	26	4	42	103	52 56 57 60 62 63 65 66 69 94 97 98 102 103 104 105
62Ser	34	7	27	21	26	5	1	34	56 61 63 64 65 66
63Gln	56	17	39	12	21	9	18	56	52 53 56 57 61 62 64 65 66 67 68
64Gly	20	5	15	9	13	4	2	20	49 62 63 65 67
65Ser	47	15	32	14	28	14	4	47	52 56 61 62 63 64 66 67 68 69
66Val	100	16	84	19	36	17	48	100	48 51 52 55 56 58 59 60 61 62 63 65 67 68 69 70 74 90 97 98 120
67Gln	83	29	54	19	40	21	14	83	45 48 49 52 63 64 65 66 68 69 70 71
68Asn	46	13	33	18	28	10	5	46	52 53 63 65 66 67 69 70 71
69Ile	114	30	84	23	53	30	31	114	51 61 65 66 67 68 70 71 72 73 74 90 91 93 94 96 97 98
70Met	98	27	71	21	40	19	31	98	48 51 52 55 66 67 68 69 71 73 74 89 90 93

Table A5. Continued

Seq Element	Total	Intra Residue	Inter Residue	Sequential	Short Range	Short Non-seq	Long Range	Intra Chain	Contacted Residues
71Ser	38	5	33	12	29	17	4	38	44 48 67 68 69 70 72 73 74
72Gln	33	17	16	9	16	7	0	33	69 71 73 74
73Gln	71	24	47	17	32	15	15	71	69 70 71 72 74 75 76 89 93
74Leu	100	26	74	22	38	16	36	100	44 47 48 49 51 66 69 70 71 72 73 75 76 77 78 81 86 89 90 93 127
75Ser	22	10	12	11	12	1	0	22	73 74 76
76Ser	25	8	17	11	16	5	1	25	73 74 75 77 78 85
77Ile	63	30	33	23	29	6	4	63	74 76 78 81 85 89 93
78Gly	32	8	24	19	23	4	1	32	38 74 76 77 79 81
79Ser	33	13	20	7	8	1	12	33	37 38 40 41 43 47 78 80 81
80Gly	12	2	10	9	9	0	1	12	39 79 81
81Ala	34	6	28	13	20	7	8	34	38 40 48 74 77 78 79 80 82 85 86
82Ser	35	12	23	13	21	8	2	35	38 81 83 84 85 86
83Thr	42	11	31	7	24	17	7	42	34 35 36 38 82 85 86 87
84Ser	6	2	4	1	3	2	1	6	29 82 85 87
85Ser	29	4	25	10	21	11	4	29	38 76 77 81 82 83 84 86 87 88
86Leu	118	39	79	14	43	29	36	118	28 34 36 37 38 39 42 47 48 74 81 82 83 85 87 88 89 90
87Ser	22	3	19	10	15	5	4	22	28 29 34 83 84 85 86 88 89 91
88Gln	56	20	36	15	28	13	8	56	21 25 28 29 85 86 87 89 90 91 92
89Ala	52	5	47	16	36	20	11	52	21 70 73 74 77 86 87 88 90 91 92 93
90Ile	110	33	77	18	34	16	43	110	24 25 28 34 51 52 66 69 70 74 86 88 89 91 92 93 94

Table A5. Continued

Seq Element	Total	Intra Residue	Inter Residue	Sequential	Short Range	Short Non-seq	Long Range	Intra Chain	Contacted Residues
91Ala	70	6	64	19	38	19	26	70	20 21 24 25 26 28 34 69 87 88 89 90 92 93 94 95 113
92Asn	53	17	36	17	31	14	5	53	21 25 28 88 89 90 91 93 94 95
93Ala	77	6	71	20	45	25	26	77	69 70 73 74 77 89 90 91 92 94 95 96 97
94Val	82	15	67	20	44	24	23	82	20 21 24 28 55 56 59 61 69 90 91 92 93 95 96 97 98 112 113 120
95Ser	53	10	43	16	29	13	14	53	17 20 21 91 92 93 94 96 97 98 99 100
96Ser	34	5	29	10	25	15	4	34	69 93 94 95 97 98 99
97Ala	50	6	44	14	25	11	19	50	52 55 56 61 66 69 93 94 95 96 98 99
98Val	96	20	76	23	39	16	37	96	15 20 55 61 66 69 94 95 96 97 99 100 101 102 103 104 105 113
99Gln	52	17	35	20	30	10	5	52	20 95 96 97 98 100 101 102 104
100Gly	28	3	25	13	18	5	7	28	17 19 20 95 98 99 101 103 104
101Ser	38	11	27	13	27	14	0	38	98 99 100 102 103 104
102Gln	32	11	21	15	20	5	1	32	61 98 99 101 103 104
103Ala	35	6	29	16	21	5	8	35	61 98 100 101 102 104 105
104Ala	51	5	46	15	23	8	23	51	15 20 61 98 99 100 101 102 103 105 109 112
105Ala	29	6	23	7	19	12	4	29	61 98 103 104 108 112
106Pro	8	0	8	2	5	3	3	8	15 107 108 109
107Gly	25	6	19	9	16	7	3	25	15 19 20 106 108 109 110
108Gln	63	25	38	13	37	24	1	63	15 105 106 107 109 110 111 112
109Glu	77	21	56	15	30	15	26	77	15 16 19 20 104 106 107 108 110 111 112 113
110Gln	69	26	43	14	23	9	20	69	15 19 20 23 26 107 108 109 111 112 113 114

