12-2016

Creating Biomaterials from Plant-derived Recombinant Spider Silk-like Proteins

Congyue Peng
Clemson University, congyup@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Recommended Citation
https://tigerprints.clemson.edu/all_dissertations/1824

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.
CREATING BIOMATERIALS FROM PLANT-DERIVED RECOMBINANT SPIDER SILK-LIKE PROTEINS

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Biochemistry and Molecular Biology

by
Congyue Peng
December 2015

Accepted by:
Dr. William R. Marcotte Jr., Committee Chair
Dr. Kerry Smith
Dr. Hong Luo
Dr. Michael Sehorn
ABSTRACT

Biomimetic fibers that resemble the structure and properties of the versatile yet tough spider dragline silk are in high demand. The assembled building blocks of major ampullate dragline silk sequences from *Nephila clavipes* [Spidroin 1 (Sp1) and Spidroin 2 (Sp2)] are expressed in transgenic *Nicotiana tabacum*. The plant derived spidroin analogs (mini-spidoins) consist of native Sp1 or Sp 2 N- and C-termini flanking 8, 16, or 32 copies of their respective consensus repeat domains. Gene insertion and RNA transcription were confirmed by PCR and reverse-transcriptase PCR, respectively. Produced mini-spidoins were purified from tobacco using affinity chromatography techniques. After a freeze-drying period, mini-spidoins formed viscous fluids.

When purified proteins were treated with acid, cross-linked by glutaraldehyde and diluted in phosphate buffer (pH 7), the mini-spidoins formed a thin film at the layer interface of a counter-ion gellan gum solution. The film was pulled into fibers that displayed auto-fluorescent across a broad range of wavelengths. Fiber mechanical performance was recorded by single displacement controlled tensile test. A method with potential to continuously pull fiber was developed. Fine fibers were produced with more uniform diameter and higher tensile performance.
DEDICATION

This dissertation is dedicated to the unconditional love and support from my family.
ACKNOWLEDGMENTS

I would like to thank Dr. William R. Marcotte Jr., my advisor, my committee members Dr. Kerry Smith, Dr. Luo Hong, and Dr. Michael Sehorn for their advice and guidance. I would like to thank Dr. Michael Ellison of the Department of Material Science, Dr. John Desjardins of the Department of Bioengineering and Dr. Todd Lyda of the Department of Genetics and Biochemistry of Clemson University and Dr. Florence Teulé of the Department of Zoology and Physiology of University of Wyoming for their generous help on fiber tensile testing and fiber formation; Dr. Nishanth Tharayil from the Multi-User Analytical Lab of Clemson University for his generous help on Mass Spectrometry analysis; Dr. Indu Maiti at the Kentucky Tobacco Research and Development Center for providing us the tobacco expression plasmid; Dr. Terri Bruce for her sincere help on microscopy and Krystal Cadle and Allison Nelson for their support.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>SPIDER SILKS: FROM BIOTECHNOLOGY TO BIOMATERIALS</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INTRODUCTION .............................................</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Types of spider silks in an orb-weaving spider</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>web, the respective secretory glands and protein composition</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Orb-weaving spider major ampullate gland morphology and physiology</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Molecular biology of major ampullate spidroin coding sequences</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Transcription and translation</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Post-translational modifications</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Major ampullate spidroin protein sequence, structure and folding</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>The repetitive domain</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>The non-repetitive C-terminal domain</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>The non-repetitive N-terminal domain</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Importance of proline to the structure and function of MaSp2</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Hydrogen bond and supercontraction</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Spider silk fiber structure and self-assembly</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Recombinant production of spider silk-like Proteins and fiber assembly</td>
<td>45</td>
</tr>
</tbody>
</table>
Table of Contents (Continued)

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application of spider silk fibers and spider</td>
<td></td>
</tr>
<tr>
<td>silk fiber analogs</td>
<td>55</td>
</tr>
<tr>
<td>Summary</td>
<td>58</td>
</tr>
<tr>
<td>References</td>
<td>61</td>
</tr>
</tbody>
</table>

II. SPIDER SILK-LIKE PROTEINS DERIVED FROM
| TRANSGENIC NICOTIANA TABACUM                                         | 95   |
| Introduction                                                          | 96   |
| Materials and methods                                                | 100  |
| Results                                                               | 105  |
| Discussion                                                            | 108  |
| References                                                            | 121  |

III. POLYELECTROLYTE FIBER ASSEMBLY OF PLANT
| DERIVED SPIDER SILK-LIKE PROTEIN                                      | 129  |
| Introduction                                                          | 130  |
| Materials and methods                                                | 133  |
| Results                                                               | 136  |
| Discussion                                                            | 142  |
| Future study                                                         | 150  |
| Conclusion                                                           | 151  |
| References                                                            | 163  |

IV. ENZYME ACTIVITY AND ITS ROLE IN PROTEIN
| CROSS-LINKING OF A SPIDER AMPULLATE GLAND SPECIFIC ENZYME            | 172  |
| Introduction                                                          | 173  |
| Materials and methods                                                | 176  |
| Results                                                               | 180  |
| Discussion                                                            | 182  |
| References                                                            | 194  |

V. FUTURE STUDY                                                        | 201  |
| References                                                            | 207  |

APPENDICES                                                             | 208  |
<table>
<thead>
<tr>
<th></th>
<th>Grant Permission</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Granted permission from American Chemical Society (1)</td>
<td>209</td>
</tr>
<tr>
<td>B</td>
<td>PLOS content license</td>
<td>210</td>
</tr>
<tr>
<td>C</td>
<td>Granted permission from American Chemical Society (2)</td>
<td>211</td>
</tr>
<tr>
<td>D</td>
<td>Granted permission from Elsevier</td>
<td>212</td>
</tr>
<tr>
<td>E</td>
<td>Granted permission from Royal Society of Chemistry</td>
<td>213</td>
</tr>
<tr>
<td>F</td>
<td>Granted permission from Nature Publishing Group (1)</td>
<td>214</td>
</tr>
<tr>
<td>G</td>
<td>Granted permission from Nature Publishing Group (2)</td>
<td>215</td>
</tr>
<tr>
<td>H</td>
<td>The Journal of Biological Chemistry copyright permission</td>
<td>216</td>
</tr>
<tr>
<td>I</td>
<td>Granted permission from John Wiley and Sons</td>
<td>217</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>2.1 Oligonucleotide primers</td>
<td>112</td>
<td></td>
</tr>
<tr>
<td>3.1 Fiber assembly of single protein component</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>3.2 Fiber assembly of two protein component</td>
<td>154</td>
<td></td>
</tr>
<tr>
<td>3.3 Mechanical property of rMaSpR8 and gellan gum fiber</td>
<td>155</td>
<td></td>
</tr>
<tr>
<td>3.4 Mechanical property of chitosan containing rMaSpR8 and gellan gum fiber</td>
<td>156</td>
<td></td>
</tr>
<tr>
<td>3.5 Mechanical property of post-spin stretched chitosan containing rMaSpR8 and gellan gum fiber</td>
<td>157</td>
<td></td>
</tr>
<tr>
<td>3.6 Mechanical property of water submerged and friction force sheared chitosan containing rMaSpR8 gellan fiber</td>
<td>159</td>
<td></td>
</tr>
<tr>
<td>4.1 NsPox protein yield from amylose affinity purification and DEAE-sepharose anion exchange purification</td>
<td>188</td>
<td></td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>1.1</td>
<td>Figure 1.1 Spider silk glands and the respective silks</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Figure 1.2 <em>Nephila clavipes</em> major ampullate gland structure</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Figure 1.3 Spider silks initial conformation in liquid storage form (left) and their final conformation after fiber assembly by Raman spectromicroscopy</td>
<td>5</td>
</tr>
<tr>
<td>1.4</td>
<td>Figure 1.4 Representative stress-strain curve of <em>N. clavipes</em> major ampullate silk</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>Figure 1.5 Diagram of <em>Nephila clavipes</em> major ampullate spidroin domain structure</td>
<td>10</td>
</tr>
<tr>
<td>1.6</td>
<td>Figure 1.6 NMR structure of the C-terminal domain of <em>Araneus diadematus</em> fibroin 3 (ADF-3)</td>
<td>30</td>
</tr>
<tr>
<td>1.7</td>
<td>Figure 1.7 NMR structure of the N-terminal domain dimer of <em>Euprosthenops australis</em></td>
<td>32</td>
</tr>
<tr>
<td>1.8</td>
<td>Figure 1.8 Two component model of spider dragline silk structure under native and supercontracted state</td>
<td>39</td>
</tr>
<tr>
<td>1.9</td>
<td>Figure 1.9 Spider dragline silk multilayer compositions</td>
<td>41</td>
</tr>
<tr>
<td>1.10</td>
<td>Figure 1.10 SEM (a) and AFM (b) imaging of <em>Nephila pilipes</em> dragline silk</td>
<td>42</td>
</tr>
<tr>
<td>1.11</td>
<td>Figure 1.11 Proposed multimeric strand model of spider dragline silk fiber assembly</td>
<td>44</td>
</tr>
<tr>
<td>2.1</td>
<td>Diagrammatic representation of recombinant mini-spidroin domain architecture and purification strategy</td>
<td>113</td>
</tr>
<tr>
<td>2.2</td>
<td>Immuno-detection of recombinant mini-spidroins in crude leaf extracts</td>
<td>114</td>
</tr>
</tbody>
</table>
2.3 Immuno-detection of recombinant mini-spidroin rMaSp2R32 in crude extracts .................................................................115

2.4 Purification of recombinant mini-spidroins from crude leaf extracts ..............................................................................116

2.5 Purified rMaSp1R8 showed various degree of aggregation after freeze-drying ..................................................................117

Sup Fig. 2.1 Immuno-detection of recombinant mini-spidroins in crude leaf extracts of non-transgenic tobacco and transgenic tobacco ...........................................................................................................118

Sup Fig. 2.2 Immuno-detection of recombinant mini-spidroin rMaSp1R16 in crude leaf extracts .................................................................119

Sup fig. 2.3 Manual grinding and blender extraction of rMaSps at various tissue (T) : buffer (E) ratios ...........................................120

3.1 Images of mini-spidroin and gellan gum fiber .........................................159

3.2 Representative stress-strain curve of chitosan containing mini-spidroin and gellan gum fiber ........................................................................160

3.3 Combined fluorescence spectrum 3D volume view of fibers ..............................................................................................161

3.4 Polarized light image of rMaSpR8 gellan gum fibers ....................162

4.1 Recombinant NsPox sequence details .........................................189

4.2 Purification of NsPox by amylose affinity chromatography ..........................................................190

4.3 Purification of NsPox by anionic exchange chromatography .........................................................................................191

4.4 Absorbance vs. time plot of NsPox catalyzed 3, 3’, 5, 5’-tetramethylbezidine (TMB) oxidizatio .............................................192

4.5 Immuno-detection of non-transformed and transformed E. coli BL21 cells ...............................................................................193
CHAPTER ONE

SPIDER SILKS: FROM BIOTECHNOLOGY TO BIOMATERIALS
INTRODUCTION

Biomaterials, synthetic or natural, provide promising approaches for innovative solutions to material engineering, regenerative medicine, drug delivery and much more (Bourzac, 2015; Gould, 2015; Lavine et al., 2012). One source of natural biomaterial is fibrous protein, which covers a broad range of functional proteins such as helical based collagen (Parry et al., 2005), elastin (Mithieux et al., 2005), β-protein based amyloids (Kreplak et al., 2006) and silks (Dicko et al., 2006). Compared to silkworm silk and most synthetic fibers, spider dragline silk which serves as the life-line of an orb-weaving spider is known as the toughest fibrous material (Gosline et al., 1999). Other impressive features include biocompatibility, biodegradability and high thermal conductivity (Allmeling et al., 2006; Huang et al., 2012). The distinct structure and crystal network of dragline silk are central to the physical properties desired in the biomaterial industry (Buehler and Yung, 2010; Woolfson and Ryadnov, 2006). Unlike silkworms that can be farmed, spider farming is difficult and unproductive due to the cannibalistic nature of the spiders (Nentwig, 2013). Bioengineering recombinant spider silk-like proteins that can be processed into fibrous structures is an alternative solution for mass production of spider silk based materials. Here, I review the physiology of spider silks and the respective secretory glands, the protein sequence and domain structure of the dragline silk, physical properties of dragline silk proteins, current techniques in recombinant spider silk protein
production, protein-to-fiber conversion and advances in medical application of these spider silk-like materials.

**Types of spider silks in an orb-weaving spider web, the respective secretory glands and protein composition**

The web of the orb-weaving spider is well structured (Fig. 1.1). The tenaceous threads named spidroins (spider fibroin) are secreted from epithelial cells and stored in the lumen of the gland.

![Figure 1.1. Spider silk glands and the respective silks.](image)

Spidroins experience elongational flow and dehydration through the gland duct and emerge as a solid fiber (Fig. 1.2). Below is a summary of the seven types of orb-weaving spider silks, their secretory glands, the protein compositions and
protein secondary structure that have been reported (Fig. 1.3). Secondary structures of some of the spider silk proteins contain ordered structures such as oriented β-sheet (β-sheets are aligned and parallel to fiber axis), β-spiral, α-helix or random coils. Details on gland physiology and spidroin biochemistry will be discussed in the next few sections.

**Figure 1.2. Nephila clavipes major ampullate gland structure.** Spidroin folding transition, pH changes and enzyme activity are labeled according to the references listed below. Adapted with permission from Lefèvre, T., Boudreault, S. et al. 2008. Conformational and orientational transformation of silk proteins in the major ampullate gland of *Nephila clavipes* spiders. Copyright (2008) American Chemical Society.

b. Andersson, M., Chen, G. et al. (2014).
Figure 1.3. Spider silks initial conformation in liquid storage form (left) and their final conformation after fiber assembly as determined by Raman spectromicroscopy. Reprinted with permission from Lefèvre, T., Paquet-Mercier, F. et al.. (2011) Structure of silk by Raman spectromicroscopy: From the spinning glands to the fibers. Biopolymers 97, 322-336. Copyright (2011) John Willey and Sons. MA: major ampullate, MI: minor ampullate, Cyl: cylindriform, Ac: Aciniform, Pyr (Pir): Pyriform, Flag: Flagelliform.

Dragline silk is the life line and the radial frame thread on the web of an orb-weaving spider. Dragline silk proteins are secreted from major ampullate gland epithelial cells. Based on the morphology of the secretory droplets, the storage sac of the gland is superficially divided into Zone A and Zone B, (Vollrath and Knight, 1999) and Zone C (Andersson et al., 2013). Two primary protein components that were identified originally in this gland are major ampullate spidroin 1 (MaSp1) (Xu and Lewis, 1990) and major ampullate spidroin 2 (MaSp2) (Hinman and Lewis, 1992). Major ampullate gland spidroins have non-
repetitive N- and C-terminal domains (NTD, CTD), and a highly repetitive central domain (Fig. 1.4).

![Diagram of Nephila clavipes major ampullate spidroin domain structure.](Marcotte, unpublished)

The protein secondary structure is dominated by oriented β-sheets (Rousseau et al., 2009). Multiple isoforms of MaSp1 have been reported. Two long isoforms of MaSp1 have been identified in *Nephila clavipes*, (Gaines and Marcotte, 2008) and a ~40 kDa short isoform of MaSp1 is found in a cob weaving spider *Cyrtophora moluccensis* (Han et al., 2013). A short peptide that reacts to NTD antibody identified from the dissolved silk of *Euprosthenops australis* may also be a putative short isoform (Andersson et al., 2013). Other than major ampullate spidroins, short (~10 kDa) cysteine rich proteins (CRP1-4) have been identified from the gland protein extracts, that are part of the protein complex in the silk and are co-expressed with MaSp1 and MaSp2 (Pham et al., 2014).

Minor ampullate spidroins are produced from minor ampullate glands and form the auxiliary spiral of the web. Minor ampullate spidroins (MiSp1, MiSp2) are composed of N- and C-terminal domains and a non-repetitive linker domain of 83-174 amino acids in length (Gao et al., 2013) that links the NTD and the long
repeat domain. Four distinct units randomly repeat themselves to form the long repeat domain, within which reside two identical interruptive spacer regions (Chen et al., 2012; Colgin and Lewis, 1998). Minor ampullate spidroins predominantly form β-sheet structure during fiber formation (Guinea et al., 2012).

Pyriform glands secrete spidroins that make up the attachment disc (Kovoor and Zylberberg, 1980). The pyriform silk protein, identified from *N. clavipes*, (Perry et al., 2010) has a distinct proline rich motif (proline alternates with one of the amino acids: alanine, leucine, arginine, valine, or isoleucine ), a glutamine and serine rich motif and a non-repetitive C-terminal domain. N-terminal sequences remain uncharacterized.

Flagelliform silk or capture silk is the spiral thread that connects web frames. Produced from the flagelliform or coronate glands (Rodriguez and Candelas, 1995), flagelliform fibroin has the non-repetitive N- and C-terminal domain, flanking a repeat domain that consists of tandem repeats of two glycine rich motifs and a spacer motif rich in acidic amino acids (Hayashi and Lewis, 2001). The flagelliform protein is coated with aqueous glues rich in positively charged amines, diamines, and N-acetylgalactosamine glycoproteins that hydrate the core spidroins (Becker et al., 2003). The glue protein, the strongest natural glue, is secreted from the aggregate glands. Two glue spidroins (ASG1 and ASG2) are transcribed from the complementary strands of one single gene. Both contain a repetitive domain and a non-repeat domain that might be a chitin binding site (Choresh et al., 2009). The repetitive domain of one of the glue
spidroin contains a glycine rich motif interspersed with the distinct motif NVN VN (Vasanthavada et al., 2012).

Pseudoflagelliform silk, a flagelliform silk equivalent, is produced from the pseudoflagelliform gland of cribellate orb-weaving spiders. Instead of being covered with a wet glue protein, pseudoflagelliform silk is coated with dry cribellar fibrils (Blackledge and Hayashi, 2006b).

Prey wrapping aciniform silk is produced from the aciniform gland. The aciniform spidroin 1 (AcSp1) domain structure includes a long tandem repeat domain of a 200 amino acids motif that lacks alanine rich stretches. The repeat motif is different from any other repeat motif in other silk types but the C-terminal domain is distantly related to the C-terminal domain in other silks (Hayashi et al., 2004).

Egg sac silk, produced from the cylindrical or tubliform glands of Nephila clavipes (Candelas et al., 1986), consists of a large protein, TuSp1, (Garb and Hayashi, 2005) with tandem repeats of a 200 amino acid motif that is variable in sequence and a non-repetitive C-terminal domain. TuSp1 predominantly forms oriented β-sheets, α-helixes (Rousseau et al., 2009) and β-spirals (Dicko et al., 2004a).

**Mechanical and physical properties of spider silks**

Spider silks have superior tensile properties which can be assessed by standard tensile testing methods. Tensile testing records the elongation of a fiber
and the force applied to it. The test can be performed by loading a fiber of a certain length onto the loading cell of tensile tester. The upper carriage gradually applies pulling force to the fiber. The applied force per cross-sectional surface area of the fiber is defined as stress. The maximum stress (tenacity) is the stress applied right before the fiber breaking point. Fiber length is increased during pulling. The ratio of elongation length to the original length is defined as strain. The maximum strain at the breaking point describes the extensibility of the fiber. The stress and strain are routinely plotted onto a stress-strain curve where strain is the X-axis and the stress is the Y-axis. Fig 1.5 shows a stress strain curve obtained from *N. clavipes* major ampullate dragline silk (Marcotte, unpublished data). The slope in the linear range of this curve is termed Young’s Modulus. It is an indication of the stiffness of the fiber (Gosline et al., 1999) and a higher Young’s Modulus indicates a stiffer fiber. The overall toughness of the fiber is calculated from the integrated surface area under the stress strain curve and it describes the amount of energy required to break the fiber. The average amount of energy required to break a single strand of dragline silk from spiders of the genus *Araneus* was tested and observed to be 160 MJ/m³, compared to 6 MJ/m³ for high-tensile steel (Gosline et al., 1999). The average overall toughness of *N.clavipes* major ampullate dragline silk reaches 250 MJ/m³ (Marcotte, unpublished data).
Figure 1.5. Representative stress-strain curve of *N. clavipes* major ampullate silk.

In comparing tensile properties of five different silk fibers (dragline, tubuliform, minor ampullate, aciniform and capture spiral from *Argiope argentata*), dragline silk has the most strength but the least extensibility (Blackledge and Hayashi, 2006a). Dragline silks maintain tenacity after years of aging but extensibility and overall toughness start to decline after a year (Agnarsson et al., 2008). Dragline silk responds to elevated humidity with supercontraction, which is similar to the contraction in muscle filaments (Agnarsson et al., 2009b) and is represented as a stress followed by relaxation and regaining of stress after internal hydrogen bond rearrangement (Bell et al., 2002). Dragline silk contracts with a stress of 10 to 140 MPa during supercontraction in relative humidity over 70% (Holland et al., 2012). The rate of supercontraction is humidity dependent (Agnarsson et al., 2009a).
Other than the unprecedented mechanical properties, spider silk has the largest wavespeed range (Mortimer et al., 2014) and it conducts heat as well as copper (Huang et al., 2012). Spider silk fiber propagates light and can be used as optical fiber (Huby et al., 2013) and it distorts positive static charges (Ortega-Jimenez and Dudley, 2013).

The superior mechanical performance of spider dragline silk correlates to the molecular structure of the silk fibers (Gosline et al., 1986). This is influenced by the physiology of the secretory gland, rheology of the liquid crystalline flow, coding genes, translational and post-translational control of spidroin production and spidroin folding. The next few sections in this review will focus on these aspects.

**Orb-weaving spider major ampullate gland morphology and physiology**

Major ampullate dragline silk constitutes the frame work of the web and the safe line of the spider. The ability to withstand energy turbulence caused by the struggling victims without damaging the overall integrity of the web requires dragline silk to be both strong and elastic. Major ampullate dragline silk emerges at the spigot of the major ampullate gland, a relatively large gland in the abdomen of the spiders. Although clustered with other glands, it is relatively easy to isolate (Jeffery et al., 2011). The major ampullate gland has a long tail, an ample sac and tapered zigzagged limbs. A funnel resides between the sac and
proximal limb and a valve resides between the distal limb and the spigot (Fig. 1.2).

The protein content of the spinning dope, spidroin, is mostly secreted in the long, coiled tail. Additional spidroin content is secreted in the sac but is produced at a slower rate. The tail secretory cells are depleted of Golgi body, indicating a direct secretion channel that may allow fast protein production (Bell and Peakall, 1969). Ribosomal materials may be released to the lumen of the tail based on morphological and isotopic labeling studies during forced silking (Bell and Peakall, 1969).

The ample sac functions as a storage region, which is discerned as Zone A and Zone B based on different types and amount of the columnar secretory cells (Vollrath and Knight, 1999). A recent study using Azure blue staining showed the presence of three types of epithelial cells that differentiate the tail and sac region into Zones A, B and C (Andersson et al., 2013). Microscopic images captured additional secretion droplets that may be secreted from the epithelial cells to the lumen at the gland storage sac and proximal duct junction. But, there is no further evidence on what type of materials are in these secretion droplets (Bell and Peakall, 1969). At the end of Zone C, the spinning dope migrates into the narrowed S-shaped duct through a funnel-shaped structure. There has been a debate about whether MaSp2 is secreted in the tail and Zone A and MaSp1 is secreted in Zone B or both spidroins are secreted in tail and both A and B zones. The argument that Zone A secretes the core spidron and
Zone B secretes the coating spidroin is based on the presence of small droplets in the secretions in Zone A but not in Zone B (Vollrath and Knight, 1999) and the amino acid composition differences in Zone A and Zone B (Dicko et al., 2004b). However, an antibody-based study showed that both MaSp1 and MaSp2 are detected by MaSp1- and MaSp2-specific antibody in tail, Zone A and Zone B (Sponner et al., 2005a). Since both MaSp1 and MaSp2 can be detected in all three regions, it is hard to argue that MaSp1 is added to the spinning dope from Zone B. Sponner and colleagues found two high molecular weight proteins from the Zone B protein extracts that cross-react to neither anti-MaSp1 nor anti-MaSp2 serum. These two unidentified proteins may possibly be added to the silk from Zone B as a coating surface.

The limbs of the narrowed spinning duct are numbered 1 and 2 from the sac end and a draw down taper (Limb 3) connects the valve and spigot. The narrowed duct provides physical shear forces that may be required in fiber assembly (Lefèvre et al., 2008) and possibly another coating material to the fiber (Vollrath and Knight, 1999).

