



**Functional Spider Silk-Based Biomaterials
through Combinational DNA Libraries
Derived from Repetitive Core Genes Present
Primarily from *Latrodectus hesperus***

By Zaroug Jaleel

Research Advisor: Dr. David Kaplan

Senior Thesis, Spring 2017

Table of Contents

Abstract:	3
1. Introduction:	4
2. Methods	7
2.1.1 Plasmid and Library Construction	7
2.2 Characterization of Novel Materials	8
2.2.1 Expression and Purification of Recombinant Proteins	8
2.2.2 Protein Films and Characterization	9
2.2.4 Scanning Electron Microscope (SEM)	10
2.2.5 Atomic Force Microscopy (AFM) for Mechanical Properties	10
2.3 Statistical Analysis	11
3. Results	11
3.1 Construction of Dynamic Repetitive Core library	11
3.3 Secondary Structural Properties of Library A and Library B constructs.....	12
3.4. Recombinant Protein Film Morphology	14
3.4. Mechanical Properties	15
4. Discussion:	16
5. Conclusions:	24
6. Acknowledgements	25
7. References	26
8. Figures	29
Figure 1. Summary figure of the seven different silks and their respective proteins	29
Figure 2. Schematic of Spider silk gene domains	30
Table 1. Summary Table of primary sequences of individual monomeric gene constructs used in Library..	30
Figure 3. Establishment of the recombinant DNA libraries	32
Table 2. Summary Table of relevant library colonies.....	32
Figure 4. Amino acid composition and secondary structures of recombinant protein films	34
Table 3. Summary table of secondary structural motif counts	34
Figure 5. Mechanical characterization of selected protein films	36

Abstract:

The properties of native spider silk vary across species due to the presence of independent genes containing conserved repetitive core domains. Previous studies seeking to understand the function and material properties of these domains focused primarily on the analysis of *Nephila clavipes* dragline silk proteins, MaSp1 and MaSp2. Our work seeks to broaden the known mechanical properties of silk-based biomaterials by establishing two libraries containing genes from the repetitive core region of the native *Latrodectus hesperus* silk genome (Library A: *Masp1*, *Masp2*, *Tusp1*, *Acsp1*; Library B: *Acsp1*, *Pysp1*, *Misp1*, *Flag*). Secondary structural analysis of expressed Library A and Library B constructs through Fourier Transform Infrared Spectrometry (FTIR) revealed a higher portion of β -sheet content for recombinant proteins produced from gene constructs containing a combination of either *Acsp1/Tusp1* and *Masp1/Masp2* genes (Library A) while a higher portion of β -turn and random coil content in recombinant proteins from *Pysp1* and *Flag* genes (Library B). Mechanical characterization of selected purified protein films through Atomic Force Microscopy (AFM) suggested Library A recombinants had a higher elastic modulus when compared to Library B recombinants with Library A films having higher Young's modulus in the absence of methanol induced β -sheet formation while Library B films had a higher Young's modulus in the presence of methanol. Our work suggests that the repetitive core regions of the aforementioned genes play a significant role in the secondary structure motifs and the resulting mechanical properties within native silk, relevant for the construction of functional biomaterials. This work lays out a blueprint for the creation of more robust proteins through a similar multimerization strategy.

1. Introduction:

Materials inspired by nature are often responsible for some of the most important advances in modern science. One of nature's most enigmatic materials and one that stands at the forefront of modern research is spider silk. With a higher per-unit elastic modulus and a higher tensile strength of any man-made material, spider silk remains one of the most impressive and least understood materials found in nature (Gosline *et al.*, 1999; Vollrath and Knight, 2001; Agnarsson *et al.*, 2010; Omenetto and Kaplan, 2010). One current focus in biomaterials research is to be able to artificially replicate the spider silk protein (Scheibel, 2004; Lewis, 2006; Kluge *et al.*, 2008; Omenetto and Kaplan, 2010). However, cloning of silk presents both challenges and opportunities due to the multifaceted and versatile nature of the native protein.

The evolutionary success of spiders over the past 400 million years is due in part to the great versatility of silk (Sutherland *et al.*, 2010). Spiders require a multitude of functions in their daily lives from casting their webs (Blasingame *et al.*, 2009), catching prey (La Mattina *et al.*, 2008), covering eggs (Gnesa *et al.*, 2012), and using draglines (Ayoub *et al.*, 2007). Certain genes have been identified as responsible for determined properties of silk (Figure 1; Eisoldt *et al.*, 2011) (Vollrath and Porter, 2006). Major ampullate spidroin 1 and Major ampullate spidroin 2, denoted *Masp1* and *Masp2* respectively were found to play a key role in the formation of tightly knit β -sheet structures likely due to the presence of small peptide motifs namely poly-alanine domains (poly (A/GA)) which facilitate β -sheet formation and glycine rich domains (GGX) with X being either Y, L or Q facilitate 3_{10} helix formation (Hu *et al.*, 2006). These repetitive motifs are hypothesized to play a role in the high tensile strength of the dragline silk (Moon *et al.*, 1998; Ayoub *et al.*, 2007). The Minor Ampullate Spidroin I (*Misp1*) gene codes for prey-wrapping and auxiliary spiral silk (Colgin and Lewis, 1998). MiSp1 shares many properties with the Masp family

of silks including the presence of GGX and poly A motifs and mechanically behaves very similarly to Masp silk in both toughness and strength but has a superior extensibility than Masp silk (Blackledge and Hayashi 2006). Aciniform spidroin 1 (*Acsp1*) is responsible for the egg case and prey wrapping silk that must be hard and durable to withstand environmental stress (Hayashi *et al.*, 2004; Garb and Hayashi, 2005). Mechanically, *Acsp1*-based silk contains a higher toughness and extensibility compared to Masp silk but lacks in tensile strength (Tokareva *et al.* 2014). Tubuliform spidroin 1 (*Tusp1*) constitutes the hard outer casing of spider eggs (Hayashi *et al.*, 2004; Garb and Hayashi, 2005). On a molecular level, Tubuliform silk is unique in that it contains β -sheet structures albeit through the presence of glutamine and serine interspersed in poly A regions (Tian and Lewis, 2006) which are theorized to facilitate greater spacing of the resulting β -sheets compared to Masp1 and Masp2 poly A regions contributing to the high hydrophilicity of Tubuliform silk (Tokareva *et al.*, 2014). Pyriform spidroin 1 (*Pysp1*) comprises the adhesive attachment disks that helps anchor the dragline and the webs to substrates (Blasingame *et al.*, 2009; Perry *et al.*, 2010). Although not much is known about the mechanical properties of this class of spidroin proteins, recent papers outlining primary sequence motifs in *Pysp1* have identified large stretches of alternating proline (Px) or glutamine (QQxxxx) motifs that are unique to pyriform silk and are suggested to play a role in its adhesiveness and elasticity (Chaw *et al.*, 2017). Finally, Flag silk (Flag) comprises the capture spiral of a spider web and thus must be flexible and elastic (Hayashi and Lewis, 1998). The characteristic extensibility of Flag is that it can stretch up to 200% of the original length due to the presence of several GPGXX motifs that create a β – spiral motif comprised of two linked type II β -turns (Tokareva *et al.*, 2014; Adrianos *et al.*, 2013). The gene designs used in this study except for Flag were derived from sequenced cDNA from the black widow spider *Latrodectus hesperus*. The genes from *L. hesperus* were used due to the superior

mechanical properties of native *L. hesperus* silk in both toughness and strength compared to other spider species (Ayoub, 2007). However, due to the superior elasticity of *N. clavipes Flag* silk (Hayashi and Lewis, 1998), the cDNA from the repetitive core region of *N. clavipes Flag* silk was used as the basis for the Flag gene design used in our library study.