Table A5. Continued

Seq Element	Total	Intra Residue	Inter Residue	Sequential	Short Range	Short Non-seq	Long Range	Intra Chain	Contacted Residues
111Ser	34	3	31	16	29	13	2	34	19 23 108 109 110 112 113 114 115
112Ile	90	20	70	24	51	27	19	90	15 20 24 59 94 104 105 108 109 110 111 113 114 115 116 118
113Ala	62	6	56	17	34	17	22	62	19 20 23 24 25 91 94 98 109 110 111 112 114 115 116 117 119
114Gln	40	10	30	10	28	18	2	40	110 111 112 113 115 116 117 118 121
115Arg	78	41	37	21	36	15	1	78	59 111 112 113 114 116 117 118
116Val	84	16	68	23	48	25	20	84	20 24 31 51 55 56 57 59 112 113 114 115 117 118 119 120 121
117Asn	66	21	45	21	40	19	5	66	24 27 113 114 115 116 118 119 120 121
118Ser	44	10	34	17	32	15	2	44	24 112 114 115 116 117 119 120 121
119Ala	51	6	45	20	38	18	7	51	54 58 59 113 116 117 118 120 121 122 123 125
120Ile	98	20	78	34	61	27	17	98	24 28 31 50 51 54 55 56 59 66 94 116 117 118 119 121 122 123 124
121Ser	60	10	50	22	37	15	13	60	28 31 114 116 117 118 119 120 122 123 124
122Ser	29	8	21	9	21	12	0	29	119 120 121 123 125 126
123Ala	40	5	35	11	31	20	4	40	50 54 119 120 121 122 124 125 126 127
124Phe	75	15	60	13	35	22	25	75	31 33 34 47 50 51 120 121 123 125 126 127 128
125Ala	41	6	35	17	32	15	3	41	31 33 119 122 123 124 126 127 128 129
126Gln	39	12	27	16	27	11	0	39	122 123 124 125 127 129
127Leu	110	44	66	16	35	19	31	110	46 49 50 51 74 123 124 125 126 128 130 131
128Ile	101	36	65	26	48	22	17	101	31 33 34 36 42 47 124 125 127 129 130 131 132 133
129Ser	37	12	25	21	25	4	0	37	125 126 128 130 131
130Gln	40	20	20	13	20	7	0	40	127 128 129 131 132 133

Table A5. Continued

Seq Element	Total	Intra Residue	Inter Residue	Sequential	Short Range	Short Non-seq	Long Range	Intra Chain	Contacted Residues
131Arg	50	26	24	13	24	11	0	50	127 128 129 130 132 133
132Thr	26	7	19	12	19	7	0	26	128 130 131 133
133Ala	14	4	10	7	9	2	1	14	128 130 131 132
134Pro	3	0	3	1	3	2	0	3	135 136 137
135Ala	21	3	18	2	18	16	0	21	134 136 137
136Pro	3	0	3	2	3	1	0	3	134 135 137
137Ala	21	3	18	1	18	17	0	21	134 135 136
139Arg	1	1	0	0	0	0	0	1	
141Arg	2	1	1	0	1	1	0	2	143
143Ala	2	1	1	0	1	1	0	2	141

Table A6. Convergence of phi and psi angles in ensemble structures.