Many studies have focused on elucidating the microenvironment subtleties of the ampullate gland that facilitates fiber formation such as proton concentration or pH at each region, ion composition, the activities of special enzymes, spidroin composition of the spinning dope and other polymer matrix, etc. As more biochemical information of the fiber producing apparatus is discovered, a greater understanding of the natural fiber assembly will occur.
The proton concentration in the lumen of the gland is progressively increased along the path. This is indicated by a pH drop from 7.3 in sac Zone A to 6.3 in the proximal duct (Limb 1, 0.5 mm from funnel) (Dicko et al., 2004a). The pH of the distal duct (Limb 3) is reported to be 6.3 but it could be lower based on the uptake speed of neutral Red dye staining and the limitation of pH detectability in the narrowed duct (Dicko et al., 2004b; Knight and Vollrath, 2001; Vollrath et al., 1998). This was confirmed in a later study using ion selective microelectrodes that demonstrate pH is as low as 5.9 at proximal limb 2 and drops to 5.7 at distal limb 2 (Andersson et al., 2014).

Energy dispersive X-ray microanalysis was used to determine the concentration of diffusible ions along the duct as a percentage of total atoms recorded (Knight and Vollrath, 2001). From the proximal duct down to the end of the duct, an estimated ~20 mm distance, sodium and chlorine percentages are progressively reduced from 1.2% to 0.1% and 1% to 0.5%, respectively. Potassium, sulfur and phosphor percentages are increased from 0.1 to 0.6, 0.1 to 0.35 and 0.05 to 0.9, respectively. The slight chlorine composition decrease corresponds to the slight decrease in the total concentration of sodium and potassium, indicating that chlorine may be the counter ion to both sodium and potassium. The gradual replacement of chaotropic potassium ions for sodium in the spinning duct may promote water removal from spidroin molecules and contribute to the dehydration process known to be important in fiber assembly.
The distal duct can be quickly stained by neutral red, indicating the proton concentration in the region of duct is high. To determine if any enzyme activity is associated with high proton concentration, the specific v-ATPase inhibitor vanadium was added (Vollrath et al., 1998). Vanadium prevents neutral red staining of the distal duct, therefore v-ATPase activity is related to proton accumulation. The gland duct positively reacts with p-phenylphosphate (ATPase substrate and indicator) and is inhibited by the cytoplasmic ATPase inhibitor ouabain, indicating that c-ATPase may also be present in the duct. Therefore, the distal duct presents two types of proton pumps: c-ATPase that pumps protons into the cytoplasm from the intercellular regions and v-ATPase that actively pumps protons from the cytoplasm into the vesicles (Vollrath et al., 1998). Proton fluctuation may contribute to aggregation and proper folding of spidrons since protonation changes are often involved in the formation of protein complexes (Mason and Jensen, 2007). Maybe the two types of ATPases have to work in pairs to maintain proton homeostasis at a certain pH level right for fiber assembly.

Carbonic anhydrase catalyze the reversible hydration of carbon dioxide to bicarbonate. Such activity is found in the distal sac and in the duct by histochemical staining with Azure blue (Andersson et al., 2014). Carbon dioxide pressure is therefore, deduced to build up along the gland. Carbonic anhydrase may also contribute to the maintenance of pH gradients in major ampullate
glands since a carbonic anhydrase inhibitor, methazolamide, changed the pH from 7.6 to 7 in the gland tail and sac region.

Peroxidase activity was detected by incubating dissected major ampullate gland with 3, 3', 5, 5' tetramethyl-benzidine (TMB) and hydrogen peroxide (TMB substrate) (Vollrath and Knight, 1999). Peroxidase activity gradually increases from the border of Zone A and Zone B and peaks right in front of the funnel. Weak activity is also detected in the nascent fiber thread. An *N. senegalensis* peroxidase gene (*NsPox*) was identified by screening an EST library of the major ampullate gland using a consensus peroxidase motif (Pouchkina et al., 2003). The translated protein is 634 amino acids with a predicted pI of 5.9.

Chitin is a natural polymer often found in the cuticles of various organisms for support or strength. After KOH treatment of the dissected major ampullate gland and iodine staining, chitin is detected when iodine turns the chitin that lines the distal limb, valve and spigot into a reddish brown color. Although the rest of the duct cannot tolerate KOH treatment and is unable to be stained by iodine, an aqueous matrix of chitinous material can be deduced from the FTIR (fourier transform infrared) spectrum (Davies et al., 2013).

**Molecular biology of major ampullate spidroin coding sequences**

Two *N. clavipes* cDNA clones: *MaSp1* (Xu and Lewis, 1990) and *MaSp2* (Hinman and Lewis, 1992) were identified through cDNA library screening using probes derived from peptide sequences obtained from dragline silk peptide
cleavage and sequencing. The partial sequence of MaSp1 includes a repetitive region of motifs rich in glycine (GGX) and polyalanine along with a non-repetitive C-terminal domain (CTD). The partial sequence of MaSp2 includes a repetitive region of motifs rich in glycine and proline (GPGGY or GPGQQ) and polyalanine along with a conserved non-repetitive C-terminal domain (CTD). Using the probes created from _N. clavipes MaSp1_ and _MaSp2_ sequence, partial cDNA clones were isolated from an _Araneus diadematus_ major ampullate gland cDNA library (Guerette et al., 1996). Four different protein repeat motif sequences were deduced from the isolated cDNA clones. ADF-1 contains the repeat motifs (GA)_n_, GGYGQGY and _A_n_ blocks. ADF-2 contains the repeat motifs GGAGQGGY, GGQGGQGGYGGGLGSQGA and _A_n_. ADF-3 contains the repeat motifs ASAAAAAA, GPGGQGPYGPG, GGYGPYS, and (GPGQQ)_n_. ADF-4 contains the repeat motifs SSSSSSSSSSSS, GPGRIDGPS and GPGGY. Due to the difficulties in cloning the long repetitive regions of _MaSp1_ and _MaSp2_ and the limitation of sequencing at the time, full length sequence of _MaSp1_ and _MaSp2_ were not revealed until more than ten years after.

Full length _MaSp1_ and _MaSp2_ genes (coding region, 5’ upstream and 3’ downstream region) were isolated by screening a genomic library of _Latrodectus hespeus_ using the primers developed from partial _LhMaSp1_ and _LhMaSp2_ sequences (Ayoub et al., 2007; Garb and Hayashi, 2005). The full length _Lh MaSp1_ gene encodes a single exon (9,390 bp, 3,129 aa) and _LhMaSp2_ encodes a single exon (11,340 bp and 3,779 aa) (Ayoub et al., 2007). Both sequences
which might have evolved from a gene duplication are composed of a non-repetitive N-terminal coding region, a C-terminal coding region and a large central repetitive coding region (Fig. 1.5). The MaSp1 repeat region is represented by repeats of GGX (where X is A, Q or Y), GX (where X is Q, A or R) and poly-alanine. The MaSp2 repeat region is represented by relatively larger number of repeat unit of GPX (where X is G or S), QQGGX (where X is A or GSG) and poly-alanine. The MaSp1 repeat region is essentially devoid of proline, the amino acid with a ring structure that may play an important role in protein folding. These dragline silk motifs are conserved in all orb-weaving spiders over millions of years of selection (Gatesy et al., 2001).

Sequence analysis of MaSp1 and MaSp2 identified a few upstream transcription regulation elements: a TATA box, a putative spider silk gene specific regulatory element, a CACG motif and a putative 15 bp transcription factor binding site identified at ~110 bp upstream of the start codon (Ayoub et al., 2007; Motriuk-Smith et al., 2005).

Multiple copies of MaSp1 clones have been identified in Nephiila (Chinali et al., 2010; Gaines and Marcotte, 2008). Due to the simultaneous presence of polymorphisms and multiple gene copies, multiple MaSp1 genes may be present at different loci as a result of gene duplication instead of two alleles of the same locus. Other than duplication, gene recombination may occur between MaSp1 loci or between MaSp1 loci and MaSp2 loci (Ayoub and Hayashi, 2008). It is postulated that for each MaSp1 NTD isoform, there may be a respective MaSp1
CTD isoform to pair with it (Gaines and Marcotte, 2008). This may be true since the MaSp1 CTD isoform is found in the assembled transcriptome of a non-orbweaver *Latrodectus hesperus* (Clarke et al., 2014).

A short isoform of *MaSp1* has recently been identified from the threads and gland extract of *Cyrtophora moluccensis*, and was annotated as *MaSp1s* (Han et al., 2013). This short gene encodes a non-repetitive NTD (447 bp) and CTD (294 bp) that are similar to other known MaSp1 and a significantly small repeat region (576 bp) that is distinct to *MaSp1s*. It is neither an isoform that could be possibly translated from a downstream 5’ methionine site of the previously identified *MaSp1* nor the product of an alternative splicing event. As part of the silk threads, this novel polypeptide *MaSp1s* may be auxiliary to MaSp1 and MaSp2 in fiber assembly.

Short (~10 kDa) cysteine rich proteins (CRPs) with predicted pIs of 8.3 that form putative slipknot structures have also been identified from gland protein extracts and dragline silk fibers. CRPs are co-expressed with MaSp1 and MaSp2 and form high molecular weight multimers in non-reducing conditions but exhibit relatively lower molecular weight multimers in reducing conditions, which may relate to disulfide bond linkage in an oxidized environment (Pham et al., 2014).

**Transcription and translation:**

About 100,000 transcripts are generated from the transcriptome study of the spider *L. hesperus*, of which 27,000 are annotated. Silk-gland specific
transcripts include gene families of proteases, protease inhibitors, oxidoreductases, putative transcription factors, putative flagilliform spidroins and novel glue proteins (Clarke et al., 2014). An orb-web weaving spider transcriptome has been built from deep RNA sequencing of transcripts from two species of *Gasteracantha* and one species of *Nasoonaria*, revealing 9,300 to 10,494 annotated unigenes. MaSp1 and MaSp2 N-terminal signal peptides and non-repetitive domains are found in all three species and align with previously discovered *N. clavipes* sequences (Zhao et al., 2014). Other than transcripts, non-coding RNA that may relate to dragline silk formation was also identified. A 5S rRNA gene that codes for a 120 nt product is found to be upregulated before spidroin translation in response to *N. clavipes* major ampullate gland stimulation (Vazquez et al., 2003).

Translational regulation of major ampullate spidroins is distinct in that the glycine-rich and alanine-rich repetitive domains of MA spidroins make it difficult to produce these proteins because of the very quick depletion of the enriched amino acids and their corresponding charged tRNAs (Candelas et al., 1990). Isotope $^{14}$C- or $^3$H-glycine labeled gland extracts showed large amounts of glycine are rapidly incorporated into the gland peptides (Candelas et al., 1983). These peptides have a laddered banding patterns on SDS-PAGE gels suggesting that glycine had been incorporated into the polypeptide chain at various translational stages so that the length of the protein would vary. This could result from simultaneous translation on multiple ribosomes at different
speeds. Major ampullate spidroin repeat domains also contain alanine rich regions. The high demand of alanine tRNA during translation depletes the alanine tRNA fairly quickly. A gland specific tRNA has been found to hybrize with the constitutive silk worm alanine tRNA probe (Candelas et al., 1990), indicating that this gland specific tRNA may be involved in alanine transfer. This is also supported by a study that a cell free rabbit reticulocyte translation system, when supplemented with tRNA extracts from *N. clavipes* major ampullate gland, incorporates isotopic alanine into various sized peptides (Candelas et al., 1989).

Translational pausing may relate to the stability of mRNA secondary structure of spidroins. Computer modeling of MaSp1 mRNA secondary structure shows that poly-alanine encoding regions are more stable than poly-glycine encoding regions. This alternating secondary structure stability may cause translational pause. However, the energy stability of MaSp2 mRNA secondary structures stays the same in alanine-rich regions and glycine- and proline-rich regions (Zama, 1997, 1999). Whether a translation pause occurs in MaSp2 remains unknown.

**Post-translational modifications:**

NMR, mass spectrometry and glycan detection of dragline silk fibers and gland extractions reveal evidence of post-translational phosphorylation and glycosylation events. Solid- state $^{31}$P-NMR of *N. clavipes* dragline silk identified phosphotyrosine, trace amounts of phosphoserine and a phosphate that may be
part of a five-membered cyclic structure that includes at least one sugar (Michal et al., 1996). This cyclic phosphate isn’t shown in the spidroin extracts from the gland but appears in the dissolved fiber solutions, suggesting the cyclic phosphate glycan is incorporated after translation and secretion of spidroins. Chromatography coupled with mass spectrometric analysis of electrophoresis separated MaSp1 from *N. clavipes* dragline silk detects eight serine and tyrosine phosphorylation sites which can be verified by phosphatase treatment (dos Santos-Pinto et al., 2014). The phosphorylation sites cover the entire repeat domain but neither NTD nor CTD. L-Dopa (3,4-dihydroxyphenylalanine) and dityrosine are detected by HPLC followed by fluorescent spectrophotometry from hydrolyzed major dragline silk solutions. The presence of L-Dopa and dityrosine may be a result of tyrosine oxidation.

Glycan is detected in the solubilized dragline silk of *N. clavipes* using digoxigenin (DIG) glycan/protein double labeling, although the oligosaccharide molecule is still unknown (Guehrs et al., 2008). Silkworm fibrion P25 is known to form complexes with silkworm fibroin through non-covalent bonding. The short cysteine rich protein (CRPs) that have been identified in the gland protein extract and dragline silk fiber may be an equivalent to silkworm P25 in spiders. The small molecular weight P25 is glycosylated with high mannose-type N-linked oligosaccharide chains. The number of N-glycosylation sites will vary based on whether the silkworm silk fibroin heavy chain is disulfide bonded to light chain or not (Tanaka et al., 1999).
Major ampullate spidroin protein sequence, structure and folding

The original deduced partial amino acid sequence of MaSp1 was obtained from cDNA clones isolated from the screening of *N. clavipes* major ampullate gland cDNA library using peptide sequence identified from reconstituted dragline silk (Xu and Lewis, 1990). The original deduced partial amino acid sequence of MaSp2 was obtained from the cDNA clones isolated from rescreening of the same library using a peptide sequence that is different from MaSp1 (Hinman and Lewis, 1992). Full length *LhMaSp1* and *LhMaSp2* are 347 kDa and 420 kDa, respectively (Ayoub et al., 2007). The molecular weight calculated from the deduced amino acid sequence matches the molecular weight determined from gel filtration of gland spidroin extracts (Sponner et al., 2005a). The deduced amino acid sequence of *Cyrtophora moluccensis* MaSp1s was obtained using a similar cDNA library screening based approach (Han et al., 2013). All three major ampullate spidroins (MaSp1, MaSp2 and MaSp1s) have conserved non-repetitive N-, C-terminal domains flanking repetitive regions distinct to each spidroin. Figure 1.4 shows an illustrated N-, C-terminal domains flanking ~100 repetitive modules. The sequences are slightly different in each module but can be generalized as (Hayashi et al., 1999):

MaSp1: (GGA) GQ (GGY) (GGL) (GGQ) GAGR (GGL) (GGQ) (GA)₂ (A)₃
MaSp2: (GPGGY) (GPGQQ) (GPGGY) (GPGQQ) GPSGPS (A)₉
MaSp1s contains only seven tandem repetitive motifs and each motif varies slightly in sequence. The entire repetitive region is very short (only 192 aa) and a closer look of the motif showed that MaSp1s repeat region has all three types of MaSp1 motifs: GGX (X=A, Q, or Y), GX (X=Q, A, or R) and poly-A (Han et al., 2013).

The total protein concentration in a major ampullate gland is estimated to be 50%. This was determined from the isotopic peak intensity of the NMR spectra of isotope fed *N. clavipes* gland extracts. The isotopic carbon of labeled glucose is mainly incorporated into glycine, alanine and glutamine. No labeled proline or tyrosine is detected in dissolved silk, suggesting that proline synthesis may be discrete from the normal pathway or proline synthesis depends on precursors from food sources (Hijirida et al., 1996).

Isoelectric focusing gel analysis of gland protein extracts revealed two bands. One band with pH above 8.5 reacts to anti-MaSp1 repeat sequence serum, indicating pH of MaSp1 is above 8.5. Another band with pH between 5.1 and 5.9 reacts to anti-MaSp2 repeat sequence serum. Thus, the pH of MaSp2 is between 5.1 and 5.9. The results match well with the predicted pH from the deduced MaSp1 and MaSp2 sequences (Sponner et al., 2005a).

Hydrophobicity sequence analysis reveals that the N-terminal domain of MaSp1 and MaSp2 are hydrophilic. A small patch that is close to the end of the C-terminus of MaSp1 or MaSp2 NTD is hydrophobic. In the repeat domain,
hydrophilic poly glycine patches alternate with hydrophobic poly-alanine patches (Bini et al., 2004).

Spidroin tertiary structure has been investigated at multiple pH values since pH declines along the major ampullate duct. Spidroin folding examined by polarized light microscopy shows that spidroin folding progressively changes from disoriented at the beginning of the duct to alternatively oriented in limb 2 and to uniaxial oriented and positively birefringent in limb 3 (Knight and Vollrath, 1999). The spidroins are predominantly random coil structured at pH 7.3 in Zone A (Dicko et al., 2004b). Raman intensity shift pattern of collected N. clavipes major ampullate silk gland extracts and dragline silk indicates some α-helix structures but mainly 3_1-helix structures in the native spinning dope (the unfolded state) (Lefèvre et al., 2007b). As the dope exits the lumen and passes through the funnel, the transformation starts with the formation of a few β-sheets but remains mostly ordered as α-helix and 3_1-helix in the limb 1 and limb 2 region. From limb 3 to the valve, β-sheet is the dominant structure and α-helices structures become diminished after the valve. The conversion from unfolded helixes to β-sheets is instantaneous, with no intermediate state detected (Fig. 1.3, Lefèvre et al., 2008).

The NMR spectrum of dragline silk fibers from deuterated alanine-fed N. clavipes revealed two alanine structures in the silk fiber: 40% of alanine exists in highly oriented β-sheet stacked crystallites and about 60% of alanine is in loose, more poorly oriented β-sheet crystallites embedded in the glycine-rich matrix
Glycine is present in both disordered (GGA motif) and ordered structures (poly-alanine and poly (GA)) with approximately 28% of the glycine existing in the β-sheet structure (Holland et al., 2008a). Another 21% of the glycine is ordered but not uniaxial (Bonev et al., 2006), as a random polyproline II (PPII) type (Lefèvre et al., 2007a).

This two component theory is further supported by FTIR measurement that detects a crystalline phase that is ordered/oriented and an amorphous phase that is loose/unoriented (Ene et al., 2009). The crystallites are visible under TEM as irregular shapes, 70-100 nm in diameter and dispersed evenly in the amorphous matrix. The difference of electron energy loss in crystallite regions and amorphous regions indicates the structure differences. This difference is also seen within the amorphous region, indicating some parts of the amorphous regions are structured differently than the other parts (Thiel et al., 1994).

The two phase structural arrangements of spider dragline silk enables the integration of strength and extensibility with crystallites contributing to the strength and amorphous regions contributing to the extensibility. The number or density of the crystallites is critical to fiber tensile strength. Computer modeling of spider silk suggests that lower numbers of β-sheet crystallites leads to higher tensile strength (Xu et al., 2014). The cross-connection of the crystallites and the amorphous region creates a fibril network which prevents the damage of small cracks and keeps the fiber connected. In addition, sliding oriented alanine β-
sheet through the amorphous matrix generates friction resistance that contributes to the tensile strength of the fiber (Patil et al., 2014).

As spidroins are pulled along the tapered spinning duct, the shear force created from the spinning duct wall increases hydrostatic pressure, which is postulated to increase order of the macromolecules (Anton et al., 2015). Shear force induces rADF-3 (MaSp2 repeat analog) aggregation (Eisoldt et al., 2010). The viscosity of the major ampullate gland extracts decreases as shear rate increases. Once it reaches the critical shear rate, the spidroin extracts start to thicken. Large molecular weight protein aggregating to fibril is induced by shear force and the shear rate required to induce protein aggregation is pH related. At pH 6.4, protein aggregation can be induced with relatively small shear rates (Chen et al., 2002b).

The repetitive domain

The ~100 repeat motifs in spider draglines silks cover about 95% of the protein sequence. The sequence of the repeat motif and the number of the repeat domains is critical to protein solubility, protein folding, and tensile strength.

MaSp2 analogs rADF-3 and rADF-4 show different solubility when produced in Sf9 insect cells. rADF-3, with tandem repeat motifs ASAAAAAA, GPGGQGPYGPG, and GGYGPGS is soluble when accumulated in the Sf9 cell but rADF-4, with tandem repeat motifs SSAAAAAAA, GPGSQGPS, and GPGGY is insoluble and aggregates to nanofibrils inside the Sf9 cell. Neither of
these two constructs includes N- or C-terminal domains in this study (Huemmerich et al., 2004b). A later study included constructs of variable combinations of rADF3 and rADF4 sequence modules plus a CTD non-repetitive sequence. The engineered proteins with modules from rADF3 can be concentrated to >30% regardless of the presence or absence of the CTD non-repetitive domain. Engineered proteins with the consensus repeat module from rADF4 have less than 9% solubility regardless of the presence of CTD.

The different aggregation pattern of rADF3 and rADF4 proteins in the cell makes them a good system to evaluate how repeat domain number relates to protein aggregation and folding. rADF3 with 24 copies of repeat domain aggregates more than rADF3 with 12 copies of repeat domain (Hagn et al., 2010). rADF4 with a single repeat motif is soluble and is in random helical structure while rADF4 with two repeat motifs or more is insoluble and is in β-sheet structure (Humenik et al., 2014b). With increasing numbers of repeat domains, rADF4 aggregates into longer ordered fibrils with less branches and lower solubility (Ittah et al., 2012). With increasing repeat domain numbers, the lag phase of rADF4 aggregation is shortened and the exponential growth rate is increased. This may indicate that larger repeat domains creating stronger hydrophobic interactions accelerates the intermolecular interaction and hydrogen bond rearrangements (Humenik et al., 2014b).

The stability of one of the GGX motifs of MaSP1, GGR, is slightly different in the nonionic state versus the zwitterionic state based on both Raman
microspectroscopy and computer modeling (Ari and Ozpozan, 2015). The stability of the GGR 3\textsubscript{10}-helix peptide chain is affected by this motif at acidic pH. In the MaSp2 motif GPGGX, the hydrophobicity and hydrophilicity of the fifth amino acid affects the extensibility of the spun fiber (Teulé et al., 2007).

**The non-repetitive C-terminal domain**

The short non-repetitive C-terminus in MaSp1 shares 75% identity with MaSp2 (Beckwitt and Arcidiacono, 1994). Three hydrophobic patches appear across the sequence of the CTD with interspersed amphiphathic amino acid patches (Sponner et al., 2005c). The amino acid sequence and hydrophobicity pattern are both conserved among other spider species (Challis et al., 2006). There is also a conserved cysteine residue in the CTD that is important for fiber formation (Ittah et al., 2007, Fig. 6).

The CTD of the MaSp2 analog rADF-3 folds into five parallel helixes and forms a homodimer with one disulfide bond at the hydrophobic interface and two salt bridges from the only two paired charged amino acids (Fig. 1.6). The hydrophobic interface is folded inward so the hydrophilic amino acids are exposed to the solution.
Replacement of the charged amino acids with uncharged amino acids or acidic pH (pH= 2) disables the salt bridge and causes unstable folding of the five helices that may expose the hydrophobic region (Hagn et al., 2010).

Recombinant *N. clavipes* MaSp1 CTD produced from *E. coli* partially unfolds at pH 5, exposing the hydrophobic area to the solvent. The exposed hydrophobic area may provide a nucleation site for poly-alanine to initiate fibril assembly (Gauthier et al., 2014). Similar results are also shown for MiSp1 CTD which unfolds and switches into an irreversible β-sheet structure at pH 5.0. The elevated CO$_2$ along the gland may have a high affinity to the exposed hydrophobic region of CTD and facilitate β-amyloid like nucleation of CTD, which can be detected by Thioflavin T (ThT) fluorescence (Andersson et al., 2014).
The role of CTD in fiber assembly has been demonstrated in various studies. CTDs accelerate spidroin repetitive domain protein aggregation (Sponner et al., 2005c). MaSp2 repetitive domain aggregates without CTD but it aggregates to a greater extent in the presence of CTD. Analogs of MaSp1 with four repeats and a CTD forms fibrils in 300 mM phosphate solution in 20 minutes whereas fibril formation takes overnight if the construct does not contain a CTD (Hedhammar et al., 2008). The repeat analogs for rADF3 and rADF4 differ in solubility such that rADF3 repeats are soluble but rADF4 repeats form fibers in transgenic Sf9 cell lines (Huemmerich et al., 2004b). CTDs enhance the sensitivity of both rADF3 and rADF4 to 300 mM potassium phosphate buffer induced in vitro fibril assembly (Huemmerich et al., 2004a).

The CTD also reinforces protein alignment during fiber formation. Both rADF3 and rADF4 form randomly aligned aggregates without the CTD but form ordered fibrils when the CTD is included (Eisoldt et al., 2010; Ittah et al., 2006).