A primary goal of silk research is to create materials that mimic the properties of native silk protein *in vitro*. One of the main challenges to this comes from the complexity of the silk protein itself. Full-length spider silk is composed of a highly repetitive Glycine and Alanine core and can have monomers up to 120 kDa in length (Scheibel, 2004), which makes direct chemical synthesis of spideroin proteins extremely difficult. Furthermore, N-terminal and C-terminal regions present within native silk play a key role in the folding and subsequent properties of the protein (Figure 2; Heim et al. 2009). Previous research has created DNA libraries that have been constructed from genes present within N-terminal and C-terminal domain regions which are primarily responsible for the folding of the silk molecule (Ayoub, 2007; Huang et al., 2015; Wang., 2014). Furthermore, much of the focus of these early silk libraries has been on the dragline silk genes *Masp1*, *Masp2*. The current project differs from previous research in that the focus of the library is to broaden the mechanical properties of silk-based biomaterials by including the repetitive core regions of the egg case silk genes (*Tusp1*, *Acsp1*) and the structural web genes (*Pysp1*, *Misp1*, *Flag*).

Two spider silk libraries derived primarily from the repetitive core genes present in *L. hesperus* were constructed to study the functional relationships between these genes. The first library consisted of *Masp1*, *Masp2*, *Tusp1*, and *Acsp1* core genes and was focused on generating materials with high stiffness and durability, while the second library consisted of *Misp1*, *Acsp1*, *Pysp1*, and the *Flag* (*N. clavipes*) core genes and was focused on generating elastic materials with

low stiffness and properties like rubber. Through this strategy, we sought to broaden the mechanical properties of silk-based biomaterials.

Mechanical and secondary structural properties of selected protein chimeras were analyzed resulting in the identification of three recombinants with varied and unique Young's moduli and secondary structural motifs that serve as a basis for the formation of future silk-based biomaterials. Secondary structural properties of the resulting recombinant proteins from the library were characterized through Fourier Transform Infrared Spectrometry (FTIR) to analyze β -sheet content, while the surface morphology was analyzed by Scanning Electron Microscopy (SEM). To analyze mechanical properties, Atomic Force Microscopy (AFM) was carried out to determine the Young's Modulus (E) a measure of a materials stiffness or resistance to deformation.

2. Methods:

2.1.1 Plasmid and Library Construction

All molecular biology procedures followed standard protocols. Figure 3a depicts the generalized schematic for the cloning strategy used (Figure 3a). First, the parent plasmid pET19b3 (Xia et al 2011) was cleaved using the restriction enzymes *NdeI* and *BamHI* to eliminate the *BanII* restriction enzyme site. Annealing of the b4 sense and antisense linker (5' TATGGCTAGCGGGCTCACTAGTTAAG 3') created the b4 insert, which was then ligated into the pET19b plasmid using T4 DNA ligase (New England Biolabs, Ipswich, MA, USA). The b4 insert contained one restriction enzyme site for *BanII* as well as restriction enzyme sites for *NheI* and *SpeI* to allow for head to tail ligation of monomeric inserts obtained from the original screens.

DNA sequences for each of the seven genes were chemically synthesized (GenScript, Piscataway, NJ, USA) and were designed to include the first two monomeric subunits of the repetitive core region (Table 1). The resulting DNA sequence was then isolated from the pUC57 plasmid by gel electrophoresis and purified using a Gel Extraction kit (Qiagen, Hilden, Germany) and cloned into the *Ban*II restriction site on the pET19b4 vector overnight at 4°C. Establishment of the library was done through the isolation of the pure spidroin gene fragments via *Ban*II restriction enzyme digestion, and subsequent ligation of the fragments together mediated by the complimentary *Ban*II sticky ends followed by insertion of the recombinant fragments into the pET19b4 expression vector (GenScript, Piscataway, NJ, USA). For creation of targeted DNA sequences such as A10₂ and B10-22-17₂, head-to-tail ligation of monomeric recombinant inserts was done through the *Nhe*I and *Spe*I restriction sites. All the resulting plasmids containing the combined spidroin genes were then transformed into competent DH5 α *Escherichia coli* cells (New England Biolabs, Ipswich, MA, USA). Robust screening of 652 colonies for inserts greater than 1kb revealed 15 total colonies with unique insert arrangements.

2.2 Characterization of Novel Materials

2.2.1 Expression and Purification of Recombinant Proteins

DNA obtained from the recombinant DNA library was transformed into competent *E. coli* BL21* (New England Biolabs, Ipswich, MA) for expression as described in Zhou et al 2015. Briefly, transformants were cultured in a 1 L hyper-broth media (Sigma-Aldrich, Natick, MA, USA) containing yeast extract medium 25 g/L and a 15% glucose nutrient mix as well as ampicillin (100 μ g/mL) at 37 °C under aeration condition for 2 hours. Once OD₆₀₀ reached 0.8 and cultures were in logarithmic growth phase, protein expression was induced with 1mM Isopropyl- β --D-thiogalactopyranoside (IPTG) (Sigma-Aldrich, Natick, MA, USA) and subsequent activation of

the T7 promoter on the plasmid. After 6 h the culture was centrifuged at 7500 x g for 20 minutes and the resulting pellet was collected and stored at -20 °C. The pellet was lysed in 8 M urea (pH=8.0) overnight after which Nickel-NTA beads (Qiagen, Hilden, Germany) were added to bind His-tag on the proteins. Protein column purification was done using urea buffered at decreasing pH intervals = 8.0, 6.3, 5.9, 4.5 with the targeted protein being eluted at pH = 4.5. The resulting elutants were run on an SDS-Page gel and stained with Coomassie Blue (Sigma-Aldrich, Natick, MA, USA) and the vials containing the target proteins were prepared for dialysis in 50 mM phosphate buffer at pH=5.4 and then water to remove the urea and induce proper folding of the proteins. The resulting solution was then lyophilized to obtain the pure protein in solid form.

2.2.2 Protein Films and Characterization

Purified lyophilized protein samples were dissolved in hexafluoro-2-propanol (HFIP) (2.5% w/v) (Sigma-Aldrich, Natick, MA, USA) at room temperature overnight to dissolve the silk proteins. Six films of each sample were prepared by depositing 30 μ L of each dissolved protein solutions onto different PolyDimethylsiloxane (PDMS) disks (R = 6 mm). Solvent was removed by evaporation in a fume hood at room temperature for 1 h. To induce β -sheet formations films were immersed in 70% methanol overnight and then air dried.

Fourier Transform Infrared Spectroscopy (FTIR) of the films were carried out with an Attenuated Total Reflectance FTIR machine (Jasco, Oklahoma City, OK, USA). An air background was taken for each sample. Deconvolution of the peaks was done on the PeakFit software provided by SigmaPlot. Ten Gaussian peaks were selected across the amide I absorption band (1720-1580 cm^{-1}). Four secondary structure motifs were analyzed that corresponded to β sheet (1618–1629 cm^{-1} and 1697–1703 cm^{-1}), α -helix (1658–1667 cm^{-1}), β -turns (1668–1696 cm^{-1}) and random coil (1630–1657 cm^{-1}) (Zhou *et al.*, 2015).

2.2.4 Scanning Electron Microscope (SEM)

Selected lyophilized recombinant proteins A10₂ (*Masp1-Masp2-Masp1-Acsp1*)₂, B10 (*PySp1₃-Flag*), B12 (*PySp1-MiSp1-MaSp1-Flag-AcSp1*) and B10-22-17₂ (*(PySp1)₃-Flag-PySp1-MiSp1-PySp1-(PySp1₂-AcSp1)₂*) were dissolved in HFIP (2.5% w/v) and cast on a silicon surface. The specimens were then sputter-coated with gold, enhancing surface conductivity and then examined using a Zeiss Eco MA10 (Carl Zeiss, Oberkochen, Germany) to observe surface morphologies (Xia *et al.*, 2011).