		S(phi)	S(psi)			S(phi)	S(psi)
MET	1	0.463	0.039	LEU	31	0.995	0.966
GLY	2	0.447	0.14	ASN	32	0.999	0.993
HIS	3	0.896	0.914	ALA	33	0.994	0.993
HIS	4	0.922	0.956	LEU	34	0.996	0.998
HIS	5	0.843	0.927	ASN	35	0.997	0.998
HIS	6	0.825	0.901	TYR	36	0.997	0.999
HIS	7	0.874	0.843	VAL	37	0.999	0.998
HIS	8	0.217	0.375	SER	38	0.998	0.998
SER	9	0.439	0.199	ASN	39	0.998	0.985
SER	10	0.927	0.888	GLY	40	0.986	0.992
ALA	11	0.97	0.967	GLN	41	0.994	0.986
PRO	12	0.993	0.899	LEU	42	0.979	0.996
VAL	13	0.914	0.958	SER	43	0.999	0.998
TYR	14	0.844	0.905	SER	44	0.998	0.998
PHE	15	0.96	0.993	SER	45	0.998	0.996
ASN	16	0.826	0.982	ASP	46	0.998	0.998
SER	17	0.998	0.996	VAL	47	0.999	0.999
GLN	18	0.999	1	ALA	48	0.999	0.995
THR	19	0.997	0.998	SER	49	0.995	0.998
LEU	20	0.999	0.999	THR	50	0.998	0.998
THR	21	0.998	0.998	VAL	51	0.998	0.999
ASN	22	0.997	0.999	ALA	52	0.999	0.997
ASN	23	0.999	0.995	ARG	53	0.997	0.998
LEU	24	0.997	0.998	ALA	54	0.997	1
ALA	25	0.999	0.998	VAL	55	0.998	0.997
SER	26	0.999	0.999	ALA	56	0.997	0.998
SER	27	0.998	0.998	GLN	57	0.999	0.996
LEU	28	0.998	0.997	SER	58	0.996	0.997
GLN	29	0.993	0.989	LEU	59	0.997	0.999
SER	30	0.989	0.993	GLY	60	0.998	0.962

Table A6. Continued

		S(phi)	S(psi)			S(phi)	S(psi)
LEU	61	0.956	0.998	ALA	91	0.999	0.999
SER	62	0.998	1	ASN	92	0.999	0.996
GLN	63	0.999	0.999	ALA	93	0.996	0.998
GLY	64	0.999	0.996	VAL	94	0.998	0.998
SER	65	0.998	0.999	SER	95	0.998	0.996
VAL	66	0.999	0.999	SER	96	0.998	0.997
GLN	67	0.999	0.996	ALA	97	0.994	0.998
ASN	68	0.997	0.998	VAL	98	0.996	0.997
ILE	69	0.999	0.998	GLN	99	0.996	0.981
MET	70	0.997	0.985	GLY	100	0.939	0.936
SER	71	0.987	0.997	SER	101	0.999	1
GLN	72	0.993	0.977	GLN	102	0.959	0.988
GLN	73	0.982	0.998	ALA	103	0.97	0.98
LEU	74	0.993	0.995	ALA	104	0.965	0.962
SER	75	0.997	0.996	ALA	105	0.971	0.984
SER	76	0.996	0.997	PRO	106	0.998	0.992
ILE	77	0.993	1	GLY	107	0.988	0.985
GLY	78	0.961	0.9	GLN	108	0.986	0.991
SER	79	0.956	0.379	GLU	109	0.993	0.99
GLY	80	0.317	0.944	GLN	110	0.997	0.979
ALA	81	0.987	0.981	SER	111	0.979	0.999
SER	82	0.997	0.999	ILE	112	0.998	0.999
THR	83	0.999	0.999	ALA	113	0.997	0.992
SER	84	0.999	0.998	GLN	114	0.989	0.998
SER	85	0.998	0.999	ARG	115	0.998	0.999
LEU	86	0.999	0.999	VAL	116	0.999	1
SER	87	0.999	0.996	ASN	117	0.999	0.999
GLN	88	0.998	0.997	SER	118	0.999	0.997
ALA	89	0.998	0.998	ALA	119	0.998	1
ILE	90	0.999	0.999	ILE	120	1	1