The non-repetitive N-terminal Domain

The NTD is the most highly conserved domain of the spider silks (Bini et al., 2004). It is hydrophilic and is conserved among the dragline, flagelliform and cylindriform silk proteins (Rising et al., 2006), across different spider species (Motriuk-Smith et al., 2005), and among waxmoth and silkworms fibrioin NTDs (Bini et al., 2004). The two *N. clavipes* MaSp1 isoforms MaSp1A and MaSp1B are 91% identical and share 72% identity with the MaSp2 NTD (Gaines and
Marcotte, 2008). Spidroin NTD is present in the epithelial secretion granules in Zone A and Zone B of the major ampullate sac, but not in Zone C (Andersson et al., 2013). NTDs can only be detected in the lumen of Zone C. In the spun dragline silk, NTD can be detected in both the inner and outer core regions (Andersson et al., 2013).

X-ray structure of MaSp1 NTD demonstrates that it is composed of five anti-parallel α-helices. The NTD can form a homodimer at the hydrophobic interface (Askarieh et al., 2010, Fig. 1.7). NcMaSp1 NTD helical structure is also obtained from NMR assignment and backbone shift experiments (Parnham et al., 2011). Computational modeling shows two salt bridges, Asp39-Arg60 and Asp40-Lys65, may contribute to the formation of an NTD dimer. Acidic pH may change the protonation state of the residues Glu 79 and Glu119 and allow water

access in the embedded regions (Wallace and Shen, 2012). High salt may weaken the salt bridge and raise the free energy of the dimer causing it to become unstable (Gronau et al., 2013).

NTD forms thermal stable homodimers at pH 6 (Hagn et al., 2011). Clusters of negative and positive charged residues are shown at the dimer interface constructed from NMR spectra. At neutral pH, the single tryptophan residue in the NTD is embedded in the hydrophobic region. Tryptophan fluorescence signal is quenched when the residue is exposed to the solvent and the signal change can be tracked spectrophotometrically. Recombinant *N. clavipes* MaSp1A-NTD and MaSp2-NTD show tryptophan fluorescence quenching at lower pH (Gaines et al., 2010). This change is mitigated in high ionic strength solutions such as NaCl, KCl and sodium phosphate. Resin-immobilized NTD is able to bind free NTD at low pH, forming stable homodimers that are difficult to separate. The area around the tryptophan residue in a stable NTD homodimer is folded in a way that restricts the movement of downstream repeat domains and forces the repeat domain to initiate the β-sheet folding (Askarieh et al., 2010).

Studies on MaSp1 NTD wild type and variants show that dimer formation is completely abolished for the D40N variant or the D40N/E84Q double variant and partially impaired for the E84Q variant (Landreh, 2010). Tryptophan fluorescence change at pH 6 is abolished for the A72R variant. This variant is also unable to form stable dimers, since the tryptophan residue remains buried at
low pH while it is largely surface exposed in the dimer (Jaudzems et al., 2012). Acidic pH induced tryptophan fluorescence shifts are reversible for *E. australis* MaSp1 NTD, albeit no tryptophan fluorescence change is found in buffer supplemented with 0.3M NaCl (Askarieh et al., 2010)

When a fluorophore is engineered in close proximity to the tryptophan residue, it captures the native tryptophan fluorescence changes that reflect the conformational changes of the polypeptide chain. Single-molecule fluorescence caused by photon induced electron transfer shows the conformation change around the native tryptophan to occur within 100 micro seconds (Ries et al., 2014). This process is pH dependent. Free energy profile reveals that lower pH may reduce the large free energy barrier required for the initiation of dimer association.

Stable NTD dimer formation may involve three consecutive steps: electrostatic changes of D40, R60 and K65 to initiate static attraction; protonation change of E79 and E119 for the proper conformation and the protonation change of E84 to stabilize the dimer (Kronqvist et al., 2014). Interestingly, the NTD behaves exactly opposite to the CTD in that acidic pH locks NTD into stable homodimers, restricting CO₂ binding (Andersson et al., 2014).

**Importance of Proline to the structure and function of MaSp2**

A major difference between the repeat domain of MaSp1 and MaSp2 is that the MaSp2 repeat is proline rich whereas the MaSp1 repeat is devoid of
proline. The imino group of proline reacts with the α-carbon to form a peptide bond and generates a kink in the polypeptide chain so that it misses a hydrogen bond donor to form a hydrogen bond. Proline is one of the more costly amino acids to produce inside a living organism (Craig et al., 1999). The average cost to produce MaSp2 (17 ATP/aa) is much higher than MaSp1 (14.5 ATP/aa). Poorly fed or low proline diet spiders produce less MaSp2 than proline-sufficient diet spiders. This can be seen from the weak anti-MaSp2 signal of poorly fed spiders’ silk (Guehrs et al., 2008). The incorporation rate of isotopic glucose into proline from earlier studies also suggests that proline synthesis takes time or it relies on precursors from food sources (Hijirida et al., 1996).

The proline kinks adversely affect β-sheet extension, forcing the peptide chain into an amorphous structure (Thiel et al., 1997). Interestingly, proline substitution at the amyloid protein β-sheet core region abolished β-amyloid aggregation since proline kinks keep the polypeptide chain from accessing the proper angle to aggregate (Wood et al., 1995).

AFM imaging and X-ray scattering analysis showed a less ordered MaSp2 rich inner core for *N. clavipes* dragline silk which had a higher extensibility compared to the MaSp1-rich outer core region. MaSp2 is less ordered due to the proline kinks in the repetitive domain, which may hold water and cause supercontraction (Brown et al., 2011). MaSp2 repeat motif of *A. aurantia* dragline silk fiber is proposed to have two types of motions in the protein backbone when
the fiber is rehydrated in which motifs are closer to the poly-alanine region moves faster than the motifs that are further to the poly-alanine region (Shi et al., 2014).

A comparison of entropy change between high proline content *A. diadematus* fiber (16%) and low proline content *N. clavipes* fiber (3.5%) during hydration shows that the entropy change of high proline fiber results from the free energy change of the restricted proline network chain. The entropy change of low proline fiber results from the internal energy change of disturbed hydrogen bonds (Savage and Gosline, 2008b). The higher proline content fiber is associated with lower strength, lower stiffness and higher extensibility (Liu et al., 2008).

Except for only a few prolines in the non-repetitive domain of MaSp1, proline is exclusively in MaSp2. Proline content in the dissolved dragline silk solutions of *N. clavipes* and *Argiope aurantia* are 4.3% and 6.4%, respectively (Lombardi and Kaplan, 1990). Comparisons of mechanical performance of dragline silk from these two species are inconclusive due to the large standard variation of the samples tested. The overall toughness of the dragline silk of these two species is very close (Brooks et al., 2005). The mechanical performance of *Nephila clavipes* dragline silks with proline content of 3.5% and *Araneus diadematus* dragline silks with proline content of 16% were also assessed. The differences of the tensile property of the draglines silks of these two species cannot be discerned from each other until the silks are rehydrated. Despite the high variation of the data, high proline content *A. diadematus*
dragline silk has a lower initial modulus and it is less stiff than *N. clavipes* dragline silk upon rehydration.

Increasing the number of recombinant proline rich MaSp2 motif elevates β-turn structure percentage and the extensibility of the spun fiber (Brooks et al., 2008b). Surprisingly, recombinant MaSp2 spun fiber has higher tensile strength than recombinant MaSp1 spun fiber. A mixture of the two spidroin analogs shows intermediate tensile strength (Brooks et al., 2008a). For MaSp1 and MaSp2 repeat domain analogs, a combination of MaSp1 and MaSp2 in 4:1 ratio gives best overall toughness (An et al., 2012). Proline content also affects the degree of the change in birefringence in that higher proline content is associated with greater birefringence change (Savage and Gosline, 2008a).

**Hydrogen bond and supercontraction:**

The spinning dope that enters the duct contains about 70% water, which will be removed when the spinning dope proceeds along the length of the spinning duct. The removal of water is relevant to the spidroin migration speed and water diffusion coefficient (Kojic et al., 2004). Adding the chaotropic ion potassium in the form of KCl induces β-sheet formation since chaotropic potassium facilitates removal of water from potential hydrogen bonding sites and making them available for polypeptide chain interaction (Chen et al., 2002a).

As three to four hydrogen bonds cluster in a β-sheet crystal in the space of a few nanometers, polypeptide chains interconnected with weak hydrogen bonds
form a strong network that resists breakage (Keten and Buehler, 2008; Keten et al., 2010). Hydrogen bonds are postulated to form between serine or tyrosine and glycine in β-sheet crystals and the unstructured helical and β-turn regions (Hayashi et al., 1999). The strength of the hydrogen bond between alanine and glycine is identical in the β-sheet and helical region (Holland et al., 2013).

*N. clavipes* major ampullate dragline silk supercontracts when the environment switches from dry to wet. However, minor ampullate fibers are unable to supercontract in water (Guinea et al., 2012). Switching to a wet condition like 70% relative humidity (RH), major ampullate dragline silk supercontraction causes a measured 40-80 MPa stress. Water absorption during supercontraction at 70% RH is 1.6% (Blackledge et al., 2009).

Hydrogen bonds function as an internal network spacer of the amorphous region (Simmons et al., 1996) and hydrogen bond rearrangement in the amorphous region is believed to be the contributing factor to supercontraction. Water induces spidroin polypeptide chain structure changes by disturbing the hydrogen bonds that support the amorphous area in dry conditions, causing this area to deform leading to supercontraction of the fiber (Ene et al., 2009, Fig. 1.8). This is confirmed by wide angle X-ray diffraction analysis of *N. clavipes* dragline silk (Grubb and Ji, 1999). The methyl composition of the alanine rich crystal is unchanged but the amorphous region is isotopically reoriented. A slight change of the β-sheet orientation relative to the fiber axis can be detected when the fiber is wet (Parkhe et al., 1997). Further evidence from AFM and X-ray scattering

Data of *N. clavipes* dragline silk and NMR data confirms that the size of β-crystals does not change upon wetting (Brown et al., 2011). From dry to wet, the MaSp1-rich outer core and the MaSp2-rich inner core both become stiffer (Brown et al., 2011).

**Spider silk fiber structure and self-assembly:**

Early evidence that spider dragline silk includes multiple proteins were based on the observation that the shape of the proteinic granules is different in
the distal or proximal half of the major ampullate gland and duct. However, these
different granules may contain different substances. The elongated spherical
droplets in Zone B may contain glycoproteins that are rich in acidic groups that
become the outer layer of the fibers (Knight and Vollrath, 1999; Vollrath and
Knight, 1999). Light reflection of AFM imaging of *N. clavipes* dragline silk
supports the notion that dragline silk consists of outer layer regions and a core
region as shown in the fiber cross section (Brown et al., 2011). The glycoprotein
containing 150-250 nm thick skin layer can be detached from the core region
mechanically or chemically with trypsin (Augsten et al., 2000). TEM and AFM
microscopy study of *N. clavipes* dragline fibers showed a much more detailed
layer structure with the skin layer further differentiated into an outer most layer, a
soft glycoprotein layer and a rigid skin layer (Fig. 1.9, Sponner and Vater, 2007).
The outer most layer is a lipid membrane coat (10-20 nm) that is hard to preserve
during sample sectioning and can be removed by gentle washing. The soft
glycoprotein layer is 40-100 nm thick with a high affinity to osmium tetroxide and
nanofibrils seen in this layer that could be washed off with detergent. The protein
component in the soft glycoprotein layer is above 200 kDa and reacts to none of
the silk specific antibodies. The rigid skin layer (50-100 nm thick) shows sharp
contrast under the microscope to the core regions, can only be
removed by freeze-thawing, contains minor ampullate spidroin (MiSp)-like
proteins above 200 kDa and only reacts with anti-dragline silk antibodies but not
the anti-MaSp1 or anti-MaSp2 antibodies.
The innermost layer is the core region which consists of an outer core region and an inner core region. The core region of the fiber consists of a bundle of nanofibrils. SEM and AFM of *N. pilipes* dragline silk reveals that each fibril runs parallel to the fiber axis (Du et al., 2006). Each fibril is a network of interconnected β-sheet crystals and the surrounding relaxed amorphous regions (Fig. 1.10, Du et al., 2006). The outer core region only reacts to MaSp1 specific
antibodies and the inner core region reacts to both MaSp1 and MaSp2 antibodies and is a mixture of MaSp1 and MaSp2 (Sponner et al., 2007).

Simulation calculation of shear forces in the spider major ampullate gland is greater than the shear force in silkworm silk glands, suggesting shear force is

Figure 1.10. SEM (a) and AFM (b) imaging of Nephila pilipes dragline silk. Reprinted with permission from Du, N., X.Y. Liu, et al. (2006). Design of superior spider silk: from nanostructure to mechanical properties. Biophys J 91: 4528-4535. Copyright (2006) Elsevier.
essential in spider silk assembly. Spidroins extension flow rate and the extensional strain both have a long lag phase followed by an exponential increase phase (Breslauer et al., 2009).

A spider pulling at a speed of 10-20 mm/s is optimum for producing fibers with a perfect combination of crystal size (6.4nm in the fiber axis direction), orientation (aligned parallel to the fiber axis) and mechanical performance (maximum tenacity) (Du et al., 2006). Calcium deposition is found only on the β-sheet crystallites of the dragline fiber (Thiel et al., 1994). Natural dragline silk fibers are positively birefringent (Vollrath et al., 1998).

Fiber assembly is a complex process that relies on many factors including spidron production, biological, chemical and physical influences as the spinning dope proceeds along the duct, the reeling speed, and the nutrient condition of the spiders. Figure 1.11 (Gaines et al, 2010) illustrates a multimeric strain formation model from the stabilized NTD homodimer at acidic pH for fiber assembly. Here I am going to focus on some additional information that is related to the fiber assembly process.

Self-assembly of MaSp2 analog rADF4 into fibrils inside the cytosol of Sf9 cells most likely involves a nucleation process since all the rADF4 produced from the cells are assembled into the fibrils (Huemmerich et al., 2004b). The first step involves the accumulation of monomers and nanofibril formation. The second step involves spontaneous conversion of nanofibrils into β-sheet crystals and then fibers once they reach critical concentration (Ittah et al., 2012). The nucleation
phase starts with the hydrophobic interaction of the poly-alanine motif with other nearby poly-alanine motifs causing the assembly of an oligomer nucleus. The oligomer chain elongation phase involves deposition of monomers to the nucleus seeds resulting in conformational changes that favor deposition of the next molecule (Humenik et al., 2014b). β-sheet aggregation kinetics can be monitored by $^{13}$C solid-state MAS NMR which shows that β-sheet rich nanofibrils can be detected after 24 hours of incubation of major ampullate gland extracts at different pH. The length of the lag phase in the aggregation kinetics curve varies at different pHs (Xu et al., 2015). Another study with rADF3 proteins showed that the lag phase of nucleation can be shortened by mixing the rADF3 solution with the previously formed “seed” nucleus (Humenik et al., 2015). The size of the
oligomers is affected by pH, the ionic strength and shear force (Leclerc et al., 2013).

Although a lot of information has been obtained in the past decade, many aspects of natural fiber assembly are still a mystery. Because dragline spider silk exhibits superior mechanical properties and is the toughest known material, mass production of spider silk is relevant. The inability of spider silk to be supplied from spider farming is unfortunate, therefore, recombinant spidroin production and assembly provide hope to mass produce the dragline silk-like materials. The next few sections will discuss the current progress of recombinant spidroin production, fiber assembly, their mechanical properties and biomedical applications on spider silks and their analogs.

Recombinant production of spider silk-like proteins and fiber assembly

Ever since partial cDNA sequences of MaSp1 (Xu and Lewis, 1990) and MaSp2 (Hinman and Lewis, 1992) and their deduced amino acids sequences are discovered, attempts have been made toward recombinant production of spidroin-like proteins in different expression systems and their assembly into fibers. Details of recombinant production of spidroin-like proteins in a variety of host systems have been recently reviewed, including the type of spidroins that are produced, host system expression efficiency, and obstacles of cloning highly repetitive spidroin-like genes (Heidebrecht and Scheibel, 2013; Rising et al.,
2011). Here, I will discuss gene construction, host system, post-translational control, fiber assembly and post-spin processing; highlighting research that has not only produced spidroin-like proteins but also successfully assembled them into fibers with tested mechanical properties.

**Gene design considerations:**

Because the spidroin repeat domain contributes to the mechanical strength and toughness of the fiber, recombinant proteins intended for spinning into high performance fibrous material contain the highest numbers of consensus repeat domains of MaSp1 and MaSp2 that can be accommodated by the host system (Fukushima, 1998; Lewis et al., 1996; Xia et al., 2010).

To improve mechanical performance of spun fibers, modified repeat domain and computer assisted design have been considered. Computer modeling can design spidroin analogs with predicted mechanical performance, taking into account the length and structure of the polymer block and the calculation of polymer network conductance with simulating shear flow (Lin et al., 2015). Changing the overall charges of rADF4 was achieved by replacement of glutamic acids with lysine, resulting in a modified polypeptide that carries charges opposite to the original polypeptide. This was done to allow the modified rADF4 and the original rADF4 to bind to either negatively or positively charged therapeutic molecules (Doblhofer and Scheibel, 2015).

To target specific areas for higher amounts of protein synthesis in the host system, gene design with a signal peptide that delivers spidroins into the targeted
area is usually considered. This has typically been done in cases where plants are chosen as host systems. A synthetic MaSp1 analogs (DP1B) targeted to apoplast, ER lumen and vacuole in *Arabidopsis* leaf is constructed with sporamin-targeting determinant peptides and an ER retention peptide (Yang et al., 2005). MaSp1 and MaSp2 analogs and rADF-3 were targeted to ER in tobacco leaves by fusing the constructs with an ER retention signal (KDEL) (Menassa et al., 2004). The maximum production yield for ER-directed MaSp1 and MaSp2 analogs are 0.68 and 3.05 mg/kg fresh leaf tissue, respectively.

Spidroin analog fused with a functional domain can be spun into nano/micro structures that carry specific features from the function domain. Examples include:

1) cell binding: An RGD cell-binding domain and spidroin analog fusion protein can form a scaffold to serve as physical support for human bone marrow derived stem cell differentiation (Bini et al., 2006). Other cell binding domains (RGE, IKVAV, YIGSR) fused to spidroins show adhesion to keratinocytes, endothelial and Schwann cells (Widhe et al., 2013).

2) affinity binding domains: IgG-binding domains Z and C2, albumin-binding domain ABD, biotin-binding domain M4 fused with 4RepCT (four MaSp1 repeat analog with CTD) demonstrate binding affinity to their respective target molecule (Jansson et al., 2014).

3) domain that improves mechanical feature: A MaSp2 and flagelliform silk protein (Flag) repeat domain fusion protein increases tensile extensibility and
overall toughness of the spun fibers (Teulé et al., 2012). Fusion proteins of 13 MaSp1 repeat sequence with MaSp2 glycine and proline rich motif have been successfully expressed in *E. coli* earlier, but I am not able to compare the mechanical properties of the fibers made from these proteins due to the lack of a process for spinning these proteins into fiber (Prince et al., 1995).

4) domain that controls tertiary structure: MaSp1 repeat analog is attached by a cAPK recognition site and β-sheet formation of the fusion protein can be controlled by the phosphorylation/dephosphorylation at (Winkler et al., 2000).

5) domain that facilitates post translational protein elongation: A flagelliform protein flanked by self-splicing intein has been targeted to ER in tobacco leaves. After translation, the intein is cleaved leaving a junction that could ligate with another junction, forming various sized flagelliform proteins (Hauptmann et al., 2012).

**Host system considerations:**

Host system selection impacts the quality and quantity of the recombinant protein production, and has permitted assessment of issues related to expressing highly repetitive sequences of spidroin analogs, the solubility and/or premature aggregation, translational pause and post-translational modification. Spidron analogs have been produced in host systems including bacteria (*E. coli* and *Salmonella*), yeast, plants, insect cells, mammalian cells, transgenic silkworm and mice, with exciting possibilities and some obstacles.

*E. coli and Salmonella derived spidroin fiber analogs*
While wild type \textit{E. coli} cells can produce soluble proteins encoding a portion of the MaSp1 repeat domain and CTD (Arcidiacono et al., 1998), engineered glycyl-tRNA rich \textit{E. coli} is able to produce an MaSp1 analog that can be spun into fibers with tensile strength reaching 0.6 GPa and 25\% extensibility (Xia et al., 2010). Enhanced protein production using a SUMO fusion system in \textit{E. coli} is capable of producing various numbers of concatenated rADF-3 and rADF-4 fragments with both NTD and CTD present (Heidebrecht et al., 2015). Due to the various numbers of rADF-3 and rADF-4 repeats, the tensile properties of the fibers assembled from these MaSp2 analogs are diversified.

The \textit{Salmonella} type III secretion system has been explored to direct fusion proteins into the media (Widmaier et al., 2009). Spidroin analog rADF1-3 is produced with an N-terminal secretion signal that directs fusion spidrons to the media and the system includes a transcriptional switch that only turns the gene on when the cell is in secretion mode.

MaSp1 analogs (with NTD and CTD) (Gaines and Marcotte, 2011) and multiple copies of MaSp1 analog (DP1B) were produced using the methylotrophic yeast \textit{Pichia pastoris}. Although multiple copies of target gene insert allows high yield, the recombinant protein size may vary due to recombination events (Fahnestock and Bedzyk, 1997). MaSp1 and MaSp2 repeat domain analogs have also been produced from \textit{P. pastoris} and spun into amyloid like nanofibrils, round micelles and films (Bogush et al., 2009).

\textit{Plant derived spidroins and spider silk-like fiber}
As an autotrophic eukaryotic host system, plant reactors are scalable, cost-effective and provide a choice of protein accumulation locations (Hauptmann et al., 2013; Scheller and Conrad, 2005). MaSp1 analogs 1f5 and 1f9 have been produced from tobacco leaf (Piruzian et al., 2003). Tissue specific expression other than in the leaves is usually accomplished by engineering a signal peptide that directs the analog protein into the specific tissue. MaSp1 analog can be delivered to the ER of tobacco, potato leaf, or tuber, resulting in fusion protein production at 2% of total soluble protein for proteins up to 100 kDa (Scheller et al., 2001). An elastin-MaSp1 repeat domain fusion protein has also been directed to tobacco and potato ERs, yielding 80 mg/kg leaf fusion protein (Scheller et al., 2004). The ER targeted elastin-like peptide and MaSp1 repeat fusion protein is retained in the ER of tobacco leaves and purified by membrane based method (inverse transition cycling) which can be made into polymer films (Hauptmann et al., 2015).

*Insect cell and mammalian cell derived spidroins and spider silk-like fiber*

Insect cell lines have been used for assessing solubility and fiber assembly of rADF-3 and rADF-4 spidroin analogs. rADF-4 spontaneously forms polymers in the cell. These polymers are easily purified, only accumulate as nanofibrils (Huemmerich et al., 2004a; Ittah et al., 2012) and provide a model system to study the association of fiber solubility with domains.

Mammalian cells such as bovine mammary epithelial alveolar cells and baby hamster kidney cells are prone to pathogens yet they produce spidroin
analogs rADF3 and MaSp1 that can be spun into fibers with comparable
toughness to natural spider silk fibers (Lazaris et al., 2002). A spidroin analog
has also been produced in the milk of transgenic mice at about 11.7 mg/L (Xu et
al., 2007).

Silkworm is a natural silk producer and can be cultivated. Expression of
chimeric spidroin fibroin protein in *Bombyx mori* cell lines (Miao et al., 2006;
Zhang et al., 2008) results in cocoon production of fusion protein in transgenic
silkworm. This method allows collection of recombinant fibers directly from the
cocoon. If materials in another shape or format is demanded, the fibers from the
cocoon have to be solubilized and then to be processed into other shapes. A
dragline silk spidroin and silkworm fibroin H-chain fusion protein has also been
produced in transgenic silkworm. The tenacity of the pulled hybrid silk from
transgenic silkworm cocoon was double that of native silkworm cocoon (Kuwana
et al., 2014). A silkworm fibroin gene with its coding region replaced by tandem
blocks of flagelliform elastic motif and MaSp2 repeats (Teulé et al., 2011) or
partial MaSp1 (Wen et al., 2010) has also been created and transformed into
silkworms. The produced chimeric proteins are spun into fibers with improved
toughness and extensibility. The silkworm fibroin *BmFib-H* gene knock out
silkworm strain is currently available. The production of silkworm fibroin in the
*BmFib-H* gene knock out strain is minimized so that it can produce recombinant
spidroin proteins to a larger extent (Ma et al., 2014).
Post translational modification and in vitro modification

Prokaryote host expression systems lack some of the post translational modification mechanisms that exist in eukaryote host expression systems. Numerous studies suggest that native spidroins may be phosphorylated and glycosylated after translation but fusion proteins produced from prokaryotes miss these functional groups. The fusion proteins from eukaryote hosts may be able to preserve the functional groups through host cell post-translational modification mechanisms.