2.2.5 Atomic Force Microscopy (AFM) for Mechanical Properties

Atomic Force Microscopy studies were carried out by Dr. Lauren Baugh. Films were created from selected samples A10₂ (*Masp1-Masp2-Masp1-Acsp1*)₂, A261 (*Tusp1₂-Acsp1-(Masp1)₂*) and B10-22-17₂ (*(PySp1)₃-Flag-PySp1-MiSp1-PySp1-(PySp1₂-AcSp1)₂*) as described in 2.2.2. Briefly, lyophilized proteins were dissolved overnight in HFIP (2.5% w/v) and cast on precleared glass slides at room temperature and allowed to dry in a vacuum chamber to minimize bubble formation. A Veeco (Town of Oyster Bay, New York) Dimension 3100 Atomic Force Microscope was used to take force-displacement measurements over a 20 x 20 μm area using a 16 x 16 grid. A 5μm borosilicate glass beaded probe with a 0.06 N/m spring constant (Novascan, Ames, Iowa) was used to take all modulus data. To avoid interference with the underlying glass coverslip, a 750nm ramp depth was used for all measurements. Using the force-displacement curves, the Young's Modulus was calculated using the Hertz model (Hu et al 2010, Yamada et al 2003, Rotsch et al 1999, Domke and Radmacher 1998) assuming a Poisson's ration of 0.5. Topographical images of each of the samples were taken with the same AFM, but using a non-contact tip (model FESP, Bruker, Billerica, Massachusetts). Representative images are 20 x 20 μm and were processed using Nanoscope Analysis 1.5 software.

2.3 Statistical Analysis

Quantitative analysis was done in at least in triplicate. Average values were plotted with error bars representing the standard deviation. Comparisons between $k > 2$ samples was done through a one-way ANOVA and a post-hoc Tukey Kramer test was done for multiple comparisons. Statistical analysis between two treatments of the same sample was done using a two-sample paired students t-test of the means. All statistical analysis was done with a significance level of $p < 0.05$. Graphs were plotted on Microsoft excel.

3. Results

3.1 Construction of Dynamic Repetitive Core library

Establishment of two libraries was conducted according to theorized material properties of the original protein constructs (Figure 3a) with the core amino acid motifs are shown in Table 1. Recombinant proteins from Library A were designed to have high durability and tensile strength due to the combined material properties of the egg case silk components and the dragline silk components respectively. In contrast, Library B was designed to have more bio-rubber characteristics due to the elastic and adhesive properties of the respective gene products. The seven monomeric genes were chemically synthesized (Table 1) and subsequently multimerized using a “concatemerization” strategy utilizing a newly synthesized pET19-b4 expression vector. The vector contained the restriction site *BanII* for initial formation of the ligated library monomers (Figure 3a). Subsequent creation of targeted recombinants (A10₂, B10-22-17₂) with the aim of creating more robust proteins with higher molecular weights similar to native spider silk was carried out using the two flanking linker directional restriction enzyme sequences *NheI* and *SpeI* present in the pET19-b4 vector which allowed for directional polymerization and recombination for the initial library screens. This cloning strategy avoided the incorporation of extra amino acid

residues at the junction of two monomeric genes, thereby preserving sequence integrity of the resulting protein.

Upon randomized ligation of the repetitive core genes into their respective libraries, transformants were screened for presence of library inserts. Validation of library insertion was done by restriction enzyme digestion (*NheI* and *SpeI*) and visualization of separated fragments. 652 colonies were screened and 15 were selected for further investigation, including 9 from Library A and 6 from Library B, due to their varied gene patterns and for their higher molecular weights (Table 2). To obtain greater molecular weights more representative of native spider silk, targeted polymerization studies utilizing the *NheI/SpeI* restriction complex in the expression plasmid were conducted resulting in three major polymerized recombinants with more robust molecular masses A10₂ (Masp1-Masp2- Masp1-Acsp1)₂, A261 (Tusp1₂-Acsp1-(Masp1)₂) and B10-22-17₂ ((PySp1)₃-Flag-PySp1-MiSp1-PySp1-(PySp1₂-AcSp1)₂). In total, fourteen constructs (A10, A11, A15, A46, A52, A217, A219, A261, A10₂, B10, B12, B17, B10-22-17, B10-22-17₂) were produced, purified and their approximate molecular weights determined by SDS-Page (Figure 3b). The results confirmed a varied combination of inserts of different molecular weights between 22 – 88 kDa were obtained (Figure 3b) and selected for further secondary structure characterization studies.

3.3 Secondary Structural Properties of Library A and Library B constructs

Fourier Transform Infrared Spectroscopy (FTIR) analysis was done to analyze secondary structure content in selected constructs (A10, A11, A15, A46, A52, A217, A219, A261, A10₂, B10, B12, B17, B10-22-17, B10-22-17₂). Four major secondary structural motifs were analyzed namely α -helices, β -sheets, β -turns and random coils. Because β -sheet concentration is an important characteristic of material durability and toughness, β -sheet induction was carried out by

addition of methanol. Primary sequence motifs were also analyzed from the translated sequencing data of five selected constructs (B10-22-17₂, A10₂, A261, A219, B10-22-17) with an emphasis on amino acid composition and four structural domains common amongst spider silk proteins. These domains were GPGXX motifs that correlate with elastic beta spiral structures, GGX motifs that correlated with amorphous 3₁₀ helix domains, and finally poly A/GA and AAQAA/AASQSA motifs that correlated with crystalline β -sheet structures (Table 3). The presence of these motifs and the observed FTIR secondary structures after methanol treatment were compared.

Amino acid composition analysis for the primary sequences of the five selected constructs (B10-22-17₂, A10₂, A261, A219, B10-22-17) revealed Library B recombinants (B10-22-17₂, B10-22-17) contained a larger percentage of positively charged (lysine, arginine) and negatively charged (glutamate, aspartate) amino acid residues than all three Library A recombinants (A10₂, A261, A219) (Figure 4c). A219 had a higher percentage of its amino acid composition comprised of glycine and alanine compared to the other recombinants (Figure 4c). A261 on the other hand had the highest percentage of glutamine compared to other recombinants (Figure 4c).

FTIR spectra were analyzed across the Amide I absorption band (1580-1720cm⁻¹) due to the high signal to noise ratio in that region compared to the Amide II absorption band, as well as the literature concerning silk films in the amide I region (Ravindra et al., 2001; Kong, 2007; Zhou et al., 2015; Hu et al., 2006; Huang et al., 2011). Addition of methanol induced β -sheet formation in all the experimental samples (Figure 4b). This was supported experimentally by a decrease in the gaussian peak size for each recombinant protein at 1650cm⁻¹ and 1660cm⁻¹ with a corresponding increase in peak size at 1620cm⁻¹ which correlated with a decrease in random coil and α -helix with an increase in β -sheets, respectively (Figure 4b).

Within the library samples, it was found that Library A recombinants ($31.3\% \pm 1.9$) on average contained higher percentage β -sheet compositions than Library B recombinants ($24.1\% \pm 5.7$) excluding the A219 and B10-22-17₂ recombinants (Figure 5a1). A219 was unique amongst library A fragments with a significantly lower β -sheet composition ($16.8\% \pm 0.6$) and a significantly higher β - turn composition ($22.8\% \pm 0.06$) (Figure 5a1, 5c). From the primary sequence, A219 was also unique in that it had highest number of GPGXX motifs (6) in Library A as well as the highest number of combined poly-A/GA motifs (11) (Table 3). A10₂ and A261 both had unique secondary structural motifs. A10₂ contained the highest amount of GGX motifs as well as one of the highest β -turn percentages amongst Library A ($21.2\% \pm 0.6$) (Table 3). A261 was the only sample amongst the ones analyzed with AAQAA/AASQSA motifs (Table 3). Likewise, it also contained the highest percentage β -sheet amongst all analyzed recombinants ($34.4\% \pm 0.9$). In Library B, B10-22-17 had one of the smallest uninduced β -sheet compositions ($8.1\% \pm 0.4$) (Figure 5a1), yet the largest methanol-induced fold change in β -sheet percentage of over 200%. Likewise, B10-22-17 had the largest percentage β -turn composition amongst both libraries ($26.9\% \pm 1.1$) (Figure 5a3). B10-22-17₂ on the other hand had the reverse trend with one of the highest uninduced and induced β -sheet percentages (29.0 ± 0.16 and 32.5 ± 0.9) and one of the lowest β -turn percentages (11.6 ± 3.6) amongst Library B recombinants. Both samples had the highest number of GPGXX motifs (8) amongst all proteins with the only difference between them being the presence of an extra GGX motif in B10-22-17₂ (Table 3).