Table A6. Continued

		S(phi)	S(psi)
SER	121	1	0.998
SER	122	0.997	0.998
ALA	123	0.999	0.998
PHE	124	0.999	0.999
ALA	125	0.999	0.999
GLN	126	0.998	0.998
LEU	127	0.999	0.998
ILE	128	0.998	0.993
SER	129	0.998	0.997
GLN	130	0.998	0.994
ARG	131	0.97	0.947
THR	132	0.781	0.607
ALA	133	0.958	0.885
PRO	134	0.977	0.771
ALA	135	0.813	0.901
PRO	136	0.973	0.882
ALA	137	0.655	0.861
PRO	138	0.961	0.871
ARG	139	0.59	0.858
PRO	140	0.969	0.924
ARG	141	0.957	0.979
PRO	142	0.985	0.9
ALA	143	0.392	0.113
SER	144	0.77	0.4

Copyright Permission Letters

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 03, 2022

This Agreement between Dalhousie University -- Jeffrey Simmons ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number 5361501351322

License date Aug 03, 2022

Licensed ContentPublisherElsevier

Licensed ContentPublicationMaterials Today

Licensed Content Title Decoding the secrets of spider silk

Licensed Content Author Lukas Eisoldt,Andrew Smith,Thomas Scheibel

Licensed Content Date Mar 1, 2011

Licensed Content Volume 14

Licensed Content Issue 3

Licensed Content Pages 7

Start Page	80
End Page	86
Type of Use	reuse in a thesis/dissertation Portion
figures/tables/illustrations	
figures/tables/illustrationsNumber of	1
Format	both print and electronic
Elsevier article?	
Will you be translating?	No
Title	From Soluble Protein to Anchoring Filament: Understanding the Structural and Mechanical Foundations of Pyriform Spider Silk
Institution name	Dalhousie University
Expected presentation date	Aug 2022
Portions	Figure 1
Requestor Location	Dalhousie University 136 Coronation Avenue Sir Charles Tupper Medical Building, PO Halifax, NS B3H 4R2 Canada Attn: Dalhousie University

Publisher Tax ID GB 494 6272 12

Total 0.00 CAD

Terms and Conditions
Are you the author of this No

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject tothe terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) hasappeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for whichpermission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may bealtered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier’s permissions helpdesk here). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance,please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in thecombination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensedimmediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensedmaterial.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, andtheir respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed,assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writingsigned by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in anypurchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions describedin this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peerreviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes authorincorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately via their non-commercial person
    - homepage or blog
    - by updating a preprint in arXiv or RePEc with the accepted manuscript via their
    - research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group directly by
    - providing copies to their students or to research collaborators for their personal use
    - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period via non-commercial hosting platforms such as their institutional repository via commercial sites with which Elsevier has an agreement In all cases accepted manuscripts should:
    - link to the formal publication via its DOI bear a CC-
    - BY-NC-ND license - this is easy to do
    - if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published

commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Aug 03, 2022

This Agreement between Dalhousie University -- Jeffrey Simmons ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5361520135549
License date	Aug 03, 2022
Licensed ContentPublisher	Springer Nature
Licensed Content Publication	Nature Chemical Biology
Licensed Content Title	Toward spinning artificial spider silk
Licensed Content Author	Anna Rising et al
Licensed Content Date	Apr 17, 2015
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic

Portion	figures/tables/illustrations
Number offigures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	30 - 99
Author of this Springer Nature content	no
Title	From Soluble Protein to Anchoring Filament: Understanding the Structural and Mechanical Foundations of Pyriform Spider Silk
Institution name	Dalhousie University
Expected presentation date	Aug 2022
Portions	Figure 2
Requestor Location	Dalhousie University 136 Coronation Avenue Sir Charles Tupper Medical Building, PO Halifax, NS B3H 4R2 Canada Attn: Dalhousie University
Total	0.00 CAD
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee.
Please apply to
Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

	Duration of Licence
Post on a website	12 months
Presentations	12 months
Scope of Licence	
Books and journals Lifetime of the edition in the language purchased	

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:***Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**

Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jul 29, 2022

This Agreement between Dalhousie University -- Jeffrey Simmons ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5358360184397
License date	Jul 29, 2022
Licensed ContentPublisher	Springer Nature
Licensed ContentPublication	Nature
Licensed Content Title	Liquid crystalline spinning of spider silk
Licensed Content Author	Fritz Vollrath et al
Licensed Content Date	Mar 29, 2001
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic

Portion	figures/tables/illustrations
Number offigures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	30 - 99
Author of this Springer Nature content	no
Title	From Soluble Protein to Anchoring Filament: Understanding the Structural and Mechanical Foundations of Pyriform Spider Silk
Institution name	Dalhousie University
Expected presentation date	Aug 2022
Portions	Figure 1
Requestor Location	Dalhousie University 136 Coronation Avenue Sir Charles Tupper Medical Building, PO Halifax, NS B3H 4R2 Canada Attn: Dalhousie University
Total	0.00 CAD
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and

completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.
1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee.
Please apply to
Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