Other post-translational modification involves artificially carrying out reactions of the functional group in the fusion protein either after the fusion protein has been translated or after purification. Via in vitro modification, elastin-like peptide (ELP) – MaSp1 repeat fusion proteins tagged with lysine or glutamine can be multimerized by transglutaminase to near native sized ELP-silk-like protein (Weichert et al., 2014). Fusion spidroin analogs are also able to be crosslinked by disulfide bonds (Grip et al., 2009). In vitro phosphorylation and dephosphorylation (Winkler et al., 2000) or reduction and oxidation (Szela et al., 2000) can control the engineered spidroin analogs’ aggregation.

Fiber assembly:

Spidroin analogs produced from almost all non-native host cells need to be converted into fibers or fibrils through artificial fiber assembly process, except transgenic silkworm cocoons. Recombinant ADF4 produced in sf9 cells can self-assemble into fibrils (Humenik et al., 2014b). Concentrated recombinant
flagelliform elastin motif fused with spidroin analog can form films when exposed to air and fibers can be drawn directly from the film (Teulé et al., 2007). Some recombinant spidroins require manual processing. Here, I discuss the methods developed for fiber assembly and the formation of other material formats.

Wet spinning:

Recombinant spidroin components have been artificially made into fibers through microfluidic extrusion of concentrated proteins into a coagulation bath. A detailed protocol has been published (Teulé et al., 2009). The process involves dissolving spidroin analogs in harsh solvent 1,1,1,3,3,3,-hexafluoro-2-propanol (HFIP) at a concentration relatively equivalent to the concentration of natural spinning dope and pump the solution into a methanol or isopropanol water coagulation bath. Fibers are readily drawn from the coagulation bath. Most spidroin analog fibers are spun this way (An et al., 2012; Geurts et al., 2010; Heidebrecht et al., 2015, Hsia et al., 2012). A second method does not require re-dissolving recombinant spidroins in HFIP. The concentrated recombinant spidroins are buffer exchanged and pumped directly into methanol coagulation bath and fibers are readily to be drawn from the coagulation bath (Lazaris et al., 2002).

A third method involves microfluidic devices that are cast to mimic the shear flow and extension flow of the spider gland duct. A system was designed with three consecutive modules mimicking the mixing of spinning dope with buffers and acids, the elongational flow in the tapered duct and the flow in the
narrow but even distal portion of the gland duct (Rammensee et al., 2008). In the processing of rADF3 and rADF4 spidroin analogs, the microfluidic systems show that if pH is lowered before the elongational flow begins, no fiber is assembled. Elevation of phosphate either before or right at elongational flow does not affect fiber assembly. A microfluidic device fabricating the gland flow was also built for micro-volume spinning dope (Kinahan et al., 2011) and a proposed microfluidic device by Rising and Johansson (2015) elaborates the recent findings in the spidroin self-assembly mechanisms (Rising and Johansson, 2015).

A fourth spinning process involves applying high voltage to the spinning dope (electrospinning). High voltage generates electrostatic force for the molecules in the spinning dope which escape from the surface tension forming a jet cone that “flies” to the electrode at the other end (Doshi and Reneker, 1995; Rutledge and Fridrikh, 2007). In a certain voltage range, the free charge is distributed at the spinning jet surface if a non-polar solvent is used and at the center if a polar solvent is used (Kornev, 2009). Computer modeled MaSp1 and MaSp2 repeat analog produced from *P. pastoris* can be electrospun into antiparallel β-sheet fibrils (Bogush et al., 2009). Poly(ethylene oxide) (PEO) may be added to the spinning dope to assist fiber formation through electrospinning and then later removed when soaked in water (Jin et al., 2002) without damaging the fiber scaffold (Baker et al., 2012).

*Diversified material shape and form*
The type of biomaterials that spidroin analogs are processed into goes beyond fibers. Films, adhesives, hydrogels, lyogels, sponges, foams, coatings and nanoparticles are under development and testing (Huemmerich et al., 2006; Jones et al., 2015; Schacht and Scheibel, 2011; Xu et al., 2013).

Nanostructure created by coupling rADF4 motifs with oligonucleotides can be self-assembled into β-sheet fibrils (Humenik et al., 2014a), with the potential of attaching a functionalized domain (Humenik and Scheibel, 2014). Hydrogel made from the spidroin analog rADF4 as bioink can be 3D printed into scaffolds for cell support (Schacht et al., 2015).

**Post-spin processing:**

Natural spidroin spinning dope contains about 70% water and dehydration happens along the spinning duct presumably by diffusion (Kojic et al., 2004). The dehydration process in artificial spinning processes involves a water bath or an aqueous alcohol bath accompanied with pulling the fibers manually. Post-spin drawing improves fiber mechanical features and β-sheet composition of the regenerated fibers from re-dissolved spidroins (Seidel et al., 2000). This is also seen in the post-spin water drawing of spidroin analogs and the overall toughness is increased two-to three-fold (Teulé et al., 2011).

**Application of spider silk fibers and spider silk fiber analogs**

Other than their extraordinary strength to meet the needs in material engineering, spider silks and analogs have become popular in biomedical
applications such as scaffolds in regenerative tissue culturing, crossing the membrane carrier for drug delivery, cancer cell targeting, coatings for implants, etc. The biocompatibility and aseptic handling of materials developed from spider silks and analogs are essential for these biomedical applications. Spider silks and analogs derived materials are hypoallergenic and no cell toxicity was detected (Dams-Kozlowska et al., 2013; Hauptmann et al., 2015). Spider silk analog fibers implanted into rat tissues have minimal inflammatory reactions (Gomes et al., 2012). For aseptic handling, spidroin analog fibers or particles can be autoclaved in water and sterilization does not change conformation or cytotoxicity (Lucke et al., 2015). The hypoallergenic nature of spider silks and analogs may lead to promising breakthroughs in the biomedical field (Widhe et al., 2012).

*Scaffold for tissue regeneration:*

Spider silks and analogs have been tested to provide framework support for regenerative medicines. This includes their use as artificial nerve grafts that provide guidance for peripheral neuron regeneration (Allmeling et al., 2008; An et al., 2015; Roloff et al., 2014) or neural stem cell differentiation (Lewicka et al., 2012), as a scaffold for bone mineral nucleation to produce artificial bone transplant (Cao and Mao, 2007) or skin cell lines to grow artificial skin substitutes (Wendt et al., 2011), and wound dressing (Baoyong et al.).

*Drug delivery*
Two spidroin-like analogs, 1F9 (MaSp1) and 2E12 (MaSp2) showed affinity to be incorporated into liposomes. Lipid binding causes a transient conformational change from polyproline II type to β-sheets and α-helixes to protect the hydrophobic core (Antonenko et al., 2010). rADF-4 or rADF-4 and polystyrene composite polymer films are able to encapsulate and release small polypeptides or low molecular compounds at ideal release kinetics (Agostini et al., 2015; Hardy et al., 2013) for controllable intake and release (Doblhofer and Scheibel, 2015).

A promising strategy in the battle against cancer could be to target tumor cells with particles that specifically bind to tumor cells and deliver therapeutic molecules to those cells. For example, particles made with MaSp1 analog fused with human epidermal growth receptor 2 (HER2) binding peptide showed greater binding affinity with HER2 expressing cells (Florczak et al., 2014). Spider silk fusion protein particles carry poly(l-lysine) motif (binds to pDNA) and tumor-homing peptides (THP) (recognizes target tumor cells) could deliver pDNA to the targeted melanoma cells (Numata et al., 2011).

Inflammation caused by implant surfaces may be minimized by a layer of hypoallergenic coating. Spider silk and analogs could be used as coating materials for silicone implant (Djedovic et al., 2015) to reduce capsule formation (Zeplin et al., 2014). An elastin-MaSp1 repeat domain fusion protein has been used to coat polystyrene plates for biocompatibility for tissue regeneration.
(Scheller et al., 2004) and glycopolymer coated rADF4 promoted cell binding (Hardy et al., 2014).

Features of spider silk and analogs can also be enhanced by mineralization. Calcium phosphate (Yang et al.), silica (Belton et al., 2012; Canabady-Rochelle et al., 2012; Zhou et al., 2015) or aluminum oxide (Lee et al., 2014) deposition on spider silk analogs improved the mechanical performance of the spun fibers.

Fusion protein created with spidroin analogs and another functional group produces functionalized fibers. The potential of the applications of functionalized fiber are promising and could be unlimited as new functional groups emerging. Spider silk analog protein and antimicrobial peptide fusion protein has been created into implant materials that can inhibit bacterial infection (Gomes et al., 2012). Scaffold made from spidroin analogs and a laminin cell-binding motif boost pancreatic islets survival.

Other than the biomedical applications, potential usage of spider silk and analogs includes thermal conductor (Zhang et al., 2014), optic sensors (Huby et al., 2013; Pal et al., 2015), sound channel (Mortimer et al., 2014) and air filter (Lang et al., 2013).

**Summary**

The advances in recombinant spider silk biotechnology and the versatile biomedical application make this ancient material an inspiration of many modern
innovations. Despite the efforts that have been expended to understand the mechanism of natural spider silk assembly and the efforts to bio-mimic this process, we still do not fully understand the details and are unable to reproduce this process given by nature. Many pieces in this puzzle are still missing.

This dissertation will focus on the production of mini-spidroins with the intact NTD and CTD flanking various abbreviated numbers of repeat domains (mini-spidroins). The availability of mini-spidroins with intact terminal domains will be used to develop scalable fiber production and to investigate fiber formation and the assembly mechanism. Fiber mechanical and physical property will be evaluated.

Chapter Two will demonstrate gene construction of mini-spidrion based on native *N. clavipes* MaSp1A and MaSp2 N-, C- terminal domain and 8, 16 or 32 of the consensus repeat domain of each spidroin; transformation confirmation; affinity purification of mini-spidroins from transgenic tobacco *Nicotiana tabacum* leaves and the collection of concentrated mini-spidroins.

Chapter Three will demonstrate mini-spidroin fiber assembly through polyelectrolyte complex formation, which takes advantage of static charge attraction of opposite charged polyions. Tensile strength of the spun fibers has been recorded. Fiber physical appearance and optical properties have been examined by microscopy. Dispersing mini-spidroins in chitosan matrix, which emulates the soluble chitinous material detected in spider major ampullate duct, results in meters long fibers. The fiber can be pulled endlessly if the
protein/chitosan matrix is continuously supplied. Directing the spun fiber into a water bath and pulling it from the water bath against the glass wall removes excessive water and improves mechanical performance of the fibers.

Chapter Four will demonstrate the cloning and characterization of peroxidase that was identified from the EST sequence of *N. senegalensis* (*NsPox*). The synthesized *NsPox* gene was codon optimized and was subcloned into the pMAL-c2x vector. After sequence confirmation, pMAL-c2x with an NsPox insert was transformed into *E. coli* BL21 cells. NsPox will be purified through affinity purification and subsequent proteinase cleavage. NsPox enzyme activity and possible NsPox catalyzed N-terminal domain crosslinking will be investigated.
REFERENCE


CHAPTER TWO

SPIDER SILK-LIKE PROTEINS DERIVED FROM
TRANSGENIC NICOTIANA TABACUM

This preparation is submitted for review to Transgenic Research with the following authorship: Congyue Annie Peng, Julia Russo, Charlene Gravgaard, Heather McCartney, William Gaines, William R. Marcotte, Jr.
INTRODUCTION

Spiders can make up to six different proteinaceous silks that display a variety of remarkable physicochemical properties (Altman et al. 2003; Gosline et al. 1986). Spider silks, particularly dragline silk, have attracted much attention for their potential exploitation in the development of new materials (Allmeling et al. 2008; Hedhammar et al. 2010; Schacht and Scheibel 2011). The combination of strength and elasticity in dragline fibers suggests that materials based on these fibers will outperform currently available man-made materials such as steel, nylon and Kevlar (Gosline et al. 1999). Spider silks are also biocompatible and biodegradable which makes them ideal materials for a variety of biomedical applications (Hedhammar et al. 2010). However, to take advantage of these remarkable qualities, sufficient amounts of the constituent proteins will need to be readily available. This cannot be accomplished by harvesting the spinning dope from spiders as a single spider produces less than 1 mg per day (Seidel 1998). In addition, farming of spiders is impractical due to territorial and cannibalistic behavior (Nentwig, 2013).

The spider silk with the greatest mechanical toughness is major ampullate or dragline silk. It is made of two large, repetitive proteins, the major ampullate spidroins 1 and 2 (MaSp1 and MaSp2) (Hinman and Lewis 1992; Xu and Lewis 1990). The full length DNA sequences that encode these proteins have been identified (Ayoub et al. 2007). The general architecture of these
proteins is similar in that both have large, central glycine- and alanine-rich repeat domains that consist of approximately 100 copies of shorter imperfect repeat blocks (Bini et al 2004; Ayoub et al 2007). MaSp1 and MaSp2 repeat blocks differ in the length (~30 and ~40 amino acids, respectively) and the nature of the repeat motifs. While each repeat block in MaSp1 and MaSp2 contains a poly-alanine tract, the amino acid triplet GGX is common in the repeat block of MaSp1 whereas MaSp2 contains proline rich motifs (GPGGY and GPGQQ) in the repeat domain (Bini et al. 2004). The repeat domain of MaSp1 is virtually devoid of proline (Hinman and Lewis. 1992). The repeat domains of other silks (e.g. tubuliform and flagelliform silks), while conserved within and across species, differ dramatically from the MaSps (Huang et al. 2006). The differences in repeat domains have been shown to be directly associated with the differences in tensile properties of various spider silks (Xia et al. 2010)

The repeat domains are flanked on both the N- and C-terminal sides by short non-repetitive sequences of approximately 150 amino acids each (N- and C-terminal domains; NTD and CTD). The NTDs and CTDs are conserved not only between MaSp1 and MaSp2 but also among other silks within species and across species (Beckwitt and Arcidiacono 1994; Garb et al. 2010; Motriuk-Smith et al. 2005). This suggests that these non-repetitive domains have a common function across the silking making spiders and it may relate to silk self-assembly process.
Recent studies on the functional role(s) of the NTD and CTD strongly suggest they are involved in regulating the self-assembly process. The CTD forms homodimers through one disulfide bond and two salt bridges (Ittah et al. 2007). Changes at the dimer interface and salt bridge region are critical in protein transition from soluble protein to solid fiber. NTD also forms homodimer through salt bridges (Hagn et al. 2010). The strength of the salt bridges is relevant to the ionic strength of the sodium chloride solution as demonstrated by a computer simulation study (Gronau et al. 2013). Reducing pH induces a local structure change that helps to stabilize NTD dimerization (Gaines et al. 2010) and initiate protein secondary structure transition from coil to β-sheet (Dicko et al. 2004). Inclusion of this domain in the recombinant protein construct may provide a valuable control point of the assembly process that is lacking in repeat domain-only fiber production. In addition, one or both terminal domains may also function as solubility enhancers thereby allowing the high protein concentrations found in the lumen of the spider silk gland (Huemmerich et al. 2004). While it has been demonstrated that fibers can be formed from repeat-only domains, inclusion of the NTD and CTD may help to stabilize proteins expressed in heterologous systems and may also allow the development of novel spinning technologies (Ittah et al. 2006).

Due to the limitations inherent in harvesting silks directly from spiders, there has been widespread use of genetic engineering for the production of recombinant spidroin-like proteins. The expression systems employed thus far
include prokaryotes such as *E. coli* and *Salmonella*: (Arcidiacono et al. 1998; Lewis et al. 1996; Widmaier et al. 2009; Xia et al. 2010); lower eukaryotes such as yeasts (Fahnestock and Bedzyk 1997; Gaines and Marcotte 2011); transgenic plants such as *Arabidopsis*, potato and tobacco (Scheller et al. 2001; Yang et al. 2005); mammalian cell lines (Lazaris et al. 2002); insects (silkworm: (Miao et al. 2006; Teulé et al. 2011), and mammals (mice and goats) (Service 2002; Xu et al. 2007). Each of these approaches has met with some success although none has yet surmounted the problem of scalable expression and purification of recombinant spidroin-like proteins for materials development.

The use of a large biomass plant like tobacco (Scheller et al. 2001) provides advantages over other approaches for recombinant spidroin-like protein production. With the demonstration that these proteins can be purified from crude plant extracts (see below), optimization of the processes involved will make the approach essentially unlimited in scalability.

Here, we present the assembly of mini-spidroin genes that encode native MaSp1 and MaSp2 NTD and CTD sequences flanking an abbreviated number of the respective consensus repeat domains. When expressed, the mini-spidroin protein is tagged with a self-cleavable intein fused to a chitin binding domain (CBD) (Chong et al. 1997). We demonstrate that upon introduction into *Nicotiana tabacum*, the genes are expressed and mini-spidroin proteins can be purified by chitin affinity chromatography and intein activation. Mini-spidroins were dialyzed and concentrated through freeze-drying, resulting in gelatin-like liquids. Immuno-
detection of these liquids showed various degree of polymerization of mini-
spidroin.

MATERIALS AND METHODS

Mini-spidroin gene construction and generation of transgenic plants

A synthetic multi-cloning site (MCS), generated by annealing and
extension of overlapping oligonucleotides, was used to replace the MCS in
pT7T3α-19. The synthetic MCS was used to assemble the mini-spidroin genes
and contained the following restriction enzyme sites: HindIII-Ascl-Ncol-Xbal-
Xmal/Smal-Agel-NgoMIV-BgIII-BamHI-NotI-EcoRI. Isolation of native NTD
sequences for MaSp1A and MaSp2 has been described previously (Gaines and
Marcotte, 2008). Shorter fragments containing the native NTD sequences
lacking the predicted signal peptide were generated using primers described in
Table 2.1 such that each NTD contained a Smal/Xmal site at one end and an
Agel site at the other end.

Alignment of published MaSp1 and MaSp2 repeat domains provided a
consensus amino acid sequence for each:

\[ NH_2-GGAGQGGYYGGLGGQGAGRGGQGAGAAAAAA-COOH \] for MaSp1

\[ NH_2-GPGQQGPGGYGPGQQGPGGYGPGQQGQGPGQGQGSAA- \]
\[ COOH \] for MaSp2
Reverse translation provided a nucleotide sequence optimized for tobacco codon usage. Consensus nucleotide repeats were created by annealing and extension of overlapping oligonucleotides described in Table 2.1 such that the repeat domains contained an *Age*I site at one end and an *NgoMIV* site at the other.

Isolation of native CTD sequences for MaSp1A and MaSp2 has been described previously (Gaines and Marcotte 2008). Shorter fragments containing the native CTD sequences were generated using primers described in Table 2.1 such that each CTD contained an *NgoMIV* site at one end and an *BgII* site at the other end.

The *Mxe* gyrA intein and adjacent CBD were amplified from plasmid pTWIN1 (New England Biolabs) using primers described in Table 2.1 to produce a fragment that contained a *BgII* site on one end and a *BamHI* site on the other end. After amplification and cloning, the single *NgoMIV* and *Age*I sites contained in the region encoding the intein-CBD were mutagenized to facilitate later cloning. These changes result in a single amino acid change in the region containing the intein and CBD.

The nucleotide fragments described above were assembled into the synthetic MCS. The translation start and stop codons were provided by the *NcoI* site in the MCS and the pTwin1-derived intein-CBD fragment, respectively. To generate multimers of the repeat domains, monomeric copies of the repeat sequence (containing *Age*I and *NgoMIV* ends) were gel-purified, self-ligated and
subsequently re-digested with Agel and NgoMIV. Tetramers of the repeat domain were then gel-purified and ligated between the NTD and CTD coding regions. For larger order multimers (8 copies and 16 tandem copies of the repeat domain), shorter multimers were pyramided on top of one another.

Plant expression plasmids pKM12 (Dey and Maiti 1999) and pKLP36 (Maiti and Shepherd 1998) were obtained from Dr. Indu Maiti (Kentucky Tobacco Research and Development Center). Both contain caulimovirus promoter regions (mirabilis mosaic and peanut chlorotic streak, respectively), an MCS and rbcS terminator as well as a kanamycin cassette for selection of transgenic plants. The expression plasmid MCS was replaced with the synthetic MCS described above to generate pKM12’ and pKLP36’. Assembled mini-spidroin constructs were inserted into pKM12’ and pKLP36’ as Ascl/NotI fragments. Constructs were named for the mini-spidroin and repeat number (e.g. pKM12-Sp1R8 contains the MaSp1 NTD, eight copies of the Sp1 repeat domain and the MaSp1 CTD followed by the intein CBD domain). Plasmid pKM12’ and pKLP36’ containing either MaSp1 (with eight or 16 copies of the repeat domain) or MaSp2 (with eight, 16, or 32 copies of the repeat domain) were generated and introduced into Agrobacterium. Agrobacterium-mediated transformation into Nicotiana tabacum and selection were essentially as described (Fisher and Guiltinan 1995). Primary transformants (T0) were selected on kanamycin-containing medium, transferred to soil and grown to maturity. Seeds from T0
Plants were germinated on kanamycin-containing medium and the seedlings (T1) were transferred to soil.

Protein extraction and purification

Leaf tissues (routinely multiples of 50 g) were pulverized either manually or using a blender in Tris-buffered saline (1X TBS; 20 mM Tris-HCl pH 7.4, 140 mM NaCl) supplemented with 0.1% Tween 20 (1X TBST) and the protease inhibitors PMSF (1 mM) and TLCK (0.1 mM). The ratio of tissue (T, in g) to extraction buffer volume (E, in ml) was 1:10 (T:E) unless noted otherwise. For small scale extractions and screening purposes, four 0.5 cm leaf punches were used. Tissue debris was removed from the crude extracts by either centrifugation (for leaf punches) or filtration through Miracloth (EMD Millipore) followed by 3 MM filter paper. The chitin affinity purification scheme is outlined in Fig 1B, 1C. Chitin resins (New England BioLabs Inc.) were washed with 10x resin volumes of nano-pure water followed by 5x resin volumes of 1X TBS, pH 7.4 prior to addition to the crude extracts (1-2 ml resin/100 ml crude extract) and incubated at 4°C overnight with mixing. The following day, the chitin beads were collected and washed with ≥ 20 volumes of 1X TBST until most or all of the chlorophyll was visually removed. Mini-spidroins were released from the fusion proteins by incubating the chitin resins with three resin volumes of intein activation buffer (20 mM Tris-HCl, pH 8 supplemented with 500 mM NaCl, 30 mM DTT and 0.1% Tween 20) at 16°C for 40 hours with mixing. Eluate was collected and the chitin
beads were washed twice with one resin volume each of 1X TBST. Eluates were pooled and concentrated 10-fold using Pellicon XL 50 Ultrafiltration Cassettes (EMD Millipore). The concentrated mini-spidoins were dialyzed against 5 mM ammonium bicarbonate at 4 ºC overnight. Dialysis was continued for an additional 8 hours with fresh dialysis buffer every four hours. The mini-spidoins were recovered from dialysis tubing and pre-frozen at -80°C. Freeze drying was performed using a Labconco lyophilizer (FreeZone 2.5) with the temperature and vacuum settings at -50°C and 0.08 mBar, respectively, until all visible ice crystals were gone. The resulting material was a viscous gelatin-like fluid.

Immuno-detection

Samples were separated on 8% SDS-polyacrylamide gels and electro-transferred to PVDF membranes. Recombinant MaSp1A-NTD was used to generate a custom polyclonal antibody (Rockland Immunochemicals, Inc.). The crude antibody was affinity-purified using Affi-10-immobilized MaSp1A-NTD and then cross-adsorbed against total protein isolated from non-transgenic tobacco (Affi-10 and Affi-15 immobilized). The resulting antibody recognizes both MaSp1 and MaSp2 NTDs and was used as the primary antibody. The secondary antibody was goat anti-rabbit conjugated with alkaline phosphatase. The detection reagent for alkaline phosphatase was Lumi-Phos WB (Pierce) and membranes were imaged using a Fujifilm LAS-1000plus imager.
RESULTS

Mini-spidroin coding regions were assembled as described in Materials and Methods. The encoded mini-spidroins consisted of MaSp1 or MaSp2 NTD and CTD sequences flanking either 8, 16 or 32 copies of the respective repeat domain and followed by the intein-CBD (Fig 2. 1A). Attempts to generate larger repeat domains were unsuccessful. Because the coding regions differ only in the number and nature of the repeat domains, these are hereafter referred to as rMaSp1R8, rMaSp1R16, rMaSp2R8, rMaSp2R16 and rMaSp2R32. All cloning junctions were confirmed by sequencing. At least 12 T1 plants for each construct were evaluated for the presence of the transgenes using PCR for both NTD and intein-CBD regions and all were positive.