3.4. Recombinant Protein Film Morphology

To study surface morphologies of the recombinant proteins, films were again cast on HFIP and aliquotted onto PDMS surfaces. SEM was carried out to observe morphology changes in protein films between uninduced and induced samples. Library A had a smoother texture than their

Library B counterparts (Figure 3c). Upon addition of methanol to induce β -sheet formation there was a slight bubbling of the films (Figure 3c). Library B colonies demonstrated a shift in morphology from smooth to spherical structure formation. Within the B12 and B17 samples, protein deposits were noticed underneath the film that contributed to the roughness (Figure 3c).

3.4. Mechanical Properties

To study the mechanical properties of the recombinant proteins, nano-indentation studies were done by AFM to study the mechanical properties of the recombinant proteins. Three representative proteins were chosen A261, A10₂ and B10-22-17₂ due to a) high expression efficiencies in BL21* *E. coli*, b) robust molecular weights which made them more representative of native spider silk proteins c) their high β -sheet compositions which suggest enhanced material properties and d) their diverse composition of the original monomeric proteins with A10₂ containing large amounts of major ampullate silk proteins, A261 containing large amounts of egg case proteins, and B10-22-17₂ containing a diverse array of all the Library B genes with a large amount of PySp1, the adhesive silk protein (Figure 5a). To control for molecular weight effects the two Library A colonies were also chosen due to their similar molecular weights around 56 kDa, with Library B having a slightly larger MW of 81 kDa.

Overall, the two Library A samples A261 and A10₂ despite having a lower molecular weight had a significantly higher elastic modulus than B10-22-17₂. For B10-22-17₂ it was found that the addition of methanol significantly increased the Young's Modulus of the resulting film from 3.4 GPa to 4.0 GPa (Figure 5c). However, for A261 and A10₂ the presence of methanol decreased the observed Young's modulus of the films from 24.1 and 31.7 GPa pre-methanol treatment to 15.1 and 20.0 GPa post-methanol treatment for A261 and A10₂ respectively (Figure

5c). Between Library A samples A10₂ had a higher elastic modulus in both the methanol induced samples and the uninduced samples compared to the corresponding A261 sample (Figure 5c).

When the nanoscale Young's modulus of the recombinant films was compared to the Young's modulus of spun fibers of the native full-length proteins that comprised the library as outlined in Blackledge and Hayashi., 2005, it was observed that the Young's modulus for A10₂ and A261 were both significantly higher than all the native proteins. On the other hand, B10-22-17₂ had a significantly smaller Young's modulus when compared to the native proteins. The only exception was with the highly elastic native Flag protein in which B10-22-17₂ contained a significantly higher Young's modulus (Figure 6d).

4. Discussion:

Establishment of two DNA libraries was conducted with a focus on creating novel biomaterials through combining different repetitive core modules present within the *L. hesperus* and *N. clavipes* genome. The library proteins chosen were screened to contain a diverse array of incorporated motifs and a degree of robustness aimed to mimic varied native silk-like properties such as high extensibility and low stiffness, as well as versions with high stiffness and high toughness. Acsp1 was included in both libraries because it contains a high Young's modulus and toughness (Blackledge and Hayashi., 2006) useful for the stiffness and strength properties desired in Library A recombinants as well as a high extensibility (Blackledge and Hayashi., 2006) useful for providing the tough yet elastic properties desired in Library B.

The libraries were created using a concatamerization strategy which would allow for further polymerization of the multimerized library colonies to obtain higher molecular weights with mixed motifs based on silk modules. The DNA sequences used for the formation of the library were derived from the monomeric units of the repetitive core regions for each of the seven genes

from *L. hesperus* and *N. clavipes* (Flag) cDNA. Prior research demonstrated the superior mechanical properties of *L. hesperus* Major ampullate silk (Ayoub et al., 2007). While sequence homology is conserved across many spider species *L. hesperus*, is unique as it contains several higher order repetitive units throughout its genome (Ayoub et al., 2007). These higher order units consist of stretches of different amino acid motifs including GGX and poly A repeats and can be from 70-2000 amino acids in length. The sequences selected for the library were designed to include sections of these higher order units. Likewise, small differences in amino acid sequences, such as high levels of S and Q in tubuliform silk can contribute to enhanced material properties of *L. hesperus* silk which are reflected in the sequences used for library design and construction.

SEM was used to shed insight into the surface morphologies of the recombinant silk proteins in film format. Recombinant A10₂, B10, B12, B17 and B10-22-17 were analyzed via SEM due to their high purity and expression yields. The Library A recombinants had film morphologies that were smoother and consistent in texture while Library B formed rougher films (Figure 5). It is possible that this disparity is due to the higher presence of *MaSp* genes in A10₂ fraction than in Library B that provides films with greater structural stability.

Characterization studies through FTIR were used to examine changes in secondary structure composition of films generated from the recombinant proteins. Deconvolution of the amide I spectra showed that the addition of methanol resulted in a decreased peak size at 1650cm⁻¹ and 1660cm⁻¹ with an increased peak size at 1620cm⁻¹ suggesting a change in structure from random coils and α -helix to a more ordered β -sheet structure (Figure 5b). This remains consistent with expectations as addition of methanol has been shown to induce β -sheet formation in polypeptides (Ravindra et al 2001, Kong 2007, Zhou et al 2015, Hu et al 2006, Huang et al 2011).

The constructions from the library show similar secondary structural compositions when compared to their native spidron proteins as reported in previous publications. Native Major ampullate silk, from which MaSp1 and MaSp2 are derived from, is comprised of approximately 34% β -sheet structures (Lewis *et al.*, 2006). In Library A constructs, it was observed that most of the recombinant proteins contained β -sheet concentrations approximately near this value with a mean value of $31.3\% \pm 1.9$ (Figure 5a1). Interestingly, the recombinants containing purely MaSp1/MaSp2 motifs – the main constituents of dragline silk, only showed a maximum β -sheet percentage of 16.8%, which was significantly lower than the expected percentage of β -sheets in dragline silk genes. Interestingly, conjugation of these dragline silk recombinant construct motifs with those found in TuSp1 or AcSp1 increased the percentage of β -sheets in resulting protein films (A15, A52, A217, A10₂, A261). The results could be explained by a model in which AcSp1 and TuSp1 proteins enhance the β -sheet percentage through addition of certain serine and glutamine-heavy amino acid repeat motifs in TuSp1 and GGX motifs from AcSp1. Indeed, when compared to A219, A10₂ contained a higher content of GGX motifs while A261 contained poly-A motifs containing S and Q with both proteins containing higher β -sheet percentages compared to A219 (Table 3). These data suggest that the interplay between these motifs plays a role in the secondary structural properties of the resulting protein. This is likely due to the large number of glycine/alanine residues present in MaSp1/MaSp2 which provide flexibility to the polypeptide backbone allowing it to more readily adopt a β - turn conformation, while in the presence of the bulkier S and Q side chains in AcSp1/TuSp1 provide just enough steric interference to allow for the adaptation of the β -sheet conformation. Indeed, it was also observed in recombinant A261 with a larger TuSp1/AcSp1 to MaSp1/MaSp2 ratio that there was a natural inclination of the protein to adhere closer to the 34% β -sheet structure present in native dragline silk (Figure 5a1).

These data suggest that the core domains in spider silk in the *MaSp1/MaSp2* genes do not play as large of a role in the presence of large β -sheet percentages in their respective proteins when present in isolation. Our hypothesis is that the interactions with the N-terminal and C-terminal domains would allow the full-sized protein to adopt the proper conformation needed to induce the characteristically high β -sheet.