	Duration of Licence
Post on a website	12 months
Presentations	12 months
Scope of Licence	
Books and journals Lifetime of the edition in the language purchased	

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:*** Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication) **Other**

Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 03, 2022

This Agreement between Dalhousie University -- Jeffrey Simmons ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5361490108633
License date	Aug 03, 2022
Licensed Content Publisher	Elsevier
Licensed Content Publication	Journal of Molecular Biology
Licensed Content Title	Diversity of Molecular Transformations Involved in the Formation of Spider Silks
Licensed Content Author	Thierry Lefèvre,Simon Boudreault,Conrad Cloutier,MichelPézolet
Licensed Content Date	Jan 7, 2011
Licensed Content Volume	405
Licensed Content Issue	1
Licensed Content Pages	16
Start Page	238

End Page	253
Type of Use	reuse in a thesis/dissertation Portion
figures/tables/illustrations	
figures/tables/illustrations	Number of 1
Format	both print and electronic
Elsevier article?	
Will you be translating?	No
Title	From Soluble Protein to Anchoring Filament: Understanding the Structural and Mechanical Foundations of Pyriform Spider Silk
Institution name	Dalhousie University
Expected presentation date	Aug 2022
Portions	Figure 9
Requestor Location	Dalhousie University 136 Coronation Avenue Sir Charles Tupper Medical Building, PO Halifax, NS B3H 4R2 Canada Attn: Dalhousie University
Publisher Tax ID	GB 494 6272 12

Total 0.00 CAD

Terms and Conditions
Are you the author of this
No

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject tothe terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) hasappeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for whichpermission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may bealtered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier's permissions helpdesk here). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance,please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in thecombination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensedimmediately upon issuance of the license at the end of the licensing process for the

transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensedmaterial.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, andtheir respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed,assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writingsigned by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in anypurchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions describedin this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may

only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peerreviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes authorincorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately via their non-commercial person
    - homepage or blog
    - by updating a preprint in arXiv or RePEc with the accepted manuscript via their
    - research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group directly by
    - providing copies to their students or to research collaborators for their personal use
    - for private scholarly sharing as part of an invitation-only work group on

commercial sites with which Elsevier has an agreement

- After the embargo period via non-commercial hosting platforms such as their institutional repository via commercial sites with which Elsevier has an agreement In all cases accepted manuscripts should:

- link to the formal publication via its DOI bear a CC-
- BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

20. **Other Conditions**:

v1.10

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Aug 03, 2022

This Agreement between Dalhousie University -- Jeffrey Simmons ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number 5361490306512

License date	Aug 03, 2022
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Angewandte Chemie International Edition
Licensed Content Title	Spider Silk: From Soluble Protein to Extraordinary Fiber
Licensed Content Author	Thomas Scheibel, David Keerl, Markus Heim
Licensed Content Date	Apr 28, 2009

Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	From Soluble Protein to Anchoring Filament: Understanding the Structural and Mechanical Foundations of Pyriform Spider Silk
Institution name	Dalhousie University
Expected	
Licensed Content Volume	48
Licensed Content Issue	20
Licensed Content Pages	13
presentation date	Aug 2022
Portions	Figure 4
Requestor Location	Dalhousie University 136 Coronation Avenue Sir Charles Tupper Medical Building, PO Halifax, NS B3H 4R2 Canada Attn: Dalhousie University

Publisher Tax ID	EU826007151
Total	0.00 CAD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a standalone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to**

include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts, You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE,

BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.
- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal

action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library

http://olabout.wiley.com/WileyCDA/Section/id-410895.html **Other**

Terms and Conditions:

v1.10 Last updated September 2015

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Aug 03, 2022

This Agreement between Dalhousie University -- Jeffrey Simmons ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5361500560326
License date	Aug 03, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Communications
Licensed Content Title	Cobweb-weaving spiders produce different attachment discs for locomotion and prey capture
Licensed Content Author	Vasav Sahni et al
Licensed Content Date	Oct 2, 2012
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations

Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	30 - 99
Author of this Springer Nature content	no
Title	From Soluble Protein to Anchoring Filament: Understanding the Structural and Mechanical Foundations of Pyriform Spider Silk
Institution name	Dalhousie University
Expected presentation date	Aug 2022
Portions	Figure 1
Requestor Location	Dalhousie University 136 Coronation Avenue Sir Charles Tupper Medical Building, PO Halifax, NS B3H 4R2 Canada Attn: Dalhousie University
Total	0.00 CAD
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and

completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee.
Please apply to
Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

	Duration of Licence
Post on a website	12 months
Presentations	12 months
Scope of Licence	
Books and journals Lifetime of the edition in the language purchased	

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:***Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication) **Other**

Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 03, 2022

This Agreement between Dalhousie University -- Jeffrey Simmons ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5361500683110
License date	Aug 03, 2022
Licensed Content Publisher	Elsevier
Licensed Content Publication	International Journal of Biological Macromolecules
Licensed Content Title	The molecular structure of novel pyriform spidroin (PySp2) reveals extremely complex central repetitive region
Licensed Content Author	Kangkang Wang,Rui Wen,Shuangzhu Wang,Luyang Tian,JunhuaXiao,Qing Meng
Licensed Content Date	Feb 15, 2020
Licensed Content Volume	145
Licensed Content Issue	n/a
Licensed Content Pages	8
Start Page	437

End Page	444
Type of Use	reuse in a thesis/dissertation Portion
figures/tables/illustrations	
figures/tables/illustrations	Number of 1
Format	both print and electronic
Elsevier article?	
Will you be translating?	No
Title	From Soluble Protein to Anchoring Filament: Understanding the Structural and Mechanical Foundations of Pyriform Spider Silk
Institution name	Dalhousie University
Expected presentation date	Aug 2022
Portions	Figure 1
Requestor Location	Dalhousie University 136 Coronation Avenue Sir Charles Tupper Medical Building, PO Halifax, NS B3H 4R2 Canada Attn: Dalhousie University
Publisher Tax ID	GB 494 6272 12

Total 0.00 CAD

Terms and Conditions
Are you the author of this
No

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject tothe terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) hasappeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for whichpermission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may bealtered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier's permissions helpdesk here). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance,please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in thecombination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensedimmediately upon issuance of the license at the end of the licensing process for the

transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensedmaterial.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, andtheir respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed,assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writingsigned by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in anypurchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions describedin this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may

only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peerreviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes authorincorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately via their non-commercial person
  - homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript via their
  - research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group directly by
  - providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on

commercial sites with which Elsevier has an agreement

- After the embargo period via non-commercial hosting platforms such as their institutional repository via commercial sites with which Elsevier has an agreement In all cases accepted manuscripts should:

- link to the formal publication via its DOI bear a CC-
- BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

20. **Other Conditions**:

v1.10

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jul 29, 2022

This Agreement between Dalhousie University -- Jeffrey Simmons ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5358360670484
License date	Jul 29, 2022
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Journal of Biomolecular NMR
Licensed Content Title	Conformationally selective multidimensional chemical shift ranges in proteins from a PACSY database purged using intrinsic quality criteria
Licensed Content Author	Keith J. Fritzsching et al
Licensed Content Date	Jan 19, 2016
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations

Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	30 - 99
Author of this Springer Nature content	no
Title	From Soluble Protein to Anchoring Filament: Understanding the Structural and Mechanical Foundations of Pyriform Spider Silk
Institution name	Dalhousie University
Expected presentation date	Aug 2022
Portions	Figure 10A
Requestor Location	Dalhousie University 136 Coronation Avenue Sir Charles Tupper Medical Building, PO Halifax, NS B3H 4R2 Canada Attn: Dalhousie University
Total	0.00 USD
Terms and Conditions	

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee.
Please apply to
Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

	Duration of Licence
Post on a website	12 months
Presentations	12 months

Scope of Licence
Books and journals Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE

MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. ***BOOKS ONLY:***Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION**

(Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

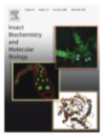
Other Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Recombinant Pyriform Silk Fiber Mechanics Are Modulated by Wet-Spinning Conditions

Author: Jeffrey R. Simmons, Lingling Xu, Jan K. Rainey
Publication: ACS Biomaterials Science & Engineering
Publisher: American Chemical Society
Date: Oct 1, 2019



Complete gene sequence of spider attachment silk protein (PySp1) reveals novel linker regions and extreme repeat homogenization

Author: Ro Crystal Chaw,Christopher A. Saski,Cheryl Y. Hayashi
Publication: Insect Biochemistry and Molecular Biology
Publisher: Elsevier
Date: February 2017