Purified rMaSp1A-NTD protein was used to generate and affinity purify an anti-NTD antibody. The purified antibody detects rGST-MaSp1A NTD (Sup. Fig. 2.1, Lane 1) but does not cross react with any protein from wild-type tobacco leaf crude extract (Sup. Fig. 2.1, Lane 2). We have been able to produce rMaSp1 proteins that contain 8 and 16 copies of the Sp1 repeat domain and rMaSp2 proteins containing 8, 16, and 32 copies of the Sp2 repeat domain in transgenic tobacco. At least two T1 plants per construct (up to six) were tested for protein expression by immunoblot. For each construct, at least two protein expression-positive lines were confirmed. The protein-positive T1 plants were grown to maturity and T2 plants were used for all subsequent analyses.
Fig. 2.2 shows the antibody readily detects both rMaSp1 and rMaSp2 proteins containing 8 and 16 copies of the repeat domains in crude extracts from transgenic plants. Interestingly, it appears that some proportion of the C-terminal intein of the rMaSp2 fusion proteins is spontaneously activated, resulting in self-cleavage of intein CBD domain from the fusion protein at some point in the extraction process. This is evidenced by the lower bands in the rMaSp2 transgenic lanes (Fig. 2.2, Lanes 2 and 4) that are consistent with removal of ~28 kDa (the size of the intein-CBD domain is 27.8 kDa). It is also possible to see some bands at higher molecular weights, particularly in the rMaSp2 samples, that likely correspond to multimers of either full-length fusion proteins, mini-spidroins lacking the intein-CBD or combinations thereof.

Fig. 2.3 shows immuno-detection of rMaSp2 proteins containing 32 copies of the repeat domain in crude extracts from both pKM12 and pKLP36 transgenic lines. It seems that spontaneous activation of intein-CBD in the case of rMaSp2R32 protein is much less than that seen for the rMaSp2R8 and rMaSp2R16 proteins. The signal strength of the rMaSp2R32 bands of the two individual pKLP36 lines (Fig. 2.3, Lanes 2, 3) indicates the protein amount varies in each transgenic line even when under control of the same promoter.

We also assessed our ability to affinity purify the recombinant mini-spidroins and subsequently remove the CBD tag by intein activation (Fig. 1B, 1C). Fig. 2.4 shows immunoblots of representative rMaSp1R8 and rMaSp2R8 samples (panels A and B, respectively) taken at various points in the purification
process. It is variable to see laddering in different extractions of the rMaSp proteins, particularly in the rMaSp2 samples. It is not clear at this time if this is due to differences in individual extractions from a given plant or to differences among sibling plants in a particular line. Regardless, it is clear that the crude extract was essentially depleted of full-length fusion protein post-chitin bead binding (Fig. 2.4, Lanes 2 and 3). It is also possible to demonstrate that the full-length fusion proteins are efficiently bound to the chitin beads (Fig. 2.4, Lanes 4). We routinely note the presence of cleaved fusion protein in the chitin bead fraction for the rMaSp2 proteins suggesting that additional spontaneous intein activation is occurring after binding but prior to incubation with intein activation buffer. After activation of the intein and concentration of the pooled eluates/wash fractions, the majority of the protein is present as monomer rMaSp protein (Fig 2.4., Lanes 5). There was some residual protein that remained bound to the beads post-cleavage (Fig 2.4 compare Lanes 4 and 6). Expression levels for untargeted (cytoplasmic) mini-spidroin yields in tobacco leaves are estimated to be 0.7% and 1.9% TSP for rMaSp1R8 and rMaSp2R8, respectively.

Purified mini-spidroins were dialyzed and lyophilized. After extensive freeze drying, the mini-spidroins formed viscous gelatin-like fluids. Immuno-detection shows an immunoblot of samples taken before dialysis, post dialysis but before freeze-drying and post freeze-drying. The post freeze-drying fluids shows highly concentrated rMaSp1R8 protein spontaneously forms aggregates (Fig. 2.5, Lane 4). This observation is consistent with what is seen with highly
concentrated spinning dope isolated directly from major ampullate glands (Sponner et al., 2004; Sponner et al., 2005; Guehrs et al, 2008).

DISCUSSION

Plants are attractive hosts for the production of recombinant proteins for a variety of reasons including increased likelihood of appropriate post-translational modifications compared to prokaryotic expression systems and elimination of the risk of contaminating human/mammalian pathogens which is a serious concern in many eukaryotic expression systems. Plants expressing significant levels of recombinant proteins also have the potential to provide a cost-effective way to scale up production.

However, target protein production levels can vary greatly across different transgenic lines (Fig. 2.3) and even among individuals within a single transgenic line. In some cases, recombinant protein is not detectable after a few generations even when the seedlings are germinated under selection. This may be due to gene silencing, a common obstacle for plant-derived proteins and can be especially true if the gene contains highly repetitive sequences (Stam et al. 1997). And, as has been noted recently, the cost and efficiency of downstream technical aspects of “molecular farming” can present significant challenges. In an elastin-like peptide (ELP)–MaSp1 repeat fusion protein study, transglutaminase was required to crosslink lysine and glutamine residue in vitro.
to produce a multimerized nearly native sized ELP-silk-like protein (Weichert et al. 2014).

Despite these concerns, plants continue to be a frequent choice for heterologous gene expression. Not surprisingly, targeting recombinant protein production to various locations/organelles in plants has been shown to have an effect on expression levels. Expression of a spider silk-like protein (tandem MaSp1 repeat domains) in either the apoplast or ER lumen of Arabidopsis leaves resulted in 8.5% and 6.7% total soluble protein (TSP), respectively (Yang et al. 2005). That same study found that expression levels in seeds can be even higher (18% TSP in the ER lumen). Our yields in tobacco leaves are estimated to be 0.7% and 1.9% TSP for rMaSp1R8 and rMaSp2R8, respectively. This is significantly lower than the levels cited above for Arabidopsis but our expression levels are comparable to the 2% TSP yield of an ER-targeted MaSp1 repeat domain construct produced in tobacco leaves (Scheller et al. 2001).

Full-length recombinant proteins are readily detectable in crude leaf extracts (Fig 2). Interestingly, the intein domain in the rMaSp2-R8 and rMaSp2-R16 fusion proteins appears to spontaneously activate at some level. This is evidenced by the presence of a polypeptide containing the MaSp2-NTD at a molecular weight consistent with removal of the intein-CBD domain (Fig 2, Lanes 2 and 4; Fig 4B, Lanes 2 and 4). The rMaSp2-R32 extracts also contain a lower MW band (Fig 2.3) consistent with loss of the intein-CBD but the activity is much lower. While the rMaSp1-R8 and -R16 samples in Fig. 2.2 do not display this
same activity, it was observed in some rMaSp1 samples (Sup. Fig. 2.2). It is not clear if this activity is present *in planta* but, if that is the case, one might expect all (or at least most) of the recombinant protein to be cleaved. There is clearly full-length protein present in all crude extracts, suggesting that intein activation may be occurring at some point in the extraction/purification procedure or there may be a unique structural arrangement that can form under some circumstances in the rMaSp fusions that leads to intein activation/cleavage.

We also see rMaSp-NTD positive bands at higher molecular weights in the crude leaf extracts and at some steps in the purification process that likely correspond to self-assembled multimers. This type of aggregation, that may also display higher molecular bands, was reported by others (Menassa et al. 2004; Sponner et al. 2004, Sponner et al. 2005; Guehrs et al. 2008) and these could represent associations of full-length fusion proteins, recombinant proteins lacking the intein-CBD or combinations thereof.

As part of the development of our purification protocol, we compared hand and blender grinding of leaf tissue at two different tissue:extract ratios (T:E = gram:ml). Based on immuno-detection (Sup. Fig. 2.3), hand grinding at a T:E of 1:5 is slightly better than the other conditions tested and there was little difference in the laddering observed. The lower leaf tissue to buffer ratio does reduce the amount of extraction buffer which would be an advantage in scalable productions. Other than multimerization and intein spontaneous cleavage, protein degradation is minimized in our current process.
The rMaSp proteins expressed here are qualitatively different than those in many other studies in that we have incorporated native spider silk NTD and CTD domains (in addition to the intein-CBD self-cleavable affinity tag). Spider silk-like proteins with only the repeat domain frequently display solubility issues due to the high hydrophobicity of the repeat domains or premature aggregation (Rammensee et al. 2008; Zhang et al. 2008) whereas our mini-spidroins with flanking NTD and CTD seems to be readily soluble. Dramatically, the highly-concentrated, purified proteins appear to retain water and remain as a gelatin-like fluid instead of powdered form after lyophilization. The NTD and CTD were speculated to enhance solubility of spidron proteins (Gao et al. 2013) (Huemmerich et al. 2004), and our observations are consistent with that hypothesis.

The availability of both MaSp1 and MaSp2 mini-spidroins with variable copies of the repeat domains enables us to evaluate their assembly into fibers as individual proteins and in various combinations. Thus, we can examine the distinct features of MaSp1 and MaSp2 proteins and the effect of repeat domain numbers on the mechanical properties of a broad range of biomaterials including fibers, hydrogels, coating materials, etc. These materials are likely to have many applications in medical and engineering related fields.
Table 2.1. Oligonucleotide primers. Terminal restriction sites are underlined. Internal overlap of repeat domain primers are in bold.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaSp1A NTD F</td>
<td>5’-GCCCCCGGGCAAAACACCCCATGGTCA-3’</td>
</tr>
<tr>
<td>MaSp1A NTD R</td>
<td>5’-GCCCACCGGTAACCACCGTAAGATACTTCATT-3’</td>
</tr>
<tr>
<td>MaSp2 NTD F</td>
<td>5’-GGCCCGGGCAAGCAAGGTCTCCATGG-3’</td>
</tr>
<tr>
<td>MaSp2 NTD R</td>
<td>5’-GCCACCGGTACCAGCGGTATGATACTTCATT-3’</td>
</tr>
<tr>
<td>MaSp1 Repeat F</td>
<td>5’-GTCGCCGGTTGGAGCTGGACAGGAGATATGGAGGACACCTGGAGGTAGGTCGGATGAGG-3’</td>
</tr>
<tr>
<td>MaSp1 Repeat R</td>
<td>5’-CTCCCTCTACCTCGACCCCTGAC-3’</td>
</tr>
<tr>
<td>MaSp2 Repeat F</td>
<td>5’-GCCACCGGGTGACAGCGACAGGAGGTCTCCAGGCTCTTGAGGTTACGGACCAG-3’</td>
</tr>
<tr>
<td>MaSp2 Repeat R</td>
<td>5’-GCCACCGGGTGACAGCGACAGGAGGTCTCCAGGCTCTTGAGGTTACGGACCAG-3’</td>
</tr>
<tr>
<td>MaSp1 CTD F</td>
<td>5’-CGGCCGGGCCCTCGCCAGCTGCTT-3’</td>
</tr>
<tr>
<td>MaSp1 CTD R</td>
<td>5’-GCCGATCTATGATGATGATGATGATGCTCGAGACHAAGAGCTTGATTGAGGTC-3’</td>
</tr>
<tr>
<td>MaSp2 CTD F</td>
<td>5’-GGCCGGGCTCCGCTGCAGGCTCTTCTC-3’</td>
</tr>
<tr>
<td>MaSp2 CTD R</td>
<td>5’-GCCGATCTATGATGATGATGATGCTCGAGAAATGGC-3’</td>
</tr>
<tr>
<td>Intein-CBD F</td>
<td>5’-GCCGATCTTTCCAATATTGCTACGAGGAGATGCA-3’</td>
</tr>
<tr>
<td>Intein-CBD R</td>
<td>5’-GCCGATCCCCTTCCTG-3’</td>
</tr>
</tbody>
</table>
Figure. 2.1 Diagrammatic representation of recombinant mini-spidroin protein domain architecture and purification strategy. A, mini-spidroin structure. Each construct contains NTD and CTD flanking repeat domain multimers (R8 shown) followed by the intein-CBD domain used for protein purification. B and C, purification strategy. rMaSp-intein-CBD protein pre-intein activation (B) and post-intein activation (C).
Figure 2.2 Immuno-detection of recombinant mini-spidroins in crude leaf extracts.
Leaf punches were hand-ground in microfuge tubes in the presence of 250 µl of extraction buffer. Each lane represents 25 µl of crude extract. Band corresponding to full length MaSp-intein-CBD fusion protein marked by asterisk Lane 1, rMaSp1R8 (73 kDa); Lane 2, rMaSp2R8 (81 kDa); Lane 3, rMaSp1R16 (92 kDa); Lane 4, rMaSp2R16 (108 kDa). Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A NTD. Molecular weight marker sizes (in kDa) are shown to left.
Figure 2.3 Immuno-detection of recombinant mini-spidroin rMaSp2R32 in crude leaf extracts. Leaf punches were hand-ground in microfuge tubes in the presence of 250 µl of extraction buffer. Each lane represents 25 µl of crude extract from individual transgenic tobacco lines. rMaSp2R32 is 162 kDa and 134 kDa, with and without the intein-CBD, respectively. Band corresponding to full length MaSp-intein-CBD fusion protein marked by asterisk Lane 1, pKM12'-Sp2R32 -1.2.1); Lane 2, pKLP36'-Sp2R32-2.3; Lane 3, pKLP36'-Sp2R32-3.5. Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A-NTD. Molecular weight marker sizes (in kDa) are shown to left.
Figure. 2.4 Purification of recombinant mini-spidroins from crude leaf extracts. A, rMaSp1R8; B, rMaAp2R8. Lane 1, positive control rGST-MaSp1A NTD fusion protein (43 kDa), 12 ng; Lane 2, crude extracts (25 µl); Lane 3, crude extract post chitin binding (25 µl); Lane 4, chitin beads post binding (15 µl of 50% slurry); Lane 5, pooled intein cleavage buffer and wash post concentration (5 µl); Lane 6, chitin beads post intein activation and wash (15 µl of 50% slurry). Band corresponding to full length MaSp-intein-CBD fusion protein marked by asterisk. Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A NTD. Molecular weight marker sizes (in kDa) are shown to left.
Fig. 2.5 Purified rMaSp1R8 showed various degree of aggregation after freeze-drying. Lane 1, positive control rGST-MaSp1ANTD fusion protein (43 kDa), 12 ng; Lane 2, Pooled rMaSp1R8 in cleave buffer; Lane 3, Pooled rMaSp1R8 dialyzed against 5 mM ammonium bicarbonate; Lane 4, rMaSp1R8 post freeze-drying at -45°C. Molecular weight marker sizes (in kDa) are shown to left.
Supplementary Figure 2.1 Immuno-detection of recombinant mini-spidroins in crude leaf extracts of non-transgenic tobacco and transgenic tobacco. Leaf punches were hand-ground in microfuge tubes in the presence of 250 µl of extraction buffer. Each lane represents 25 µl of crude extract. Lane 1, positive control rGST-MaSp1A NTD fusion protein (43 kDa), 12 ng; Lane 2, non-transgenic tobacco; Lane 3, rMaSp1R8 (73 kDa); Lane 4, rMaSp2R8 (81 kDa). Band corresponding to full length MaSp-intein-CBD fusion protein marked by asterisk. Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A NTD. Molecular weight marker sizes (in kDa) are shown to left.
Supplementary Figure 2.2  Immuno-detection of recombinant mini-spidroin rMaSp1R16 in crude leaf extracts. Leaf punches were hand-ground in microfuge tubes in the presence of 250 µl of extraction buffer. Each lane represents 25 µl of crude extract. Lane 1, pKM12’ - Sp1R16 - 3.1 (92 kDa); Lane 2, pKM12’ - MaSp1R16 - 7.6 (92 kDa); Lane 3, pKM12’ - MaSp1R16 - 8B.5 (92 kDa); Lane 4, pKLP36’ - MaSp1R16 - 10.2 (92 kDa). Band corresponding to full length MaSp-intein-CBD fusion protein marked by asterisk. Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A NTD. Molecular weight marker sizes (in kDa) are shown to left.
Supplementary Figure 2.3  Manual grinding and blender extraction of rMaSps at various tissue (T):buffer (E) ratios. Tissue samples were ground either by hand or in a blender and 15 µl of extracts was applied per lane. Lane 1, rMaSp1R8 manually ground at T:E of 1:10; Lane 2, rMaSp1R8 manually ground at T:E of 1:5; Lane 3, rMaSp1R8 blender ground at T:E of 1:10; Lane 4, rMaSp1R8 blender ground at T:E of 1:5; Lane 5, rMaSp2R8 manually ground at T:E of 1:10; Lane 6, rMaSp2R8 manually ground at T:E of 1:5; Lane 7, rMaSp2R8 blender ground at T:E of 1:10; Lane 8, rMaSp2R8 blender ground at T:E of 1:5. Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A NTD. Molecular weight marker sizes (in kDa) are shown to left.

<table>
<thead>
<tr>
<th></th>
<th>rMaSp1R8</th>
<th></th>
<th>rMaSp2R8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

Primary antibody was affinity-purified polyclonal rabbit anti-rMaSp1A NTD. Molecular weight marker sizes (in kDa) are shown to left.
References


CHAPTER THREE

POLYELECTROLYTE FIBER ASSEMBLY OF PLANT DERIVED SPIDER SILK-LIKE PROTEIN
INTRODUCTION

Spider silks are natural proteinaceous biomaterials, durable, strong and hypoallergenic. The toughness of a native *Nephila clavipes* dragline silk collected from Georgetown, South Carolina is $518.0 \pm 202.4 \text{ MJ/M}^3$ (sample number N=13), with failure stress ($\sigma_{\text{max}}$) $4.65 \pm 1.13 \text{ (GPa)}$ (Marcotte, unpublished data). It is forty times more than that of the high tensile steel and five times more than that of the man-made material Kelvar. The major protein components of dragline silk and their properties have been reviewed extensively (Atkins, 2003; Lewis, 2006; Omenetto and Kaplan, 2010). Two types of proteins, namely Major ampullate Spidroin 1 (MaSp1) and Major ampullate Spidroin 2 (MaSp2) were found in *Nephila clavipes* major ampullate dragline silk (Hinman and Lewis, 1992; Xu and Lewis, 1990). Depending on the feeding condition, MaSp1 content can be five times as much as the MaSp2 protein in the spider silk (Guehrs et al., 2008).

The average cost to produce MaSp1 is 14.5 ATP/aa and the average cost to produce MaSp2 is 17.0 ATP/aa (Guehrs et al., 2008). The higher cost for MaSp2, due to the presence of proline in the repeat domain, forces starved spiders to produce less MaSp2 protein to adapt to the poorly fed condition. Spidroins are produced as a highly concentrated solution of unfolded $\alpha$-helices in the storage sac of the ampullate gland but fold into $\beta$-pleated sheets as the spinning dope funnels into the tapered spinning duct (Knight et al., 2000). Spidroins are large proteins (250-300 kDa) composed of roughly 100 non-perfect repeat domains flanked by short non-repetitive N- and C- terminal domains (Ayoub et al., 2007).
MaSp1 and MaSp2 spidroins have distinct repeat domains. For MaSp1, the consensus repeat motifs are (GGX)$_n$, (GX)$_n$, (GA)$_n$ and poly-alanine (Xu and Lewis, 1990). For MaSp2, the consensus repeat motifs are (GPGXX)$_n$ and poly-alanine (Hinman and Lewis, 1992;). The poly-alanine motifs assemble into β-sheet crystals during fiber assembly and are the major contributor to fiber strength (Bini et al., 2004; Hayashi et al., 1999; Parkhe et al., 1997).

The non-repetitive N- and C-terminal domains (NTD and CTD, respectively) are conserved among species that could be traced back to ancient times (Beckwitt and Arcidiacono, 1994; Garb et al., 2010). Unlike the repeat domain, most of the amino acids in the NTD and CTD have hydrophilic side chains, consistent with the high protein concentration formed in the storage sac of ampullate gland. CTDs form homodimers through inter-molecular salt bridges and the monomers are covalently connected by a disulfide bond (Sponner et al., 2005). The CTD plays an important role in β-sheets alignment during fiber formation (Ittah et al., 2006; Ittah et al., 2007). NTDs also homodimerize and this association is enhanced through a conformational change in response to the decreased pH in the ampullate gland duct (Askarieh et al., 2010; Gaines et al., 2010). This leads to the formation of long multimeric strands that then coalesce into fibers through extensive hydrophobic interactions (Askarieh et al., 2010).

Spidroin post-translational phosphorylation may also contribute to fiber assembly. NMR $^{31}$P spectrum and amino acid analysis of $N$. clavipes dragline silk detected phosphotyrosine from the dragline silk fiber (Michal et al., 1996). Eight
tyrosine and serine phosphorylation sites that are confirmed by phosphatase
treatment have been mapped to MaSp1 sequences that cover the entire repeat
domain (dos Santos-Pinto et al., 2014). The possibility of enzyme catalyzed
phosphorylation/dephosphorylation to control β-sheet assembly of a spidroin-like
repeat domain suggests that post-translational phosphorylation may enhance
spidroin solubility by preventing the hydrophobic interaction of the alanine core
region (Winkler et al., 2000).

The use of spider silks as promising biomaterials will require mass
production of spider silk-like proteins. Recombinant protein bioengineering may
reach large scale and maintain a stable supply of the material. We previously
demonstrated affinity purification of recombinant spider silk-like mini-spidroins
from transgenic tobacco that could potentially reach large quantities. The mini-
spidroin monomer produced from tobacco has the native Nephila clavipes
dragline silk NTD and CTD which flank 8, 16, or 32 copies of a MaSp1 or MaSp2
consensus repeat domains. Additionally, these mini-spidroins can spontaneously
form multimers with various molecular sizes (Peng et al., in review).

Considering the charges that amino acid side chains carry on a protein
molecule, spidroins in solution are polyelectrolytes. Oppositely charged
polyelectrolytes form complexes at their interface. Although this approach has
been used to form fibers from poly-L-lysine and gellan gum (Yamamoto et al.,
2001), poly α, L-glutamic acid and chitosan (Ohkawa et al., 2001), and lysozyme
amyloid nano-fibrils and gellan gum (Meier and Welland, 2011), it remains to be
determined if it will facilitate fiber formation of any spider silk-like proteins. The current spider-silk like fiber assembly involves dissolving the protein powder in a strong solvent, hexafluoroisopropanol (HFIP), followed by injection of the solution into a dehydration/coagulation bath, usually 80%-90% of methanol or ethanol (Hsia et al., 2012; Teulé et al., 2009).

Here, we present our studies to produce mini-spidroin fibers through polyelectrolyte complex formation. Mini-spidroins, after treatment with acetic acid, cross-linked with glutaraldehyde and diluted in phosphate buffer, formed a thin film at the interface of gellan gum solution. The film can be immediately pulled into fibers with forceps. Fibers were imaged under various microscopes. The tensile properties were measured through single displacement controlled tensile testing.

MATERIALS AND METHODS

Gellan gum and mini-spidroin fiber assembly

The gelatin-like mini-spidroin samples resulting from freeze-drying (Peng et al., in review) were mixed with 25% acetic acid and 25% glutaraldehyde to make a spinning solution. For every 25 mg (~25 μl) of gelatin-like liquid (estimated concentration 123.8 ± 15.9 μg/μl and 192.8 ± 20.9 μg/μl for rMaSp1R8 and rMaSp2R8, respectively), 0.2 μl of 25% acetic acid and 5 μl of glutaraldehyde 25% solution was added. The mixture was incubated at ambient
temperature for ≥5 hours and then diluted to 150 μl with 10 mM phosphate buffer (pH 7). Aliquots of mini-spidroin spinning solution (10 μl) were carefully overlaid with 150 μl gellan gum solution (in water) that had been kept at 55°C. For rMaSp1 and rMaSp2 alone, the gellan gum concentration was 0.5%. For mixtures of rMaSp1 and rMaSp2, the gellan gum concentration was either 0.5% or 0.1%. Fibers were drawn from the interface of the gellan gum and mini-spidroin mixture, air-dried and kept at room temperature. Bovine serum albumin (BSA, 10 mg/ml in water) was used as a negative control.

Chitosan based mini-spidroin and Gellan gum fiber assembly

Chitosan (medium molecular weight, Sigma Aldrich #448877) was dissolved in 1% acetic acid to a final concentration of 5 mg/ml. To crosslink chitosan, 1 ml of glutaraldehyde (25% water solution) was added into 20 ml of the chitosan solution and incubated at room temperature overnight. Mini-spidroin rMaSp1R8 or rMaSp2R8 proteins samples were crosslinked with glutaraldehyde and diluted in phosphate buffer as described above. Fibers with a single mini-spidroin component were pulled from the mixture of 20 μl crosslinked chitosan solution and 10 μl of either rMaSp1R8 or rMaSp2R8 solution, layered with 0.5% gellan gum solution. Fibers with a combination of rMaSp1R8 and rMaSp2R8 (70:30) were pulled from the mixture of chitosan solution (20 μl), rMaSp1R8 (10 μl) and rMaSp2R8 (4.2 μl), layered with 0.5% gellan gum solution. Fibers were
drawn from the interface of the gellan gum and mini-spidroin mixture. The fibers were air dried and kept at room temperature.