For the B10-22-17 recombinant of Library B, with an abundance of PySp1 as well as the Flag gene from *N. clavipes*; the resulting protein had the highest percentage β - turn as well as the highest change in β -sheet composition after methanol treatment. The Flag protein contains a high per unit elasticity by being able to stretch more than 200% (Hayashi and Lewis, 2000; Gosline et al 1996). It is theorized that the high β - turn percentage, due to repetitive GPGXX motifs, gives Flag this ability (Hayashi and Lewis, 2000). Both B10-22-17 and B10-22-17₂ have eight of these motifs, the highest of all the library recombinants. PySp1 is the adhesive silk necessary for the binding of the webs and draglines to their substrates. Although little is known of its secondary structure and mechanical properties, it has been theorized to be mainly constituted by random coils (Perry et al., 2010) which give piriform silk its flexibility and adhesive properties. This would explain the large random coil presence in the B10 recombinant that had the highest percentage of its molecular mass comprised of PySp1. However, upon the addition of a PySp1₂-AcSp1 unit to B10-22-17 the β -turn percentage decreased while the β -sheet percentage increased. This could be explained by the Flag protein now occupying a smaller percentage of the total protein mass thus diluting the effects of the eight GPGXX motifs. This in conjunction with an increase in the β -sheet-favored AcSp1 gene could explain the reversal of secondary structures between these two recombinants. In both libraries, we can see that the amalgamation of different genes creates a family of varied and unique proteins, providing a source for biomaterial designs.

Materials with a high Young's modulus can withstand a large amount of stress before deformation. Materials with a low Young's modulus are elastic and deform easily. Spider dragline silks are known for their high stiffness and tensile strength due to the composition as structural supports for spider webs (Moon *et al.*, 1998). It is known that Major ampullate silks derived from *L. hesperus* have a high Young's moduli (Blackledge *et al.*, 2005). Tubuliform and Aciniform silks which comprise the egg case and prey wrapping silks, respectively, must also be resistant to deformation to protect eggs and to prevent prey from escaping (Hayashi *et al.*, 2004; Tokareva *et al.*, 2014), and likewise have high per unit Young's moduli similar to Major ampullate silk, but with an increased extensibility (Blackledge and Hayashi, 2006). In contrast, Flag silk which comprises the capture spiral is one of the most elastic materials found in nature and have a low Young's modulus. Little is known about the mechanical properties of PySp1 apart from its high adhesiveness with recent papers reporting that Pyriform Silk can adhere to substrates with a force up to 40 mN (Wolff *et al.*, 2015). Our work seeks to characterize the role of these spider silk components on recombinant silk-based biomaterial properties.

Upon AFM analysis of three representative recombinant protein films A10₂, A261 and B10-22-17₂ - it was found that B10-22-17₂ had a five-fold lower Young's modulus when compared to A10₂ and A261 (Figure 5c). Between the two Library A recombinants, A10₂ protein films had a higher Young's modulus including both the methanol induced and uninduced samples when compared to the corresponding A261 films (Figure 5c). A possible explanation for the data lies in the genomic makeup of the recombinant proteins. Library A consisted of sections of genes that coded for inelastic, stiff and tough materials, the dragline silk and egg case silk genes. In contrast, Library B consisted of sections of genes that coded for the highly elastic, extensible and adhesive silks. The repetitive core region of the dragline and egg case silk genes present in the recombinant

Library A proteins contain amino acid residues and motifs that contribute to the significantly higher Young's modulus compared to B10-22-17₂ which consisted of repetitive core regions of the extensible Flag and the adhesive pyriform genes. Both A261 and A10₂ contained more β -sheet motifs, poly A/GA and AAQAA/AASQSA, while also containing fewer elastic β -spiral GPGXX motifs than B10-22-17₂ (Table 3). The interplay between these motifs could have led to the observed results. It is also possible that non-canonical amino acid motifs present in the PySp1 core gene could be playing a role in the mechanical properties of the PySp1 heavy B10-22-17₂ recombinant. Blasingame et al identified an AAARAQAQAE motif in *L. hesperus* Pysp1 that is unique amongst araneid pyriform silk (Blasingame et al., 2009). Recent publications by Chaw et al elucidate upon this motif by comparing the primary sequence of *L. hesperus* pyriform silk to that of other araneids. They found that the core repetitive region of most pyriform silk contains large proline rich (Px) and glutamine rich (QQQx₅) regions which are theorized to play a role in the formation of random coil structures and a subsequent high elasticity of the resulting protein. However these motifs are largely missing in the repetitive core region of *L. hesperus* in favor of the aforementioned AAARAQAQAE motif as well as an alternating ARAKAE. In their paper they speculate that the absence of these proline and glutamine rich regions results in a higher stiffness in *L. hesperus* pyriform silk as opposed to pyriform silk of other araneids (Chaw et al., 2017).

We would like to expand upon Chaw et al's analysis. The AAARAQAQAE that both Chaw et al and Blasingame et al discuss is repeated 36 times in B10-22-17₂ (Table 2). We speculate that favorable steric and electrostatic interactions between the positively charged arginine, and the negatively charged glutamate side chains could facilitate the formation of salt bridges between the charged states of the resulting recombinant protein. These salt bridges coupled with hydrogen bond interactions would strengthen intermolecular and intramolecular interactions between adjacent and

localized R and E side chains and could play a role in the adhesiveness of the native piriform silk protein. However, compared to the Library A recombinants which lack these localized R and E motifs it is also possible that these salt bridge interactions in conjunction with the bulky glutamine residues would create greater spacing between secondary structures thus leading to a less tightly packed recombinant protein, when formed into films, and thus a lower Young's modulus when compared to Library A films. Furthermore, the absence of the poly-proline repeats in the repetitive core region of *L. hesperus* piriform silk could explain the low random coil and high β -sheet and alpha helix percentages in the recombinant B10-22-17₂. The conformational rigidity of proline prevents the formation of higher order secondary structures and could explain the high random coil composition of piriform silk. However, the absence of these regions in the *L. hesperus* PySp1 construct that comprised B10-22-17₂ would give the protein backbone the flexibility to adopt a more stable helical or β -sheet conformation that would otherwise be unfavorable in the presence of the large proline repeats found in other araneid piriform proteins.

Between Library A proteins, the increased Young's modulus in A10₂ compared to A261 could be due to differences in the repetitive alanine heavy regions. In A10₂ the poly A regions are primarily a string of poly A/GA repeats (Table 3). On the other hand, A261 had fewer A/GA repeats but instead had more poly A repeats interspersed with Q and S from the presence of the *Tusp1* core gene. The presence of Q and S in the native TuSp1 protein provides greater spacing between β -sheets. In the absence of these two amino acid residues in the poly A region, such as in the sequences for MaSp1 and MaSp2 present in A10₂, the β -sheets are more tightly packed thus explaining the higher Young's modulus in the recombinant A10₂. It is also likely that the interplay between GGX motifs and other amino acid side chains could facilitate and stabilize β -sheet formation. High levels of GGX motifs in native Flag silk had the potential to form β -sheet

structures by themselves as well as facilitate alignment of other amino acid motifs to form β -sheet structures (Adrianos et al., 2013). With an increasing number of GGX motifs correlating with an increase in Young's modulus across the three recombinants this model could also play a role in the observed results.

The addition of methanol had differing effects between Library samples. For Library B the addition of methanol increased the Young's modulus significantly, likely due to the increase in β -sheets. For Library A however, the addition of methanol decreased the Young's modulus for both A10₂ and A261 significantly. We speculate that the high percentage α -helices (approximately 16%) in the uninduced conformations for both proteins could facilitate stabler coiled coil intermolecular interactions thus providing stability to the resulting material and thus a higher Young's modulus.

Finally, in comparing the nanoscale AFM of the recombinant films to the macroscale Young's modulus of the native spider silks fibers that comprise the library, it was found that both A10₂ and A261 had higher Young's moduli than all of the native spider silk proteins. B10-22-17₂ on the other hand had a lower Young's modulus than native silk proteins except for Flag, which it had a significantly higher Young's modulus (Figure 5d). However, as noted the Young's modulus measurements made in the study are on nanoscale measurements whereas Young's moduli measurements of native silks are based on macroscale using Instron tensile testing (Blackledge and Hayashi, 2006). Due to the relative stability of nanoscale material properties when compared to the macroscale level, we expect our recombinant proteins to have a lower Young's moduli at the macroscale than what is currently observed at the nanoscale through AFM. However, previous research comparing mechanical properties of recombinant spider silk films to their corresponding fibers found that fibers were significantly stronger and had a higher Young's modulus at the

macroscale level compared to their corresponding film at the macroscale level (Tucker, 2014). Furthermore, when monomeric units of the repetitive core region were incorporated in the recombinant proteins, despite their relatively short sizes of 50-80 kDa compared to native silk which can be up to 300 kDa in length, they nevertheless contained relatively large Young's moduli on the nanoscale level. Whether these high Young's moduli exist on a macroscale level is a topic for future study.