**Post-spin processing of dried and post-spin processing of wet chitosan based mini-spidroin and gellan gum fiber**

Air dried fibers were cut into ~10 cm pieces and immersed in water for 2 min. They were stretched about 15-20 mm in the water and then air dried.

The nascent wet fibers pulled from the polyelectrolyte interface were immediately submerged in a water bath. The fibers were stretched using the friction created by pulling the water submerged fiber out against the pre-wetted smooth surface of the curved side (~ 10 cm) of the glass water bath.

**Tensile testing**

Fibers were cut into ~3 cm pieces, mounted onto 3 cm$^2$ paper squares with a precut 1 cm$^2$ opening in the center and equilibrated at room temperature (20ºC and 55-65% RH) overnight. The diameter of the fibers was measured under a basic light microscope with a 20x objective. The paper square was loaded onto the tensile testing unit of a Bruker CETR UMT200 loading cell. The standard displacement-controlled tensile test protocol was used (Pre-load = 0, constant displacement rate = 0.1 mm/s and max displacement of 20 mm). The carriage pulling force (Fz) and carriage position (Z) was continuously recorded. Relaxation and thermal drift holding sequence were applied after the max
displacement was reached. Young’s Modulus and toughness were calculated by UMT-Viewer Software.

**Imaging**

For light microscope imaging, the fibers were imaged with a Nikon inverted microscope. For fluorescent confocal imaging, the fibers were immersed in 1x phosphate buffered saline (pH 7.4) (137 mM sodium chloride, 2.7 mM potassium chloride, and 10 mM sodium phosphate) mounted with 50% glycerol, and imaged using a Nikon TiE Eclipse microscope with a Csi1 spectral confocal head. Polarized light images were taken with a Leica DM 750 P polarizing microscope using a 10X lens, and a 533 nm lambda plate at the Clemson Light Imaging Facility.

**RESULTS**

When treated with acetic acid, cross-linked with glutaraldehyde and diluted in phosphate buffer, the mini-spidroin polyelectrolytes, that carry positive charges (rMaSp1R8 with predicted pI of 9.16), caused counter-ion condensation at the interface of negatively charged gellan gum. We pulled fibrous material from the thin film formed at the interface. The charge density of rMaSp2R8 mini-spidroin may be weaker since the predicted pI for rMaSp2R8 is 5.7. This can be inferred from the brittleness of fibers that were pulled from rMaSp2R8 samples under the same conditions. Fiber formation conditions with single protein
components of rMaSp1R8 or rMaSp2R8 is summarized in Table 3.1. The crosslinking of mini-spidroins by glutaraldehyde is essential for the formation of mini-spidroin/gellan gum fibers. The air-dried fibers can be rehydrated with water and retain a fiber structure but water absorption leads to swelling, which has been documented in natural dragline fibers (Holland et al., 2012). Neither water nor BSA control reactions exhibited counterion condensation; thus, no fibers were assembled (Table 3.1).

We also pulled fibers from mixtures of rMaSp1 and rMaSp2 proteins. Since rMaSp1R8 only formed long fibers when crosslinked in the presence of acetic acid, all the rMaSp1R8 samples used for forming composite fibers were treated with acetic acid. For composite fiber formation, rMaSp1R8 and rMaSp2R8 were independently crosslinked with glutaraldehyde and diluted in phosphate buffer (pH 7), as described above for single protein fiber formation. The two proteins were then mixed as indicated in Table 3.2.

The charge density ratio of gellan gum to protein affects the physical property of the fiber complexes (Meier and Welland, 2011). So, protein spinning solution was layered with two different gellan gum concentrations (0.5% or 0.1%). We observed that using gellan gum concentration higher than 0.5% with mini-spidroins resulted in large globules at the interface that could only be pulled into short thick fibers. Regardless whether rMaSp2R8 was crosslinked in the presence of acetic acid or not, long composite fibers were formed from the mixture of the two proteins at the interface with 0.5% gellan gum but not 0.1%
gellan gum. Only short fibers were formed at the lower gellan gum concentration (0.1%) so all future experiments were done with 0.5% gellan gum.

In the wet spinning process of mini-spidroin fibers pulled from a gellan gum interface, we observed that as soon as the fibers were pulled out into the air, droplets started to develop along the fibers which resulted in fibers with uneven diameters (Fig. 3.1A). A similar phenomena was reported in the fiber pulling process from the interface of chitosan and poly α,L-glutamic acid solution (Ohkawa et al., 2001). The droplets may represent excess water that is carried with or captured in the fiber complex. The diameter of the dried mini-spidroin/gellan gum fibers ranged from 20-80 µm and up to 50 to 100 µm at the droplet sections. The observed fiber surfaces are rough, and sometimes scale-like in appearance (Fig. 3.1B).

To test the fiber tensile strength, we performed a displacement controlled standard tensile test protocol developed for the microtensile tester Bruker CETR UMT200. Since the cross-section surface area is critical in fiber tensile tests, we recorded each fiber diameter using a Fisher Scientific Micromaster light microscope. Average and standard deviation was determined from repetitive measurements. rMaSp2R8 protein alone produced very brittle fibers (with maximum strength of 12.23 ± 2.35 MPa) and most of the samples were broken during sample handling. rMaSp1R8 fibers are stiffer (with Young’s Modulus 0.883 ± 0.316 GPa) but have higher strength (maximum strength 37.86 ± 11.53 MPa). The physical properties of the composite fibers reflects a mixture of the two
proteins, less stiff (with Young’s Modulus 0.347 ± 0.221) but weaker (maximum strength 15.13 ± 4.95 MPa) (Table 3.3).

We further investigated the potential of chitosan as a matrix for mini-spidroins to form fibers since a soluble chitinous polysaccharide was detected in *N. clavipes* gland duct matrix through Fourier transform infrared (FTIR) spectroscopy (Davies et al., 2013). To assess the potential physiological relevance of this observation, we added crosslinked chitosan into the mini-spidroin spinning solution, and then layered with gellan gum. Interestingly, we were able to pull meters long fibers. Fig. 3.1A, b) showed three pieces of meters long fibers wrapped inside a 150 mm x 15 mm Petri dish. It appeared as though the fibers could be continuously made had we kept an adequate supply of spinning solution. Compared to the tensile properties of rMaSp2R8 fiber (Table 3.3), the toughness of chitosan containing rMaSp2R8 fibers improved with chitosan present in the spinning solution despite a small overlap of sample variation (from 0.584 ± 0.227 to 1.926 ± 1.15 MJ/M³ for rMaSp2R8 fibers and from 0.551 ± 0.356 to 1.242 ± 0.45 MJ/M³ for composite fibers) (Table 3.4). The maximum strength of rMaSp2R8 fibers improved from 12.23 ± 2.35 to 30.32 ± 10.91 MPa and the maximum strength of composite fibers improved from 15.13 ± 4.95 to 45.60 ± 18.30 MPa.

Post-spin processing of man-made spider silk-like fibers in water was shown to increase fiber tensile strength (Teulé et al., 2012). We cut the dried fiber into ~ 10 cm pieces and submerged them in water for 2 min. The rehydrated
fiber was pulled in water and then air dried. The tensile strength of these post-spin processed fibers was measured. We compared the tensile strength of chitosan containing mini-spidroin fiber before and after these fibers were post-spin stretched in water. The overall toughness was increased, from 1.75 ± 0.668 to 3.186 ± 1.208 MJ/M³ for rMaSp1R8, from 1.926 ± 1.15 to 2.933 ± 1.255 MJ/M³ for rMaSp2R8 and from 1.242 ± 0.45 MJ/M³ for composite fiber (Tables 3.4 and 3.5).

To produce uniform fibers in large quantity without interruption, the fiber was submerged in water immediately after being pulled from the interfacial layer. The submerged fiber was then pulled from water and stretched along a smooth glass surface. The friction generated from the wet fiber against the glass surface may serve as a mimic to the shear force along the narrowing spinning duct. Using this approach, the droplets that occur on the fiber were much smaller in size and the fibers were much more uniform and also smaller in diameter (from 44.8 ± 10 to 28.6 ± 7 μm for rMaSp1R8 fiber; from 34.9 ± 13 to 32.9 ± 6 μm for rMaSp2R8 fiber and from 40.6 ± 12 to 32.3 ± 4 μm for composite fiber). The fibers were air dried and then tested for tensile strength. The overall toughness of the fiber was also improved, from 1.75 ± 0.668 to 3.442 ± 1.779 MJ/M³ for rMaSp1R8, from 1.926 ± 1.15 to 3.010 ± 1.065 MJ/M³ for rMaSp2R8 fiber and 1.242 ± 0.45 to 3.453 ± 2.239 MJ/M³ for composite fiber (Table 3.4, 3.6). Although post-spin processing increased the maximum strength and the overall toughness of the rMaSp2R8 fiber, the rMaSp2R8 fibers tend to be stiffer as
reflected by the Young’s Modulus. The Young’s Modulus of unprocessed as-spun chitosan containing fiber is 0.489 ± 0.2 GPa, the post-spin stretch in water increases the Young’s Modulus to 1.402 ± 0.54 GPa and the post-spin water submerge and friction shear increase the Young’s Modulus to 1.487 ± 0.355 GPa (Tables 3.4, 3.5 and 3.6).

Fig.3.2 is a plot of the stress-strain curves representing the low, medium, and high strength of each group of the fiber. The strength of the rMaSp2R8 fibers shift to higher strength when it is post-spin stretched in water (yellow line). The strength of rMaSp1/2 composite fibers shift to higher strength when it is spun through the water bath and friction sheared (magenta line). The trend of the strength for rMaSp1R8 fibers is not very clear but they tend to be more extensible when post-spin stretched and when water submerged and friction force sheared (blue line).

For the microscopy studies, the fibers were immersed in 1x phosphate buffered saline (pH 7.4) mounted with 50% glycerol, and imaged using a Nikon TiE Elipse microscope with a Csi1 spectral confocal head. Fig. 3.3 shows a combined 3D volume view of the mini-spidroin fibers auto-fluorescent across numerous excitation/ emission wavelengths: 408 nm/515 nm, 488 nm/590 nm, and 561 nm/650 nm. The fibers showed positive birefringence under polarized light microscopy (Fig.3.4).
DISCUSSION

Polyelectrolyte complex fiber formation

Unlike other recombinant spider silk-like proteins that either spontaneously form fibers (Stark et al., 2007) or can be pulled into fibers in a methanol or ethanol precipitation bath (Teulé et al., 2007), the tobacco derived mini-spidroins equipped with both N- and C-terminal domains produced no fibers under the above mentioned conditions. These mini-spidroins have fairly high solubility in aqueous solution and moisture remains in the purified protein even after freeze-drying. Aqueous mini-spidroin solution thickened to a jelly-like texture after it is being mixed with acetic acid. Therefore, a charge attraction based wet spinning method was attempted (polyelectrolyte complex formation). After crosslinking mini-spidroins with glutaraldehyde and diluting with phosphate buffer, the formation of films at the interface of mini-spidroin and gellan gum solutions was observed, from which fibers could be immediately pulled with a pair of forceps (Fig. 3.1).

Factors that affect polyelectrolyte complex fiber formation

Charge attraction based polyelectrolyte complex formation is controlled by the charge distribution of each interacting molecule and the ratio of the total charges of each molecule (Meier and Welland, 2011; Ohkawa et al., 2001). Polyelectrolyte complex formation is also affected by the order that the interacting molecules are introduced to the interface (Hamman, 2010), which one
of the two molecules is layered on the other or if it is entirely submerged into a solution of the other molecule. The gellan gum concentrations we used in our experiments are based on those from experiments with poly-glycine (Yamamoto et al., 2001) and poly nano-glutamic acid (Meier and Welland, 2011). We pulled fibers with average diameters between 30 and 90 μm for individual mini-spidroin proteins and mixtures of the two mini-spidroins (Table 3.3). At gellan gum concentrations above 0.5%, a large globule formed at the interface and pulling resulted in very short and thick fibers that are not applicable for further fiber handling. Lower gellan gum concentrations (0.1%) in our experiment with the combination of rMaSp1 and rMaSp2 mini-spidroins produced very thin but very brittle fibers (≤ 2 inches, Table 3.2). Because different shapes of materials are produced at different gellan gum concentrations, it may be possible to tailor the dimensions of the resulting mini-spidroin complexes. For example, the congealed min-spidroins produced with higher gellan gum concentration may have the potential to form hydrogels or capsules. Such capsule development has been reported using poly-L-lysine (Yamamoto et al., 2001).

Charge distribution and charge density of mini-spidroin molecule is affected by the protonation state. The positive charges that are carried on an rMaSp1R8 molecule (with predicted pl of 9.59) are estimated to be stronger than the rMaSp2R8 molecule (with predicted pl of 4.34) in acidic pH (pH 5.5 or lower). This could be the reason that rMaSp1R8 only forms longer fibers when acidified before glutaraldehyde crosslinking (Table 3.1). In contrast, rMaSp2R8 forms
fibers regardless of the addition of acetic acid but rMaSp2R8 fibers are very brittle and most of them could not survive the sample handling during fiber tensile tests (Table 3.1, Table 3.3). Intermediate results are found for rMaSp1/rMaSp2 composite fibers. When rMaSp1R8 was treated with acetic acid and crosslinked with glutaraldehyde, fibers were formed regardless of rMaSp2R8 treatment with acetic acid in the mixture (Table 3.2). Considering MaSp1 protein is dominant in natural dragline silk fiber (Guehrs et al., 2008) and the pH decrease in the duct of spider major ampullate gland (Knight and Vollrath, 2001), we used mini-spidroins both treated with acetic acid in our further experiments.

A soluble form of chitinous material was reported to be present in the spider spinning duct (Davies et al., 2013). Chitin is an abundant biopolymer with β-(1,4)-N-acetyl-D-glucosamine repeat units that provide support strength to cuticles, cell walls and shells for many organisms. Partial deacetylated chitin, chitosan, is readily dissolved in acetic acid solution. Protonated chitosan is a polyionic material that carries a positive charge. The properties of chitin and chitosan has been reviewed by Rinaudo, (Rinaudo, 2006) and the functional biomaterials formed from chitosan include fibers and capsules from poly-L-glutamic acid and chitosan (Ohkawa et al., 2001), nanofibrous scaffolds from electrospinning of recombinant spider silk protein, polycaprolactone and chitosan (Zhao et al., 2013) and nanogels from chitosan derivatives, hexanoyl chitosan and succinoyl chitosan (Zubareva et al., 2013). Chitosan is a polycationic polysaccharide that was used to form polyelectrolyte complexes with poly-
glutamic acid (Ohkawa et al., 2001). Chitosan also forms fiber complexes with gellan gum. At the gellan gum concentration and crosslinking conditions used in our experiments, chitosan and gellan gum fibers are short and brittle. This led us to explore the potential of chitosan to provide a support matrix for mini-spidroin fibers. Meter long fibers were pulled from the interface of chitosan/mini-spidroin and gellan gum (Table 3.4). We assume that fibers could be endlessly pulled if the spinning solution were continuously supplied.

Fiber tensile testing data of natural spider silk or recombinant spider silk-like proteins are known to have a high variation (Brooks et al., 2005; Savage and Gosline, 2008a). We also encountered high variance in our fiber tensile testing data. The overall toughness was 0.584 ± 0.227 MJ/M³ for rMaSp2R8 gellan gum fiber and 1.926 ± 1.15 MJ/M³ for chitosan containing rMaSp2R8 gellan gum fiber. The overall toughness was 0.551 ± 0.356 MJ/M³ for the combination of rMaSp1/2 gellan gum fiber and 1.242 ± 0.45 MJ/M³ for chitosan containing combination of rMaSp1/2 composite gellan gum fiber. The overall toughness was the same for rMaSp1R8 gellan gum fibers and rMaSp1R8 chitosan containing fiber (1.444 ± 0.976 MJ/M³ and 1.75 ± 0.668 MJ/M³, respectively). The chitosan containing rMaSp2R8 fiber and the composite fiber have a tendency to be tougher albeit there is a slight overlap of the two groups of data (Table 3.3, Table 3.4).
**Crosslinking reaction**

Glutaraldehyde present in aqueous solution has multiple forms (monomer, dimer or polymer) and crosslinking reactions are pH and concentration dependent (Migneault et al., 2004). Glutaraldehyde is known to majorly react with the nucleophilic ε-amino of lysine side chain (Hopwood et al., 1970). Lysine residues are found only in the NTDs of the MaSps with six occurrences in MaSp1A NTD (ACF19411), five occurrences in MaSp1B NTD (ACF19412) and four occurrences in MaSp2 NTD (ACF 19413). It is absent from both MaSp1 and MaSp2 repeat domains and CTDs (partial sequence P19837) and MaSp2 repeat domain and CTD (partial sequence AAA29381). Assuming lysine residues are the primary target of glutaraldehyde crosslinking, the resulting microfibrils are expected to be crosslinked by their NTD.

Glutaraldehyde crosslinking is critical in forming mini-spidroin and gellan gum fibers (Table 3.1). No fibers were formed if mini-spidroins were not crosslinked by glutaraldehyde. In our experiments, each mini-spidroin crosslinking reaction is carried out independently. To pull composite fibers, rMaSp1 and rMaSp2 were mixed after crosslinking. Surprisingly, we could not pull any fibers when rMaSp1 and rMaSp2 were mixed prior to crosslinking. However, when mini-spidroins are mixed with crosslinked chitosan matrix, we were able to pull out fibers both from glytaraldehyde crosslinked mini-spidroin (Table 3.4) and non-crosslinked mini-spidroins.
**Fiber physical and mechanical properties**

Mini-spidroin fibers showed autofluorescence across a broad range on the visible spectrum (Fig 3.3) consistent with what is seen from natural spider dragline silks (Wendt et al., 2011). The surface of the as-spun fibers is rough (Fig. 3.3). The droplets that developed along the nascent fibers dry into the swollen nodes (Fig. 3.2). We observed occasional swollen nodes on native dragline silks of *Nephila clavipes* collected in Georgetown, South Carolina. We speculate that the excess water captured from the spinning solution may cause this uneven diameter on the as-spun fibers. Fibers pulled through water and friction force sheared against a smooth glass have much smaller diameter, swollen nodes and the resulting fibers have a much more even diameter.

The mechanical properties of chitosan containing mini-spidroin gellan gum fibers are represented in the stress-strain plot with selected samples representing average, minimum and maximum strength (Fig. 3.2). The overall toughness of chitosan containing mini-spidroin gellan gum fiber (3.442 ± 1.779 MJ/M^3^ for rMaSp1R8, 3.010 ± 1.065 MJ/M^3^ for rMaSp2R8 and 3.453 ± 2.239 MJ/M^3^ for rMaSp1/2R8) is not comparable to the overall toughness of native *Nephila clavipes* (518.0 ± 202.4 MJ/M^3^, sample number N= 13, Marcotte, unpublished data). This is not surprising since native *N. clavipes* dragline silk contains roughly 100 copies of the repeat domain and our mini-spidroins only contain eight copies of the repeat domain. The overall toughness of our fiber with
eight copies of repeat domain is fairly comparable to bone, high tensile steel and elastin (Gosline et al., 1999).

**Fiber post spinning process and hydrogen bond**

The distinct mechanical properties of biomaterials are associated with their molecular structure (Meyers et al., 2013). In spider silk’s nano structure, a cluster of 3-4 weak hydrogen bonds connect the layers of β-sheet (forms a crystal of a few nanometer in size) and contribute to fiber breakage resistance (Keten et al., 2010). Hydrogen bonds also function as an internal network spacer of the amorphous region (Simmons et al., 1996). Water disrupts this network and causes the silk fiber to contract in the direction of the fiber axis (Ene et al., 2009). Therefore, effectively removing water from the wet-spinning process may restore the hydrogen bond between the layers of the β-sheet and amorphous regions. The droplets developed along our nascent fiber caused uneven fiber diameter and may also disturb hydrogen bond formation in the fibers. The change of fiber nanostructure and its hydrogen bond network may contribute to the increase of the tensile strength. The nanostructure of spider silk is based on earlier $^2$H NMR data. $^2$H labeled alanine was simultaneously detected in an oriented dense unit (ordered β-sheet crystal) and a less oriented loose unit (amorphous helical region) in spider silk fiber (Simmons et al., 1996). Computer modeling of the amino acid sequence suggests that the hydroxyl group of serine and tyrosine may connect to glycine in the amorphous region or between β-sheet layers through identical
hydrogen bonding (Hayashi et al., 1999; Holland et al., 2013). Although hydrogen bonding is the weakest covalent bond, when 3-4 hydrogen bonds form a cluster in a small β-sheet crystal of a few nanometer in size, they become resistant to breakage (Keten and Buehler, 2008). Infrared absorption showed that water disrupts the hydrogen bond that hold the loose amorphous region in shape, causing supercontraction of this region, which appears to shorten the length of the fiber (Ene et al., 2009). Excess captured water in the wet spinning process may disrupt hydrogen bonding in the fiber. Wet spider silk fiber stretching in water may help to re-establish the hydrogen bond network, as indicated in Termonia’s computer model (Termonia, 1994). The stress strain curve of a stretched wet fiber resembles that of a dried fiber. Comparing the tensile test data of chitosan containing mini-spidroin gellan gum fibers in the form of as-spun (Table 3.4), dried post-spin fiber rehydrated in water and stretched in water (Table 3.5) and post-spin water submerged and friction force sheared (Table 3.6), we saw a trend of increase in maximum strength and overall toughness in the post-spin processed samples either stretched in water or water submerged and friction force sheared. Post spinning processing also improved fiber diameter uniformity (Tables 3.5 and 3.6).

**Proline and hydrogen bond**

The difference in tensile strength improvement of rMaSp1R8 fiber and rMaSp2R8 fiber in response to post-spin processing may be due to the amino
acid difference in the repeat block of the two proteins. Proline, a ring forming amino acid found exclusively in MaSp2 repetitive domain (deprived in MaSp1 repetitive domain) may create inter-molecular spaces required for hydrogen bonding and may affect the nano-structure of the native dragline silk (Brown et al., 2011). The proline content in MaSp2 protein has been reported to relate to the supercontraction and elasticity of spider silks (Savage and Gosline, 2008b, Shi et al., 2014). Higher proline content correlates to higher fiber elasticity and stronger supercontraction (Liu et al., 2008). Savage and Gosline reported that the tensile strength of *Araneus* dragline silk with higher MaSp2 content drastically changed in wet and dry conditions whereas the tensile strength of *Nephila* dragline silk with lower MaSp2 content did not changed much in wet and dry conditions (Savage and Gosline, 2008a). The difference in tensile strength changes of rMaSp1R8 fiber and rMaSp2R8 fiber in response to water processing provide corroborating evidence that proline in rMaSp2R8 may play an important role in fiber nanostructure and hydrogen bonding.

**Future study**

Currently we are assessing the possibility of an alternative crosslinker that may allow us to have precisely control on the crosslinking reaction. Considering the various factors that may affect polyelectrolyte complex formation and the tensile strength of the fibers, we will further evaluate and optimize fiber pulling conditions with the newly employed alternative crosslinker. Since natural spider
dragline silk proteins contain about 100 repeat domains and recombinant spidroins with higher number of repeat domains produced fibers with higher tensile strength (Xia et al., 2010), we will further investigate the relationship of the repeat domain number to the tensile strength of polyelectrolyte complex fibers using tobacco derived rMaSp1/2 R16 and rMaSp2 R32 proteins (Peng et al, in review). We hope to produce high performance fibers that can be utilized in future biomedical studies. We will continue our efforts to optimize the fiber assembly process for stronger and better fibers and to further assess the possibility to make these fibers versatile in biomedical engineering applications.

CONCLUSIONS

Through counter-ion polyelectrolyte condensation, we produced tobacco derived mini-spidroin fibers. The tensile strength of mini-spidroin fibers can be improved by spinning mini-spidroins in a chitosan matrix. After a post-spin water submerging and friction force shearing, we were able to continuously produce fibers with an average diameter in the range of 28 to 33 μm and and overall toughness of 3.442 ± 1.779; 3.010 ± 1.065 and 3.453 ± 2.239 MJ/M^3 for rMaSp1R8; rMaSp2R8 and rMaSp1/2 R8, respectively. We speculate that the spider silk formation process in spider spinning duct may involve charge attraction of variously charged chitinous polymer, spidroins and glycoproteins. This is the first report that man-made spider silk-like fibers can be made through polyelectrolyte complex formation, a cost effective process that is safe for
subsequent medical applications. The autofluorescent and positive birefringent nature of these fibers is promising for broad medical and engineering applications. We will continue to explore methods that lead to stronger and tougher fibers that may outperform other man-made materials.