5. Conclusions:

Functional spider silk-based biomaterials were produced from constructs created through the random ligation of spideroin genes derived from the first monomeric subunit within the core repetitive region of *L. hesperus* and *N. clavipes*. Secondary Structure characterization of the original recombinant constructs was carried out through FTIR to determine secondary structure. It was observed that on average, Library A recombinant proteins with AcSp1 and TuSp1 contained higher overall β -sheet compositions compared to Library B samples, possibly due to the presence of more poly A repetitive regions that are more susceptible to β -sheet conformations. Mechanical analysis using nano-indentation via AFM revealed that two representative Library A samples, A261 and A10₂, both had a significantly higher Young's modulus compared to the representative Library B sample B10-22-17₂. We hypothesize that the presence of high β -sheet favored motifs in Library A combined with the highly elastic GPGXX motifs found combined with the non-canonical AAARAQAQAE motif in B10-22-17₂ could support that monomeric core units derived from the dragline and egg case silk genes can create stiffer and more deformation resistant biomaterials than monomeric core units derived from the adhesive and capture spiral silk. It was also observed that methanol induction increased the Young's modulus in the representative Library B sample while decreasing the Young's modulus in the Library A samples. One explanation is in

the presence of tightly packed and stabilized coiled-coiled interactions between α -helix regions in Library A being responsible for the high Young's modulus in uninduced Library A samples, while the increase in β -sheets after methanol induction could support the increased Young's modulus in Library B samples. The AFM results suggest that a varied array of biomaterials was achieved by the different libraries with Library A resulting in robust materials with a high Young's modulus and high stiffness and Library B containing biomaterials with a low Young's modulus and low stiffness and thus higher elasticity. To that end, future studies should focus on macroscopic qualities of fibers created from these recombinant proteins as well as biocompatibility in stem cell culture. With this work, we sought to create novel biomaterials that would mimic properties of native silk in vitro. From the observed trends of the novel recombinant proteins, the experiment led to successful creation of proteins with varied mechanical properties mimicking the diversity present in native spider silk. This suggests that materials could be further utilized to expand and enhance properties of spider silk proteins as future biomaterials.

6. Acknowledgements

I would like to thank Dr. Kaplan for his support and mentorship throughout my four years working in his lab. I would also like to thank my post-doctoral mentors Dr. Shun Zhou, Dr. Nina Dinjaski, Dr. Laura Brown and Dr. Zaira Moldes-Martin who have all provided mentorship and guidance for me throughout my research career. I would also like to thank Lauren Baugh for her help running AFM samples for us and for writing the AFM methods section for this paper. I also thank Dr. Jessica Garb for providing the cDNA sequences used in the formation of the libraries. This research was made possible through the support of NIH (NIH U01 EB014976) as well as the support of the Air Force Office of Scientific Research (AFOSR FA9550-14-1-0015).

7. References

1. Hu, X. *et al.* Araneoid egg case silk: a fibroin with novel ensemble repeat units from the black widow spider, *Latrodectus hesperus*. *Biochemistry* **44**, 10020–10027 (2005).
2. Hu, X., Wang, X., Rnjak, J., Weiss, A. S. & Kaplan, D. L. Biomaterials derived from silk–tropoelastin protein systems. *Biomaterials* **31**, 8121–8131 (2010).
3. Agnarsson, I., Kuntner, M. & Blackledge, T. A. Bioprospecting finds the toughest biological material: extraordinary silk from a giant riverine orb spider. *PLoS ONE* **5**, e11234 (2010).
4. Zhao, Y., Ayoub, N. A. & Hayashi, C. Y. Chromosome mapping of dragline silk genes in the genomes of widow spiders (Araneae, Theridiidae). *PLoS ONE* **5**, e12804 (2010).
5. Gnesa, E. *et al.* Conserved C-terminal domain of spider tubuliform spidroin 1 contributes to extensibility in synthetic fibers. *Biomacromolecules* **13**, 304–312 (2012).
6. Zhou, S. *et al.* Control of silicification by genetically engineered fusion proteins: Silk–silica binding peptides. *Acta Biomaterialia* **15**, 173–180 (2015).
7. Beckwitt, R., Arcidiacono, S. & Stote, R. Evolution of repetitive proteins: spider silks from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentarius* (Araneidae). *Insect Biochem. Mol. Biol.* **28**, 121–130 (1998).
8. Moon, M.-J., Townley, M. A. & Tillinghast, E. K. Fine structural analysis of secretory silk production in the black widow spider, *Latrodectus mactans*. *Korean Journal of Biological Sciences* **2**, 145–152 (1998).
9. Kong, J. & Yu, S. Fourier Transform Infrared Spectroscopic Analysis of Protein Secondary Structures. *Acta Biochimica et Biophysica Sinica* **39**, 549–559 (2007).
10. Florczak, A., Mackiewicz, A. & Dams-Kozłowska, H. Functionalized Spider Silk Spheres As Drug Carriers for Targeted Cancer Therapy. *Biomacromolecules* (2014). doi:10.1021/bm500591p
11. Wang, Q. *et al.* High Throughput Screening of Dynamic Silk-Elastin-Like Protein Biomaterials. *Advanced Functional Materials* **24**, 4303–4310 (2014).
12. Sutherland, T. D., Young, J. H., Weisman, S., Hayashi, C. Y. & Merritt, D. J. Insect Silk: One Name, Many Materials. *Annual Review of Entomology* **55**, 171–188 (2010).
13. Hinman, M. B. & Lewis, R. V. Isolation of a clone encoding a second dragline silk fibroin. *Nephila clavipes* dragline silk is a two-protein fiber. *J. Biol. Chem.* **267**, 19320–19324 (1992).
14. Vollrath, F. & Knight, D. P. Liquid crystalline spinning of spider silk. *Nature* **410**, 541–548 (2001).
15. Tucker, C. L. *et al.* Mechanical and physical properties of recombinant spider silk films using organic and aqueous solvents. *Biomacromolecules* **15**, 3158–3170 (2014).
16. Hayashi, C. Y., Blackledge, T. A. & Lewis, R. V. Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. *Mol. Biol. Evol.* **21**, 1950–1959 (2004).
17. Hayashi, C. Y. & Lewis, R. V. Molecular architecture and evolution of a modular spider silk protein gene. *Science* **287**, 1477–1479 (2000).
18. Omenetto, F. G. & Kaplan, D. L. New Opportunities for an Ancient Material. *Science* **329**, 528–531 (2010).
19. Perry, D. J., Bittencourt, D., Siltberg-Liberles, J., Rech, E. L. & Lewis, R. V. Piriform Spider Silk Sequences Reveal Unique Repetitive Elements. *Biomacromolecules* **11**, 3000–3006 (2010).
20. Marhabaie, M., Leeper, T. C. & Blackledge, T. A. Protein composition correlates with the