This study is a great addition to the current knowledge on spider silk-like fiber assembly. With the understanding that chitinous polymers may be a matrix in the spider silk spinning duct, that native spider silk have layers of MaSp1 and MaSp2 core region and glycoprotein skin (Augsten et al., 2000; Sponner et al., 2007) and that fibers from polyelectrolyte complexes have an inner core of one polyon and an outer layer of the counter polyon ((Meier and Welland, 2011), we speculate that the silk formation process in spider spinning duct may involve the charge attraction of variously charged chitinous polymer, spidroins and glycoproteins in order to assemble into polyelectrolyte complexes.

Acknowledgement

The authors would like to thank Dr. Michael Ellison from the Department of Material Science and Dr. John Desjardins from the Department of Bioengineering for their help on fiber tensile testing and fiber formation. We would like to thank Dr. Terri Bruce for her help on microscopy and Dr. Florence Tuelé for her constructive suggestions.
Table 3.1. Fiber assembly of single protein component (rMaSp1R8 or rMaSp2R8) in the presence or absence of acetic acid and/or crosslinker glutaraldehyde. Water and bovine serum albumin were used as negative controls. All spinning solution is layered with 0.5% gellan gum solution. Composition of spinning solution is described in the results.

<table>
<thead>
<tr>
<th></th>
<th>acetic acid (25%)</th>
<th>glutaraldehyde (25% water solution)</th>
<th>Phosphate Buffer (pH 7)</th>
<th>Fiber Length (every 10 µl spinning solution)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (dH₂O)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>No Fiber</td>
</tr>
<tr>
<td>Bovine Serum Albumin (10 mg/ml)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>No Fiber</td>
</tr>
<tr>
<td>rMaSp1R8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>≥ 4 inches</td>
</tr>
<tr>
<td>rMaSp1R8</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>≤ 3 inches</td>
</tr>
<tr>
<td>rMaSp1R8</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>No Fiber</td>
</tr>
<tr>
<td>rMaSp2R8</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>≤ 3 inches</td>
</tr>
<tr>
<td>rMaSp2R8</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>≥ 4 inches</td>
</tr>
<tr>
<td>rMaSp2R8</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>No Fiber</td>
</tr>
</tbody>
</table>
Table 3.2. **Fiber assembly of two protein component** (mixture of rMaSp1R8 and rMaSp2R8 after each independently reacted with crosslinker glutaraldehyde in the presence or absence of acetic acid) at different gellan gum concentration 0.5% or 0.1%. Composition of spinning solution is described in the methods.

<table>
<thead>
<tr>
<th>rMaSp1R8 treated with acetic acid and glutaraldehyde (%)</th>
<th>rMaSp2R8 treated with acetic acid and glutaraldehyde (%)</th>
<th>rMaSp2R8 treated with glutaraldehyde (%)</th>
<th>Phosphate Buffer (pH 7)</th>
<th>Gellan gum (%)</th>
<th>Fiber Length (Based on every 10 µl spinning solution)</th>
</tr>
</thead>
<tbody>
<tr>
<td>70</td>
<td>30</td>
<td>-</td>
<td>+</td>
<td>0.5</td>
<td>≥ 4 inches</td>
</tr>
<tr>
<td>70</td>
<td>-</td>
<td>30</td>
<td>+</td>
<td>0.5</td>
<td>≥ 4 inches</td>
</tr>
<tr>
<td>70</td>
<td>30</td>
<td>-</td>
<td>+</td>
<td>0.1</td>
<td>≤ 2 inches</td>
</tr>
<tr>
<td>70</td>
<td>-</td>
<td>30</td>
<td>+</td>
<td>0.1</td>
<td>≤ 2 inches</td>
</tr>
</tbody>
</table>
Table 3.3. Mechanical property of rMaSp R8 and gellan gum fiber.  
(Recorded by single displacement controlled tensile test).

<table>
<thead>
<tr>
<th>Fiber Composition</th>
<th>Sample number (N)</th>
<th>Diameter (mm)</th>
<th>Young's Modulus (GPa)</th>
<th>$\sigma_{\text{max}}$ (MPa)</th>
<th>$\varepsilon_{\text{max}}$</th>
<th>Toughness (MJ/m$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rMaSp1(R8)</td>
<td>10</td>
<td>0.0368 ± 0.009</td>
<td>0.883 ± 0.316</td>
<td>37.86 ± 11.53</td>
<td>7.09 ± 3.47</td>
<td>1.444 ± 0.976</td>
</tr>
<tr>
<td>rMaSp2(R8)</td>
<td>3*</td>
<td>0.054 ± 0.003</td>
<td>0.22 ± 0.14</td>
<td>12.23 ± 2.35</td>
<td>9.88 ± 5.02</td>
<td>0.584 ± 0.227</td>
</tr>
<tr>
<td>rMaSp1/2 (R8)**</td>
<td>13</td>
<td>0.072 ± 0.019</td>
<td>0.347 ± 0.221</td>
<td>15.13 ± 4.95</td>
<td>6.84 ± 3.16</td>
<td>0.551 ± 0.356</td>
</tr>
</tbody>
</table>

*rMaSp2R8 protein crosslinked with 5% glutaraldehyde resulted in brittle fibers. 15 samples were tested, only 3 sample survived sample handling.  
**rMaSp1R8 and rMaSp2R8 are crosslinked individually and then mixed in a volume ratio 70:30.
Table 3.4. Mechanical property of chitosan containing rMaSpR8 and gellan gum fiber. (Recorded by single displacement controlled tensile test).

<table>
<thead>
<tr>
<th>Fiber Compositio n</th>
<th>Sample number (N)</th>
<th>Diameter (mm)</th>
<th>Young’s Modulus (GPa)</th>
<th>$\sigma_{max}$ (MPa)</th>
<th>$\epsilon_{max}$</th>
<th>Toughness (MJ/M^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rMaSp1(R8)</td>
<td>9</td>
<td>0.0448 ± 0.010</td>
<td>0.949 ± 0.407</td>
<td>38.50 ± 8.173</td>
<td>9.06 ± 2.82</td>
<td>1.750 ± 0.668</td>
</tr>
<tr>
<td>rMaSp2(R8)</td>
<td>9</td>
<td>0.0349 ± 0.013</td>
<td>0.489 ± 0.2</td>
<td>30.32 ± 10.91</td>
<td>10.86 ± 3.87</td>
<td>1.926 ± 1.150</td>
</tr>
<tr>
<td>rMaSp1/2(R8)*</td>
<td>8</td>
<td>0.0406 ± 0.012</td>
<td>1.446 ± 0.765</td>
<td>45.60 ± 18.30</td>
<td>5.73 ± 1.36</td>
<td>1.242 ± 0.450</td>
</tr>
</tbody>
</table>

*rMaSp1R8 and rMaSp2R8 are crosslinked individually and then mixed in a volume ratio 70:30.
Table 3.5. Mechanical property of post-spin stretched chitosan containing rMaSpR8 and gellan gum fiber. (Recorded by single displacement controlled tensile test).

<table>
<thead>
<tr>
<th>Fiber Composition</th>
<th>Sample number (N)</th>
<th>Diameter (mm)</th>
<th>Young's Modulus (GPa)</th>
<th>$\sigma_{\text{max}}$ (MPa)</th>
<th>$\varepsilon_{\text{max}}$</th>
<th>Toughness (MJ/m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rMaSp1(R8)</td>
<td>9</td>
<td>0.0384 ± 0.010</td>
<td>0.827 ± 0.679</td>
<td>65.28 ± 20.45</td>
<td>9.73 ± 2.79</td>
<td>3.186 ± 1.208</td>
</tr>
<tr>
<td>rMaSp2(R8)</td>
<td>9</td>
<td>0.041 ± 0.01</td>
<td>1.402 ± 0.540</td>
<td>66.38 ± 22.58</td>
<td>8.87 ± 1.92</td>
<td>2.933 ± 1.255</td>
</tr>
<tr>
<td>rMaSp1/2(R8)</td>
<td>9</td>
<td>0.0354 ± 0.009</td>
<td>1.518 ± 0.829</td>
<td>61.84 ± 14.13</td>
<td>7.43 ± 2.69</td>
<td>2.306 ± 1.003</td>
</tr>
</tbody>
</table>
Table 3.6. Mechanical property of water submerged and friction force sheared chitosan containing rMaSpR8 gellan gum fiber. (Recorded by single displacement controlled tensile test).

<table>
<thead>
<tr>
<th>Fiber Composition</th>
<th>Sample number (N)</th>
<th>Diameter (mm)</th>
<th>Young's Modulus (GPa)</th>
<th>$\sigma_{\text{max}}$ (MPa)</th>
<th>$\varepsilon_{\text{max}}$</th>
<th>Toughness (MJ/M$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rMaSp1(R8)</td>
<td>9</td>
<td>0.0286 ± 0.007</td>
<td>1.222 ± 0.191</td>
<td>66.77 ± 16.39</td>
<td>9.86 ± 2.69</td>
<td>3.442 ± 1.779</td>
</tr>
<tr>
<td>rMaSp2(R8)</td>
<td>9</td>
<td>0.0329 ± 0.006</td>
<td>1.487 ± 0.355</td>
<td>66.87 ± 15.95</td>
<td>8.90 ± 1.68</td>
<td>3.010 ± 1.065</td>
</tr>
<tr>
<td>rMaSp1/2(R8)</td>
<td>9</td>
<td>0.0323 ± 0.004</td>
<td>1.548 ± 0.716</td>
<td>75.88 ± 28.82</td>
<td>8.39 ± 3.29</td>
<td>3.453 ± 2.239</td>
</tr>
</tbody>
</table>
Figure 3.1. Images of mini-spidroin and gellan gum fiber. **A.** Image of nascent composite fibers from the mixture of rMaSp1R8 (70%) and rMaSp2R8 (30%). a, Droplets formed along the wet (left) nascent fiber appears as swollen nodes along the dry (right) nascent fiber. b, Dried nascent fiber collected in a 150 mm x 15 mm Petri dish. **B.** Light microscope image of mini-spidroin and gellan gum fiber taken at 20x with Nikon inverted light microscope. Left, rMaSp1R8 fiber; Middle, rMaSp2R8 fiber; Right, composite fiber from the mixture of rMaSp1R8 (70%) and rMaSp2R8 (30%).
Figure 3.2. Representative stress-strain curve of chitosan containing mini-spidroin and gellan gum fiber. **Left**, as-spun, the fibers were air-dried after they were pulled; **Middle**, post-spin water stretched, the fibers were air-dried after they were pulled and then rehydrated with water and stretched in the water and then air-dried; **Right**, water submerged and friction force sheared, the fibers were pulled and then dropped into a water bath and then pulled against a smooth glass surface as the friction generated between wet fiber and glass surface serves as a shear force. Blue, rMaSp1R8 fiber; Yellow, rMaSp2R8 fiber; Maganta, composite fiber from the mixture of rMaSp1R8 (70%) and rMaSp2R8 (30%).
**Figure 3.3. Combined fluorescence spectrum 3D volume view of fibers**
under Nikon Eclipse Ti, inverted microscope with Nikon EZ-C1 spectral confocal
at excitation/ emission wavelengths 408 nm/515 nm, 488 nm/590 nm, and 561
nm/650 nm. Left, rMaSp1(R8) fiber; Middle, rMaSp2(R8) fiber; Right, rMaSp1/2
(R8) fiber.
Figure 3.4. Polarized light image of rMaSpR8 gellan gum fibers. Left, rMaSp1(R8) fiber; Middle, rMaSp2(R8) fiber; Right, rMaSp1/2 (R8) fiber. (Courtesy from Dr. Terri Bruce)
REFERENCE:


Zubareva, A., Ily‘ina, A., Prokhorov, A., Kurek, D., Efremov, M., Varlamov, V.,
of protein and peptide binding to nanogels formed by differently charged
chitosan derivatives. Molecules 18, 7848-7864.
CHAPTER FOUR

ENZYME ACTIVITY AND ITS ROLE IN PROTEIN CROSS-LINKING OF A SPIDER AMPULLATE GLAND SPECIFIC ENZYME
INTRODUCTION

The unique mechanical properties of spider dragline silk inspire researchers to produce similar materials from biotechnologically modified organisms. Research in this area covers topics from the molecular basis of spider silk gene and protein structure to fiber assembly processes and biomimicry of the spider silk spinning duct (Lewis, 2006; Vendrely and Scheibel, 2007; Vepari and Kaplan, 2007). The high tensile strength major ampullate dragline silk consists largely of two distinct spidroin proteins named major ampullate spidroins 1 and 2 (MaSp1 and MaSp2) (Hinman and Lewis, 1992; Xu and Lewis, 1990). MaSps have a conserved protein architecture consisting of short (~150 AA) non-repetitive N- and C-terminal domains flanking a large (3000-4000 AA), highly repetitive central domain. The repeat domains of MaSp1 and MaSp2 are similar in that each repeat contains a glycine-rich region that is largely amorphous (β-spiral/α-helix) and a poly-alanine region that forms β-crystallites (Hayashi et al., 1999). This two component nature, crystalline embedded in the amorphous matrix, has been confirmed by solid-state $^2$H NMR (Simmons et al., 1996), X-ray micro-diffraction (Sampath et al., 2012) and scanning transmission X-ray microscopy (Rousseau et al., 2007). Intra- and intermolecularly hydrogen bonding contributes significantly to the stability of the structure and the overall toughness of the fiber (Ene et al., 2009).

The non-repetitive N- and C-terminal domains (NTD and CTD) of MaSp1 and MaSp2 are conserved among most silk proteins made by insects and
spiders (Bini et al., 2004), suggesting an essential role of the non-repetitive domains in fiber formation. In the storage sac of the gland where the pH is near neutral, NTDs form dynamic homodimers through salt bridges and CTDs form stable homodimers through salt bridges and a disulfide bond (Hagn et al., 2010; Ittah et al., 2006; Wallace and Shen, 2012). In response to the pH decrease along the major ampullate duct, NTDs undergo a conformation change that leads to the formation of stable non-covalent dimers. There is currently no evidence for covalently linked NTD dimers (Dicko et al., 2004a; Gaines et al., 2010; Hagn et al., 2011). Whether this stable dimer is an intermediate stage or a precursor for stronger connection is unknown. CTD dimers, however, become unstable in acidic pH, exposing the hydrophobic core region that may act as a docking platform for the hydrophobic polyalanine region and contribute to nucleation (Gauthier et al., 2014).

Nucleation and formation of amyloid like microfibrils can be visualized by electronic microscopy of dissected *Nephila clavipes* major ampullate gland (Kenney et al., 2002). Nucleation is localized in Zone A region of the gland and microfibrils formation starts at the junction region of Zones A and B. Interestingly, peroxidase activity has also been detected in epithelial cells in Zone B. Peroxidase activity is stronger at the junction of Zone B and the spinning duct and traces of peroxidase activity can be detected in newly pulled fiber (Vollrath and Knight, 1999). Consistent with this observation, EST analysis of *N.*
The deduced sequence of *N. senegalensis* peroxidase (*NsPox*) is 634 amino acids in length, with a predicted molecular mass of 71 kDa and pI of 5.9. A conserved domain search reveals that it contains a heme binding site, putative substrate binding site and calcium binding site, indicating that it belongs to the peroxinectin_like family within the animal heme peroxidase superfamily (Marchler-Bauer et al., 2015). One possible role for peroxidase activity in the gland could be to create crosslinks through dityrosine formation. While early studies did not detect dityrosine in dragline silk hydrolysates (Vollrath and Knight, 1999), more recent work reveals dityrosine and 3,4-dihydroxyphenylalanine in MaSp1 (dos Santos-Pinto et al., 2014). However, there has been no causal link established between the presence of these structures and peroxidase activity.

We have used a synthetic gene to produce recombinant *NsPox* protein and investigate the enzyme activity and function. We demonstrate recombinant *NsPox* purification and immuno-detection, *NsPox* peroxidase activity at neutral pH (7.4) and acidic pH (6) and the effect of calcium ion on *NsPox* activity. Spider major ampullate gland peroxidase may function in crossing linking spidroins either through the formation of di-tyrosine or through the formation of sulfilimine bonds by crosslinking methionine and lysine residues thereby securing the dimer association through a stable covalent bond that facilitates microfibril formation. Identification of a role for *NsPox* in fiber assembly in the context of crosslinking
mini-spiderin and forming microfibrils will be a great addition to our current knowledge of native spider silk fiber assembly. The elucidation of fibril assembly by peroxidase-assisted crosslinking could lead to the production of artificial fibers from spiderin-like proteins with properties closer to native spider silks.

**MATERIALS AND METHODS**

**NsPox gene construction**

The NsPox (GenBank AF516694) (Pouchkina et al., 2003) gene sequence was codon optimized for expression in *E. coli* and tailored at the 5’ and 3’ ends to facilitate future cloning. In addition, an XbaI restriction site was inserted immediately upstream of the predicted signal peptide cleavage site and a six histidine tag followed by a stop codon and NotI restriction site was added to the 3’ end of the sequence (Fig. 4.1).

The modified sequence was synthesized by GenScript (Piscataway, NJ) and provided as a clone in pUC57. NsPox 6H was subcloned into pMAL c2X vector to create pNsPox. The XbaI/HindIII subfragment was subcloned into the pMAL c2x vector, thus removing the signal peptide sequence. In this plasmid, the peroxidase coding sequence is downstream and in frame with the maltose binging protein (MBP) coding sequence. Both junctions were confirmed by sequencing (Clemson University Genomics Institute).
Protein purification

pNsPox was transformed into E. coli BL21 cells and 30 ml of an overnight culture was used to inoculate fresh LB ampicillin medium (1 liter) and allowed to grow at 37 °C with agitation to OD$_{600}$ = 0.6. Protein expression was induced by addition of IPTG to 0.3 mM and continued to growth at 19°C with agitation for 2-3 hours. Cells were harvested by centrifugation at 2130 xg for 20 min.

For amylose affinity chromatography, the cell pellet from 600 ml of a one liter culture was re-suspended in 125 ml column buffer (20 mM Tris-HCl pH 7.4, 200 mM NaCl, 1 mM DTT) supplemented with 1 mM PMSF. The cell suspension was sonicated with short interval bursts (10 seconds) at power output 20%. Cell debris was removed by centrifugation at 13,200 xg for 25 min. Supernatant was collected and loaded onto 5 ml amylose resin (New England Biolabs #E8021S) pre-washed with column buffer. After loading, the column was washed with 10 bed volumes of column buffer. MBP-NsPox fusion protein was eluted with column buffer supplemented with 10 mM maltose. Aliquot fractions (number 1 to 7, 1.5 ml each) were run on SDS-PAGE gel and the fractions containing a band that matched the molecular weight of the full length MBP-NsPox fusion protein were pooled and buffered exchanged with 20 mM Tris- HCl (pH 8), supplemented with 100 mM NaCl and 2 mM CaCl$_2$ using Amicon-ultra concentrators with 3 kDa cutoff or 100 kDa cutoff. The protein solution retained from both concentrators were collected and analyzed by SDS-PAGE.
For anion exchange chromatography, the cell pellet from 500 ml of a one liter culture was re-suspended with 125 ml extraction buffer (5 mM Tris-HCl pH 7.4, 1 mM PMSF). The cell suspension was sonicated and cell debris removed as described above. Supernatant was collected and loaded onto 8 ml DEAE-Sepharose resin CL-6B (Sigma Aldrich) pre-equilibrated with 5 mM Tris-HCl (pH 7.4). The column was washed with 8 x bed volumes of 5 mM Tris-HCl (pH 7.4). Protein was eluted stepwise, first with 20 ml low salt elution buffer (5 mM Tris-HCl pH 7.4 supplemented with 125 mM NaCl) collected as 12 fractions (#1-12) and then with 20 ml high salt elution buffer (5 mM Tris-HCl pH 7.4 supplemented with 500 mM NaCl), also collected as 12 fractions (#13-24). Enzyme activity of each fraction was analyzed using spot assay with 3,3',5,5'tetramethylbenzidine (TMB) working solution (MP Biomedicals). Two 1.5 ml low salt fractions (#6 and 7) and five 1.5 ml high salt fractions (#18-22) that demonstrated positive peroxidase activity were pooled individually, concentrated and buffer exchanged with 5 mM Tris-HCl pH 7.4 supplemented with 125 mM NaCl using a Corning Spin-X Uf 50 kDa cutoff concentrator.

**Protein quantification, SDS-PAGE and immuno-detection**

Protein concentration was determined by Bradford assay (Bradford, 1976). For SDS-PAGE, protein samples (volumes as indicated in the legends) were mixed with Laemmli sample buffer and heated at 95°C for 5 min. The samples were loaded onto 12% polyacrylamide gel. For immuno-detection, anti-6x-His
epitope tag polyclonal antibody (Thermo Scientific Pierce PA 1983B) was used as primary antibody and goat anti-rabbit alkaline phosphatase was used as secondary antibody. Alkaline phosphatase was detected by Lumi-Phos chemiluminescent substrate (Thermo Scientific Pierce #34150) and imaged by Fujifilm LAS-1000plus imager.

**NsPox peroxidase activity spot assay**

Spot assays for cell crude extract, flow through and washes were done by mixing 15 µl 1.25 mM TMB solution with 20 - 40 µl aliquots from cell crude extract, flow through or washes. Spot assays for fractions were done by mixing 15 µl 1.25 mM TMB solution with 5 -10 µl aliquots from each fraction.

**NsPox peroxidase activity test**

Peroxidase activity was determined for the pooled, buffer exchanged and concentrated high salt ion exchange elution fractions using 3,3’,5,5’tetramethylbenzidine (TMB) (MP Biomedicals ICN15234650) as substrate. Reactions (1 ml) contained 30 µl of high salt elution, 200 µl 1.25 mM TMB and 730 µl buffer. Buffers were 20 mM Tris-HCl (pH 7.4) with or without 10 mM CaCl₂ and 20 mM MES (Ph 6.0) with or without 10 mM CaCl₂. Absorbance at 655 nm was monitored for 15 minutes.
RESULTS

MBP-NsPox fusion protein was expressed in *E. coli* and purified by amylose resin affinity chromatography. Figure 4.2A shows SDS-PAGE analysis of the steps of amylose resin purification. A band that corresponds to the full length MBP-NsPox fusion protein (113.06 kDa) accumulated in the induced cell lysate (Lane 2). That band is absent in uninduced lysate (Fig. 4.2A, Lane 1) and is largely depleted from the amylose resin flow through (Fig. 4.2A, Lane 3). The MBP-NsPox full length protein remained bound to the resin during the wash (Fig. 4.2A, Lane 4) and could be eluted by maltose in fractions 2-5 (Fig. 4.2A, Lanes 6-9). These fractions were pooled and buffered exchanged using a 3 kDa cutoff concentrator or 100 kDa cutoff concentrator and the retentates were collected. Immuno-detection indicates that four bands are recognizable with polyclonal anti-His antibody in the induced cell lysate (Fig. 4.2B, Lane 1), uninduced cell lysate (Fig. 4.2B, Lane 2) and flow through (Fig. 4.2B, lane 3). The top band corresponds in size to the intact MBP-NsPox fusion protein. Crude lysate of BL21 cells alone do not contain any anti-His-detectable bands (Fig. 4.5) suggesting the bands of lesser molecular weight represent truncation products of MBP-NsPox. As might be predicted, the presence of DTT in the extraction and elution buffers abolished peroxidase activity in the cell lysate. Peroxidase activity was recovered after DTT was removed from the eluates through buffer exchange.

Figure 4.3 shows the SDS-PAGE analysis of the anion exchange purification and shows that most of the bound protein elutes with high salt.
Immuno-detection indicates that bands correspond to full-length MBP-NsPox and the same truncated products seen with amylose purification are recognizable using the anti-His antibody, in crude lysate, low salt and high salt elutions, although the majority is in the high salt elution (Fig. 4.3B). Final protein concentration was 1.82 µg/µl for high salt elution. The protein concentration of low salt elution was 0.064 µg/µl before buffer exchange and it was barely detectable after and not used further. DEAE-sepharose ion exchange resin bound all detectable peroxidase activity as crude cell lysate was depleted of peroxidase activity after the passage over the DEAE-sepharose resin.

Peroxidase activity (spot assay) was observed in the low salt elution (fractions 6, 7; Fig. 4.3A, Lane 3) and a high salt elution (fractions 18-22; Fig. 4.3A, Lane 4), which appears in very light brown color that may be from the iron in the heme cofactor.

**Enzyme activity assay**

Peroxidase activity was assayed by monitoring oxidation of 3,3’,5,5’tetramethylbenzidine (TMB), a reaction that results in the formation of a blue product. Fig. 4.4 summarizes enzyme activity in the buffer exchanged high salt elution from DEAE-sepharose. Peroxidase activity is readily detectable at near neutral pH (pH 7.4). This would be similar to the lumen of the major ampullate gland where spidroins are stored. Interestingly, at pH 6.0, a condition that would reflect the acidification known to occur in the duct of the gland, the
enzyme becomes significantly more active. It is also of note that activity at both pH 7.4 and pH 6.0 is decreased in the presence of 10 mM CaCl₂.

**DISCUSSION**

**MBP-NsPox fusion protein purification**

The elucidation of NsPox function will help us further understand the mechanisms of fiber assembly in spider major ampullate spinning duct. We have cloned a synthesized NsPox sequence into the pMAL-c2x vector and expressed the fusion protein in *E. coli* cells. Anti-His antibody detected a product consistent with full-length MBP-NsPox product, but also three truncated products of smaller molecular weight. Immuno-detection (Fig. 4.2 B Lanes 1, 2, 3 and Fig. 4.3 B Lanes 1) showed that truncated MBP-NsPox protein may be produced during sonication while extracting protein from *E. coli* cell or during protein processing in the *E. coli* cells.

MBP-NsPox fusion protein is 113.06 kDa in size and contains a cleavage site that is recognizable by Factor Xa (Nagai et al., 1985) that recognizes the amino acid sequence IEGR or IDGR. Factor Xa cleavage should produce two fragments, one that is NsPox (70.6 kDa) and the other that is MBP (42.48 kDa). Unfortunately, Factor Xa also recognizes a secondary site (GR), which is present in the NsPox protein sequence at multiple positions. Complete cleavage of NsPox by Factor Xa at primary and secondary sites would produce fragments of 1.84 kDa, 6.54 kDa, 42.43 kDa and 18.99 kDa with only the 18.99 kDa fragment
retaining the His tag. A fragment at ~19 kDa is seen in Factor Xa cleaved fusion protein suggesting cleavage is occurring at secondary sties (Fig 4.2 B, Lane 5).

Leaky expression of MBP-NsPox protein was detected from the crude cell lysate (Fig 4.2 B Lane 2). Immuno-detection is very sensitive and the signal was saturated by loading 10 μl of crude cell lysate. It is hard to tell the quantity of the leaky expression, but the SDS-PAGE gel shows the MBP-NsPox bands had greater intensity in the induced cells (Fig 4.2 A Lane 1, 2).

A conserved domain search reveals that NsPox belongs to the peroxinectin_like family of animal peroxidases, which includes thyroid peroxidase, myeloperoxidase_like and peroxidasin_like proteins (Marchler-Bauer et al., 2015). In addition to being implicated in the formation of dityrosine, peroxidases in the peroxinectin family have been shown to convert hydrogen peroxide and chloride into hypochlorous acid that can react with L-tyrosine to generate an aldehyde intermediate, p-hydroxyphenylacetaldehyde (pHA) (Hazen et al., 1996), which can react with the ε-amino group of lysine (Hazen et al., 1999). When the chloride concentration is low, myeloperoxidase forms dytyrosine and trityrosine molecules through converting L-tyrosine into tyrosol radicals (Jacob et al., 1996). A peroxidasin has been identified from the cDNA clones of Drosophila melanogaster (Ng et al., 1992) that participates in crosslinking collagen of the Drosophila membrane matrix (Nelson et al., 1994). A human version of peroxidasin catalyzes the reaction of chloride to hypochlorous acid as...
intermediate, which in turn crosslinks methionine and lysine residue of collagen through sulfilimine bond (Bhave et al., 2012).

Due to the presence of peroxidase activity in the major ampullate gland (Vollrath and Knight, 1999), we speculate NsPox may be involved in protein crosslinking during fiber formation. Detection of dityrosine in native silk (dos Santos-Pinto et al., 2014) is consistent with this possibility, although the source of oxidation has not been linked to peroxidase enzyme activity. A peroxidase enzyme like NsPox could also catalyze the oxidation of chloride and form hypohalogenous acids that may result in covalent sulfilimine bonds between methionine and lysine in spidroin peptide chains. Interestingly, the occurrence of methionine and lysine residues in major ampullate spidroins is restricted largely to the NTD. Not including the signal peptide that would be cleaved from mature Nephila clavipes MaSps, there are six lysine and seven methionine residues in the MaSp1A-NTD, five lysines and six methionines in the MaSp1B-NTD and four lysines and eleven methionines in the MaSp2-NTD (Gaines and Marcotte, 2008). There are no known lysines or methionines in Nephila clavipes repeat domains of MaSp1 or MaSp2, although the complete sequences are not available. In Nephila clavipes CTD, there are two lysines and a single methionine in MaSp1A, two lysines and two methionines in MaSp1B and no lysines or methionines in MaSp2. This is similar to what is seen in Latrodectus where full length sequences are available (Ayoub et al, 2007). NsPox may catalyze the reaction of methionine...
and lysine to form sulfilimine bonds thereby covalently crosslinking NTD dimers, contributing to the formation of long multimeric strands during fiber assembly.

**NsPox peroxidase activity**

NsPox is a spider ampullate gland specific enzyme (Pouchkina et al., 2003). The pH of the ampullate gland and its narrowed duct progressively decreases from 7.3 in the storage sac Zone A to 6.3 in the proximal duct region, 5.9 in the middle of the duct and 5.7 at the distal duct region (Andersson et al., 2014; Dicko et al., 2004b; Vollrath et al., 1998). Acidic pH increases the rate of hypohalous acids generation by heme peroxidases and protonation of a histidine residue in the proximity of the substrate binding site (Arnhold et al., 2006). Similar results from myeloperoxidase shows that the optimum pH for TMB oxidation by myeloperoxidase is 5.4 (Suzuki et al., 1983). A higher activity for NsPox at at lower pH infers that the catalytic function of this enzyme may be enhanced during the pH changes along the duct. This result is perfectly aligned with reports that the enzyme reaction to TMB gradually intensifies from the epithelium of Zone B area to the proximal duct area (Vollrath and Knight, 1999). Acidification in the ampullate gland duct is also associated with a conformation change of in the MaSp-NTD that triggers stable dimerization (Askarieh et al., 2010; Gaines, 2010; Hagn et al., 2011). Whether NTD dimers go on to become covalently linked is unknown at this time but the enrichment for lysine and methionine residues in the NTD and the presence of peroxidase activity lend
strength to this hypothesis. NTD dimers could be covalently crosslinked. This would allow us to extend our multimeric strand hypothesis (Gaines et al., 2010) to include covalently linked NTDs as well as CTDs.

NsPox is postulated to have a heme binding site that binds to heme cofactors (amino acid position 272-450) and a calcium binding site (amino acid position 210-220) according to the conserved domain search (Marchler-Bauer et al., 2015). Including calcium to the reaction buffer lowered the activity of NsPox oxidation of TMB at both neutral pH (7.4) and acidic pH (6) (Fig. 4.4). This is exactly opposite of the result that calcium enhances lactoperoxidase activity in converting 2,2′-azino-bis-3-ethylbenz-thiazoline-6-sulphonic acid (ABTS) (Fonteh et al., 2005). Calcium atom analysis from Cryo-SEM study detected a low percentage of calcium counts (<0.0175%) in dissected major ampullate gland (Knight and Vollrath, 2001). A study of the effect of cations on spidroin folding in Nephila edulis major ampullate gland dope shows that while potassium cations facilitate the transition of coiled to β-sheet folding, adding calcium cation keeps spidroin protein stabilized in the coiled unfolded state (Dicko et al., 2004a). This would be consistent with calcium ions modulating peroxidase activity in the lumen of the gland.

Peroxidase activity of NsPox requires an oxidative environment and the reducing agent DTT completely abolished NsPox activity in the extraction buffer for our amylose resin binding purification process. However, peroxidase activity was recovered in the eluted protein fractions after the removal of DTT through
buffer exchange. Transcriptome studies of the spider *Latrodectus hesperus* reveals that gland-specific transcripts are enriched for oxidoreductase (Clarke et al., 2014), which regulates the transfer of electrons in the reduction and oxidation reaction. This observation further supports a role for redox reactions during fiber self-assembly.

**Acknowledgement**

The authors would like to thank Dr. Michael Sehorn for constructive suggestions and Janie Hamilton for her technical supports.
Table 4.1: NsPox protein yield from amylose affinity purification and DEAE-sepharose anion exchange purification.

<table>
<thead>
<tr>
<th></th>
<th>Cell Culture Vol (ml)</th>
<th>Protein volume (ml)</th>
<th>Protein concentration (mg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein purified from amylose affinity resin</td>
<td>600</td>
<td>2.7</td>
<td>0.0345</td>
</tr>
<tr>
<td>Protein purified from DEAE-sepharose resin</td>
<td>500</td>
<td>7.0</td>
<td>1.8217</td>
</tr>
</tbody>
</table>
**Figure 4.1. Recombinant NsPox sequence details.** A. Diagrammatic representation of recombinant NsPox. B. NsPox sequence details of pUC57 *Kpn*I/*Hind*III fragment. Restriction sites are underlined. Stop codon is indicated in bold.
Figure 4.2. Purification of NsPox by amylose affinity chromatography. A. SDS-PAGE analysis (12% acrylamide gel) of uninduced cell lysate, 25 μl (Lane 1), induced cell lysate, 25 μl (Lane 2), cell lysate after pass through amylose column, 25 μl (Lane 3), washes, 25 μl (Lane 4), elution fraction 1, 5 μl (Lane 5), elution fractions 2-5, 5 μl each (Lanes 6-9, respectively). Band corresponding to full length MBP-NsPox fusion protein marked by asterisk. B. Immuno-detection of induced cell lysate 10 μl (Lane 1), uninduced cell lysate, 10 μl (Lane 2), cell lysate after pass through amylose column (flow through) 10 μl, (Lane 3), pooled fraction buffer exchanged using 100 kDa cutoff concentrator, 5 μl (Lane 4), pooled fraction buffer exchanged using 3 kDa cutoff concentrator and treated with Factor Xa, 2 μl (Lane 5). For immuno-detection, primary antibody was polyclonal rabbit anti-6x-His. Molecular weight marker sizes (in kDa) are shown to left.
Figure 4.3. Purification of NsPox by anionic exchange chromatography. A. SDS-PAGE analysis (12% acrylamide gel) of induced cell crude cell lysate, 20 μl (Lane 1), cell lysate after pass through DEAE-sepharose CL-6B column, 20 μl (Lane 2), pooled low salt (125 mM NaCl) elution after buffer exchange using 50 kDa cutoff concentrator, 5 μl (Lane 3), pooled high salt (500 mM NaCl) elution after buffer exchange using 50 kDa cutoff concentrator, 5 μl (Lane 4). B. Immuno-detection image of cell lysate, 10 μl (Lane 1), pooled low salt (125 mM NaCl) elution after buffer exchange using 50 kDa cutoff concentrator, 2 μl (Lane 2), pooled high salt (500 mM NaCl) elution after buffer exchange using 50 kDa cutoff concentrator, 1 μl (Lane 3). For immuno-detection, primary antibody was polyclonal rabbit anti-6x-His. Molecular weight marker sizes (in kDa) are shown to left. Band corresponding to full length MBP-NsPox fusion protein marked by asterisk.
Figure 4.4. Absorbance vs. time plot of NsPox catalyzed 3, 3’,5, 5’-tetramethylbezidine (TMB) oxidization in neutral pH (Tris-HCl pH 7.4) and acidic pH (MES pH 6) buffer with or without supplementation with 10 mM CaCl$_2$. Blue diamond: 20 mM Tris-HCl (pH 7.4), magenta square: 20 mM Tris-HCl (pH 7.4) supplemented with 10 mM CaCl$_2$, green triangle: 20 mM MES (pH 6), purple x: 20 mM MES (pH 6) supplemented with 10 mM CaCl$_2$. Each reaction mixture includes 200 µl TMB working solution and 30 µl DEAE-sepharose high salt elution (after they were buffer exchanged to remove the high salt). Absorbance measured at 655 nm.
Figure 4.5: Immuno-detection of non-transformed and transformed *E. coli* BL21 cells. Non-transformed extracts, 10 µl (Lane 1); uninduced pMAL c2x NsPox cell crude extracts, 10 µl (Lane 2), induced pMAL c2x NsPox cell crude extracts, 10 µl (Lane 3). Primary antibody was polyclonal rabbit anti-6x-His. Molecular weight marker sizes (in kDa) are shown to left. Band corresponding to full length MBP-NsPox fusion protein marked by asterisk.
REFERENCE:


CHAPTER FIVE

FUTURE STUDY
The high tensile strength and biocompatibility of spider dragline silk makes it a desirable material in many engineering and tissue regeneration applications. Here, we present the feasibility to produce recombinant proteins in transgenic tobacco *Nicotiana tabacum* with sequences representing spider silk protein building blocks. Recombinant mini-spidroins contain native N- and C-terminal domains of major ampullate spidroin 1 (rMaSp1) or rMaSp2 flanking an abbreviated number (8, 16 or 32) of consensus repeat domains. Two different expression plasmid vectors were tested and a downstream chitin binding domain and self-cleavable intein were included to facilitate protein purification. We confirmed gene insertion and RNA transcription by PCR and reverse-transcriptase PCR, respectively. Mini-spidroin production was detected by N-terminus specific antibodies. Purification of mini-spidroins was performed through chitin affinity chromatography and subsequent intein activation with reducing reagent. Mini-spidroins, when dialyzed and freeze-dried, formed viscous gelatin-like fluids. Immuno-detection of this fluid showed various degrees of polymerization of mini-spidroin monomers.

We speculate that the ability of these mini-spidroins to retain water may relate to the post-translational modifications process such as phosphorylation and glycosylation. Protein phosphorylation detection kit such as Expedeon pIMAGO will be used to detect mini-spidroin phosphorylation. Mass spectrometry will be a better choice for the detection of which residues that may be phosphorylated inside the tobacco protein production process. Mass
spectrometry is also a choice to determine residues if and where glycosylation may occur on mini-spidroin proteins. The identification of glycol groups attached to mini-spidroin would be valuable for further investigate their role in fiber assembly.

Polyelectrolyte complex formation of mini-spidroin fibers using an anionic polyelectrolyte, gellan gum was achieved. Treated with acetic acid and crosslinked by glutaraldehyde, mini-spidroins and gellan gum solution formed a thin film at their interface. Immediate pulling of the film generated autofluorescent fibrous materials from each mini-spidroin alone or a combination of rMaSp1R8 and rMaSp2R8 (70:30). The fibers are 35 to 91 μm in diameter with overall toughness $1.444 \pm 0.976$, $0.584 \pm 0.227$, $0.551 \pm 0.356$ MJ/M$^3$ for rMaSp1R8, rMaSp2R8 and composite fiber, respectively. We added chitosan matrix to mini-spidroins and were able to continuously produce fibers as long as the spinning dope feeding were maintained. The fibers are 34 to 46 μm in diameter with overall toughness $1.75 \pm 0.668$, $1.926 \pm 1.15$, $1.242 \pm 0.45$ MJ/M$^3$ for rMaSp1R8, rMaSp2R8 and composite fiber, respectively. When as-spun fibers were air-dried and then rehydrated and stretched in water, the fiber diameter ranges from 34 to 42 μm and the overall toughness improved to $3.186 \pm 1.208$, $2.933 \pm 1.255$, $2.306 \pm 1.003$ MJ/M$^3$, respectively. To be able to post-spin process the fibers while keeping the production flow, we submerged the as-spun fiber in water immediately after they were pulled out from the interface. The fibers were then pulled out from water and stretched by the friction generated from pulling the
fibers against a smooth glass surface. We were able to produce fibers with average diameter 27 to 33 μm and overall toughness 3.442 ± 1.779, 3.01 ± 1.065, 3.453 ± 2.239 MJ/M³ for MaSp2R1, rMaSp2R8 and composite fibers, respectively. This study showed that spider silk-like fibers can be produced in large quantity through charge attraction which assembles chitosan, mini-spidroins and gellan gum into fibrous complexes. We speculate that the spider silk formation process in spider spinning duct may involve attraction of variously charged chitinous polymers, spidroins and glycoproteins.

In the spider major ampullate gland, a specific peroxidase activity was detected and the sequence of this enzyme was identified from the EST derived cDNA library. A synthetic gene for this enzyme was cloned into pMAL-c2x vector and the enzyme produced in E. coli BL21 cells. Two purification strategies were employed for NsPox purification. Amylose resin affinity purification resulted in little active enzyme. Anion exchange by DEAE-sepharose resin produced a large amount of active enzyme containing fractions that may be used for further purification. Preliminary NsPox peroxidase activity data using TMB as substrate reveals that this enzyme is more active under acidic conditions (pH 6.0). Presence of calcium ion inhibits NsPox peroxidase activity under both acidic and neutral pHs.

NsPox was detected in the uninduced BL21 cells, suggesting that leaky expression occurs. Growth medium containing 1% glucose may minimize leaky expression of NsPox BL21 cells. NsPox is a heme-binding protein, so adding iron
in the growth media may also increase the accumulation of NsPox in the bacterial cells. Several NsPox truncation bands were detected by the anti-6x-His antibody. Since the six histidine tag is engineered at the C-terminal end of NsPox, we doubt the truncation is due to incomplete translation in the *E. coli* cells. To determine if the truncation resulted from sonication, we will heat the NsPox cell pellet in extraction buffer and run on a SDS-PAGE gel without being sonicated. If sonication does cause truncations, we will employ a French press for subsequent purification. Eluates from high salt elution buffer can be used in sequential with nickel column. After obtaining NsPox in better purity, serial enzyme assays will be done to determine the kinetics of NsPox in neutral and acidic pH, with or without calcium ion as well as chloride ion.

Studies have demonstrated the evidence of protein crosslinking in the spider silk fibers (dos Santos-Pinto et al., 2014), in the silk of bees (Campbell et al., 2014) and oxidative stresses affect the stiffness of bysall fiber (Sun et al., 2001). Interestingly, amyloid peptides bind to heme and form a protein complex with peroxidase function (Atamna and Boyle, 2006). To further demonstrate the links between NsPox and protein crosslinking, the purified NsPox will be used for the crosslinking reaction of L-tyrosine to dityrosine. Production of dityrosine can be detected by fluorescence spectrophotometer at excitation/emission wavelength 315 nm/410 nm (Minamihata et al., 2011).

Another putative function of NsPox relates to the crosslinking NTD through the lysine and methionine residue in the NTDs. We will carry out reactions
containing purified rMaSp1A NTDs, NsPox in the presence of hydrogen peroxide. NTD crosslinking will be analyzed by SDS-PAGE. Any resulting higher molecular weight bands (dimer or trimer) will be analyzed by mass spectrometry to elucidate if covalent bonds are formed in the NTD dimer or even trimer association. We will also evaluate the possibility of fiber formation from mini-spidrion that we obtained from transgenic tobacco (Peng et al., In review) using NsPox as a crosslinker instead of chemical crosslinker glutaraldehyde (Peng et al., In preparation). This will provide direct evidence on the NsPox function in spider silk fiber formation.

The mini-spidroins that was used to form fibrous materials can also be used to make meshes, hydrogels and coatings through different assembly techniques. The diverse proteinaceous materials from mini-spidroin will open many opportunities in advanced material development for medical and other applications.
REFERENCES


APPENDICES
A Granted permission from American Chemical Society (1)
B  PLOS content license

Content License

The following policy applies to all of PLOS journals, unless otherwise noted.

PLOS applies the Creative Commons Attribution (CC BY) license to make its content accessible, free, immediate access to, and unrestricted reuse of original works of all types. Under this license, authors agree to make articles legally available for reuse, without permission or fee, for virtually any purpose. Anyone may copy, distribute, or reuse these articles, as long as the author and original source are properly cited.

Using PLOS Content

No permission is required from the authors or the publisher to reuse or repurpose PLOS content provided the original article is cited. In most cases, appropriate attribution can be provided by simply citing the original article.

Example citation:

If the item you plan to reuse is not part of a published article (e.g., a featured issue image), then indicate the originator of the work, and the volume, issue, and date of the journal in which the item appeared.

For any reuse or redistribution of a work, you must also make clear the license terms under which the work was published.

Figures, Tables, and Images

Figures, tables, and images are published under the Creative Commons Attribution (CC BY) license.

Data

If any dataset accompanying data is submitted to repositories with stated licensing policies, the policies should not be more restrictive than CC BY.

Submitting Copyrighted or Proprietary Content

Do not submit any figures, photos, tables, or other works that have been previously copyrighted or that contain proprietary data unless you have and can supply written permission from the copyright holder to use that content. This includes:
> images and scientific images
> tables and graphs
> software and content.
C  Granted permission from American Chemical Society (2)
D Granted permission from Elsevier
E Granted permission from Royal Society of Chemistry

---

**ROYAL SOCIETY OF CHEMISTRY LICENSE TERMS AND CONDITIONS**

Nov 22, 2015

This is a License Agreement between Congyue Peng ("You") and Royal Society of Chemistry ("Royal Society of Chemistry") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Royal Society of Chemistry, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>Field</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>License Number</td>
<td>3754390380575</td>
</tr>
<tr>
<td>License date</td>
<td>Nov 22, 2015</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Royal Society of Chemistry</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Soft Matter</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Combined structural model of spider dragline silk</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Roxana Ene, Periklis Papadopoulos, Friedrich Kremer</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Sep 9, 2009</td>
</tr>
<tr>
<td>Volume number</td>
<td>5</td>
</tr>
<tr>
<td>Issue number</td>
<td>22</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/educational</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/images</td>
</tr>
<tr>
<td>Number of figures/tables/images</td>
<td>1</td>
</tr>
<tr>
<td>Format</td>
<td>electronic</td>
</tr>
<tr>
<td>Distribution quantity</td>
<td>100</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>no</td>
</tr>
<tr>
<td>Order reference number</td>
<td>None</td>
</tr>
<tr>
<td>Title of the thesis/dissertation</td>
<td>Creating Biomaterials from Plant-derived Recombinant Spider Silk-like Proteins</td>
</tr>
</tbody>
</table>
F Granted permission from Nature Publishing Group (1)
G Granted permission from Nature Publishing Group (2)
Copyright Permission Policy

These guidelines apply to the reuse of articles, figures, charts and photos in the Journal of Biological Chemistry, Molecular & Cellular Proteomics and the Journal of Lipid Research.

For authors reusing their own material:

Authors need NOT contact the journal to obtain rights to reuse their own material. They are automatically granted permission to do the following:

- Reproduce an article for use in the author's courses (if the author is employed by an academic institution, that institution also may reproduce the article for teaching purposes)
- Reproduce an article for use in future commercial and non-commercial works
- Post a copy of the paper on the author's own website
- Link to the journal site containing the final edited version created by the publisher.

EXCEPTION: If authors select the Author's Choice publishing option:

- The final version of the manuscript will be covered under the Creative Commons Attribution License (CC-BY), the most accommodating of licenses offered. Click here for details.
- The final version of the manuscript will be released immediately on the publisher's website and PubMed Central.

Please note that authors must include the following citation when using material that appeared in an ABD journal:

"This research was originally published in {Journal Name, Author(s). Title, Journal Name. Year, Vol, Pages. © the American Society for Biochemistry and Molecular Biology."

For other parties using material for noncommercial use:

Other parties are welcome to copy, distribute, transmit and adapt the work — at no cost and without permission — for noncommercial use as long as they attribute the work to the original source using the citation above.

Examples of noncommercial use include:

- Reproducing a figure for educational purposes, such as schoolwork or lecture presentations, with attribution.
- Reprinting a condensed article as a Ph.D. dissertation, with attribution.
I  Granted permission from John Wiley and Sons

<table>
<thead>
<tr>
<th>License Number</th>
<th>3755410090433</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Nov 24, 2015</td>
</tr>
<tr>
<td>Licensed Content Publisher</td>
<td>John Wiley and Sons</td>
</tr>
<tr>
<td>Licensed Content Publication</td>
<td>Biopolymers</td>
</tr>
<tr>
<td>Licensed Content Title</td>
<td>Structure of silk by raman spectromicroscopy: From the spinning glands to the fibers</td>
</tr>
<tr>
<td>Licensed Content Author</td>
<td>Thierry Latèvre, François Paquet-Mercier, Jean-François Rioux-Dubé, Michel Pézolet</td>
</tr>
<tr>
<td>Licensed Content Date</td>
<td>Aug 31, 2011</td>
</tr>
<tr>
<td>Pages</td>
<td>15</td>
</tr>
<tr>
<td>Type of use</td>
<td>Dissertation/Thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>University/Academic</td>
</tr>
<tr>
<td>Format</td>
<td>Electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>Figure/table</td>
</tr>
<tr>
<td>Number of figures/tables</td>
<td>1</td>
</tr>
<tr>
<td>Original Wiley figure/table number(s)</td>
<td>Figure 10</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Creating Biomaterials from Plant-derived Recombinant Spider Silk-like Proteins</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Dec 2015</td>
</tr>
<tr>
<td>Expected size (number of pages)</td>
<td>200</td>
</tr>
</tbody>
</table>