- mechanical properties of spider (*Argiope trifasciata*) dragline silk. *Biomacromolecules* **15**, 20–29 (2014).
21. Blasingame, E. *et al.* Pyriform Spidroin 1, a Novel Member of the Silk Gene Family That Anchors Dragline Silk Fibers in Attachment Discs of the Black Widow Spider, *Latrodectus hesperus*. *J Biol Chem* **284**, 29097–29108 (2009).
 22. Jenkins, J. E., Creager, M. S., Lewis, R. V., Holland, G. P. & Yarger, J. L. Quantitative Correlation between the Protein Primary Sequences and Secondary Structures in Spider Dragline Silks. *Biomacromolecules* **11**, 192–200 (2010).
 23. La Mattina, C. *et al.* Spider minor ampullate silk proteins are constituents of prey wrapping silk in the cob weaver *Latrodectus hesperus*. *Biochemistry* **47**, 4692–4700 (2008).
 24. Colgin, M. A. & Lewis, R. V. Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like ‘spacer regions’. *Protein Sci.* **7**, 667–672 (1998).
 25. Lewis, R. V. Spider silk: ancient ideas for new biomaterials. *Chem. Rev.* **106**, 3762–3774 (2006).
 26. Vollrath, F. & Porter, D. Spider silk as archetypal protein elastomer. *Soft Matter* **2**, 377–385 (2006).
 27. Kluge, J. A., Rabotyagova, O., Leisk, G. G. & Kaplan, D. L. Spider silks and their applications. *Trends Biotechnol.* **26**, 244–251 (2008).
 28. Scheibel, T. Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins. *Microb Cell Fact* **3**, 14 (2004).
 29. Xu, M. & Lewis, R. V. Structure of a protein superfiber: spider dragline silk. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7120–7124 (1990).
 30. Huang, J., Wong, C., George, A. & Kaplan, D. L. The effect of genetically engineered spider silk-dentin matrix protein 1 chimeric protein on hydroxyapatite nucleation. *Biomaterials* **28**, 2358–2367 (2007).
 31. An, B. *et al.* The influence of specific binding of collagen–silk chimeras to silk biomaterials on hMSC behavior. *Biomaterials* **34**, 402–412 (2013).
 32. Gosline, J. M., Guerette, P. A., Ortlepp, C. S. & Savage, K. N. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J. Exp. Biol.* **202**, 3295–3303 (1999).
 33. Xia, X.-X., Xu, Q., Hu, X., Qin, G. & Kaplan, D. L. Tunable Self-Assembly of Genetically Engineered Silk–Elastin-like Protein Polymers. *Biomacromolecules* **12**, 3844–3850 (2011).
 34. Adrianos, S. L. *et al.* *Nephila clavipes* Flagelliform silk-like GGX motifs contribute to extensibility and spacer motifs contribute to strength in synthetic spider silk fibers. *Biomacromolecules* **14**, 1751–60 (2013).
 35. Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A. & Hayashi, C. Y. Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One* **2**, e514 (2007).
 36. Lane, A. K., Hayashi, C. Y., Whitworth, G. B. & Ayoub, N. A. Complex gene expression in the dragline silk producing glands of the Western black widow (*Latrodectus hesperus*).
 37. Robinson, M. D. *et al.* Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881–2887 (2007).

38. Ravindra Babu, K., Moradian, A. & Douglas, D. J. The Methanol-Induced Conformational Transitions of ⁿ-Lactoglobulin, Cytochrome c, and Ubiquitin at Low pH: A Study by Electrospray Ionization Mass Spectrometry.
39. Tian, M. & Lewis, R. V. Tubuliform silk protein: A protein with unique molecular characteristics and mechanical properties in the spider silk fibroin family. *Appl. Phys. A* **82**, 265–273 (2006).
40. Tokareva, O., Jacobsen, M., Buehler, M., Wong, J. & Kaplan, D. L. Structure-function-property-design interplay in biopolymers: spider silk. *Acta Biomater.* **10**, 1612–26 (2014).
41. Hu, X. *et al.* Molecular mechanisms of spider silk. *Cell. Mol. Life Sci.* **63**, 1986–1999 (2006).
42. Eisoldt, L., Smith, A. & Scheibel, T. Decoding the secrets of spider silk. *Mater. Today* **14**, 80–86 (2011).
43. Vollrath, F., Knight, D. P., Plaza, G., Real, J. I. & Guinea, G. V. Liquid crystalline spinning of spider silk. *Nature* **410**, 541–548 (2001).
44. Guinea, G. V., Elices, M., Real, J. I., Gutiérrez, S. & Pérez-Rigueiro, J. Reproducibility of the tensile properties of spider (*Argiope trifasciata*) silk obtained by forced silking. *J. Exp. Zool. Part A Comp. Exp. Biol.* **303A**, 37–44 (2005).
45. Gomes, S. *et al.* AFM study of morphology and mechanical properties of a chimeric spider silk and bone sialoprotein protein for bone regeneration. *Biomacromolecules* **12**, 1675–85 (2011).
46. Krishnaji, S. T. *et al.* Thin Film Assembly of Spider Silk-like Block Copolymers. *Langmuir* **27**, 1000–1008 (2011).
47. Li, X., Shi, C.-H., Tang, C.-L., Cai, Y.-M. & Meng, Q. The correlation between the length of repetitive domain and mechanical properties of the recombinant flagelliform spidroin. *Biol. Open* bio.022665 (2017). doi:10.1242/bio.022665
48. Huang, W. *et al.* Synergistic Integration of Experimental and Simulation Approaches for the *de Novo* Design of Silk-Based Materials. *Acc. Chem. Res.* acs.accounts.6b00616 (2017). doi:10.1021/acs.accounts.6b00616
49. Thamm, C. & Scheibel, T. Recombinant production, characterization and fiber spinning of an engineered short major ampullate spidroin (MaSp1s). *Biomacromolecules* acs.biomac.7b00090 (2017). doi:10.1021/acs.biomac.7b00090
50. Wolff, J. O., Grawe, I., Wirth, M., Karstedt, A. & Gorb, S. N. Spider's super-glue: thread anchors are composite adhesives with synergistic hierarchical organization. *Soft Matter* **11**, 2394–2403 (2015).
51. Domke, J. & Radmacher, M. Measuring the Elastic Properties of Thin Polymer Films with the Atomic Force Microscope.
52. Rotsch, C., Jacobson, K., Radmacher, M. & Meyer, T. J. Dimensional and mechanical dynamics of active and stable edges in motile fibroblasts investigated by using atomic force microscopy. *Cell Biol.* **96**, 921–926 (1999).
53. Yamada, K., Tsuboi, Y. & Itaya, A. AFM observation of silk fibroin on mica substrates: morphologies reflecting the secondary structures. *Thin Solid Films* **440**, 208–216 (2003).

8. Figures

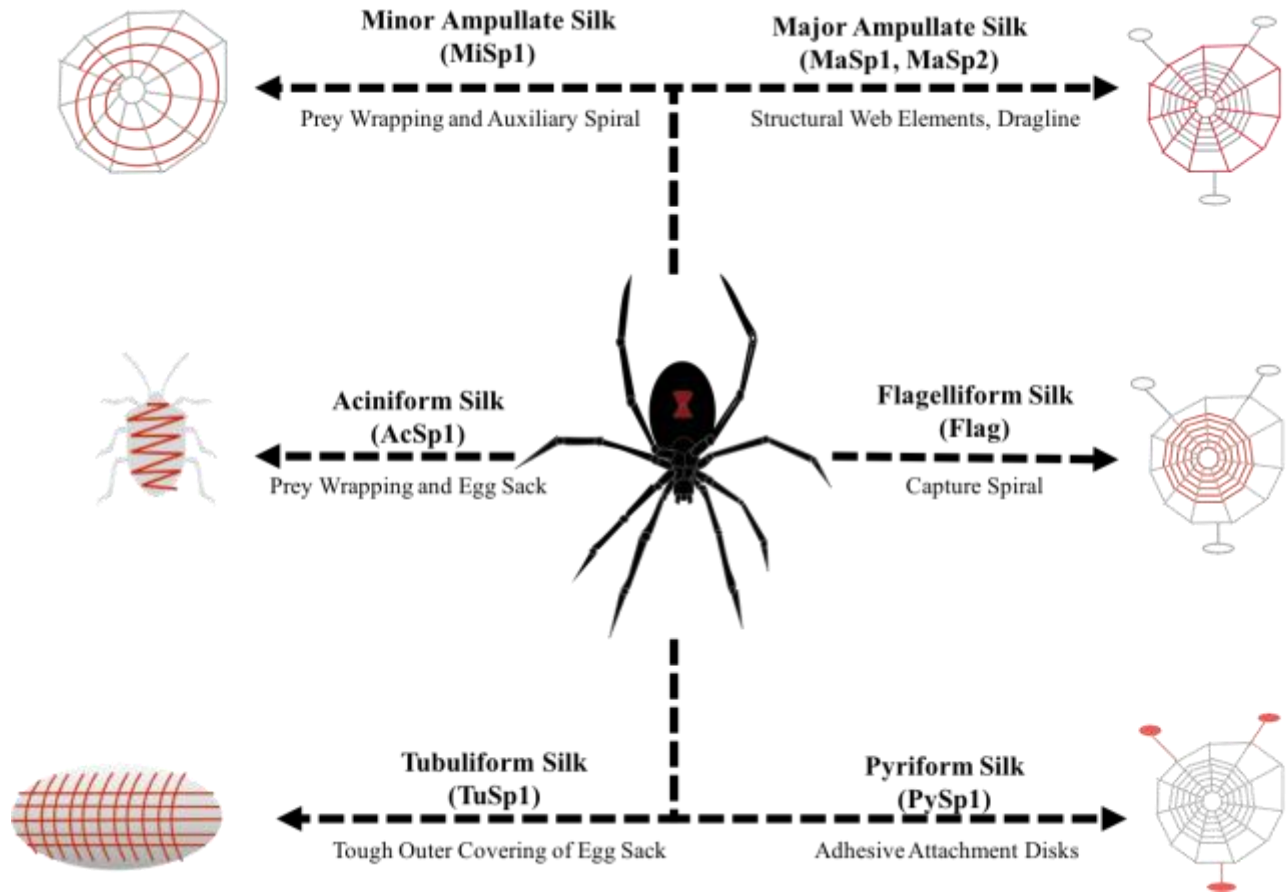


Figure 1. Summary figure of the seven different silks and their respective proteins produced by the female black widow spider (*L. hesperus*). The function of each of the silks shown in red are described (Adapted from Eisoldt et al., 2011).

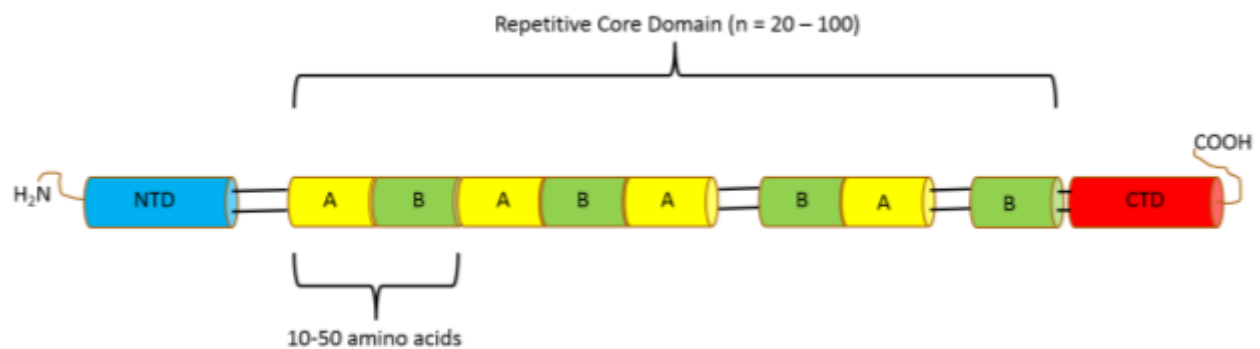


Figure 2. Schematic of Spider silk gene domains. Gene constructs used in the experiment are derived from the AB block monomers present in the repetitive core domain regions (Figure adapted from Heim et al. 2009)

Table 1. Summary Table of primary sequences of individual monomeric gene constructs used in Library

Module	Length (Amino Acids)	Sequence (N'→C')
MaSp1	35	GGAGQGGQGGYQGGYQGGAGQGGAGAAAAAAAA
MaSp2	40	GGSGPGGYQGPAAYGPSGPSGQQGYGPGGSGAAAAAAAA
TuSp1	184	ASQAASQSASSYSAAQSASFQASSSALASSSSFSSAFSSASSASAVGQVGYQIGLNA AQTLGISNAPAFADAVSQAVRTVGVGASPFQYANAVSNAFGQLLGGQGILTQENAAG LASSVSSAISSAASSVAAQAASAAQSSAFAQSQAAAQAFSQAASRSASQSAAQAGSSS TSTTTTSQA
MiSp1	42	GAGGYGQQGAGAGAGAGAGAGAGGYGQGSAGAAAGAAASAGA
AcSp1	188	FGLAIAQVLGTSGQVNDANVNQIGAKLATGILRGSSAVAPRLGIDLSGINVDSDIGSVT SLILSGSTLQMTIPAGGDDLSGGYPGGFPAGAQPSSGAPVDFGGPSAGGDVAAKLAR SLASTLASSGVFRAAFNSRVSTPVAVQLTDALVQKIASNLGLDYATASKLRKASQAVSKV RMGSDTNAYALA
PySp1	40	ARAQAQAEAAARAQAQAEAAARAQAQAEAAARAQAQAEAA
Flag	40	GPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGAGPGGA

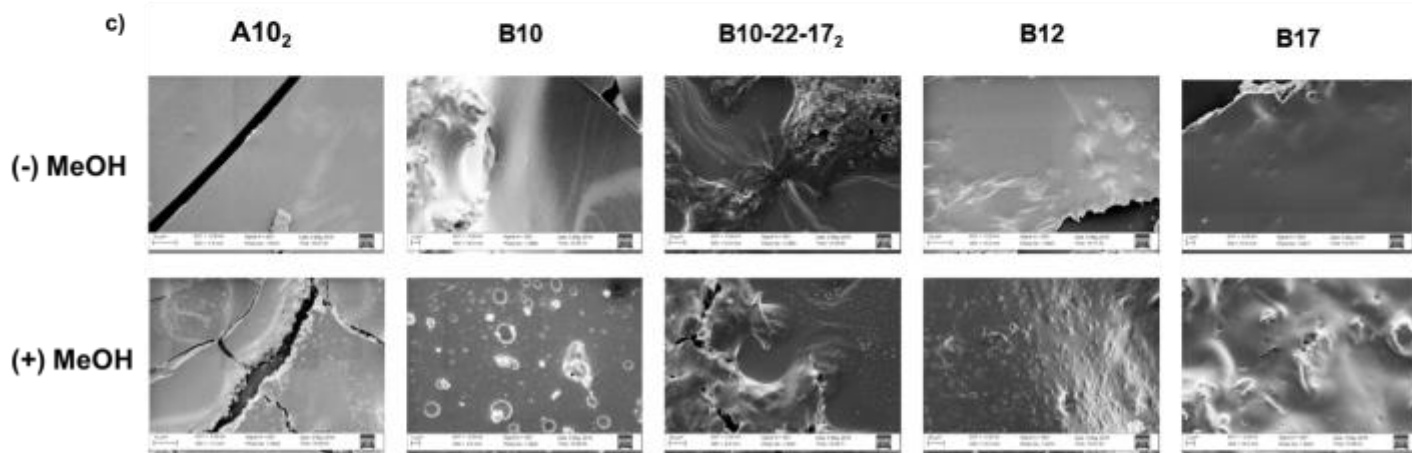
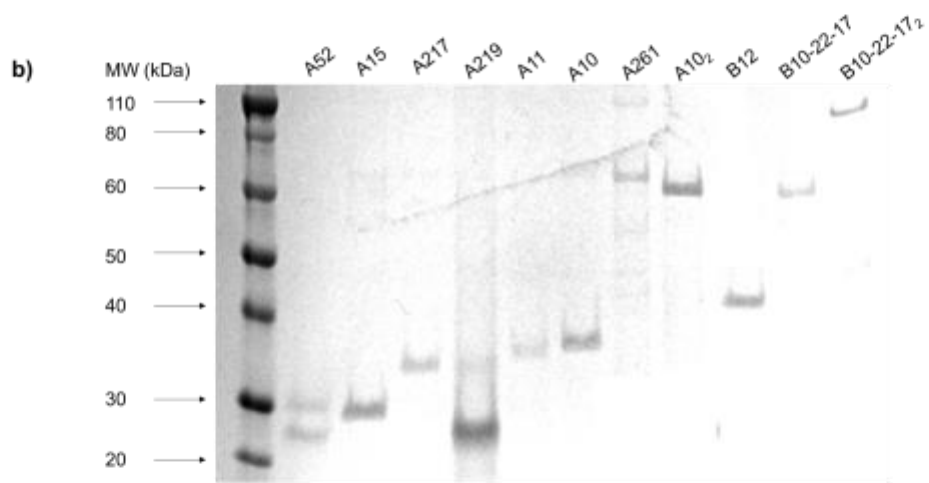
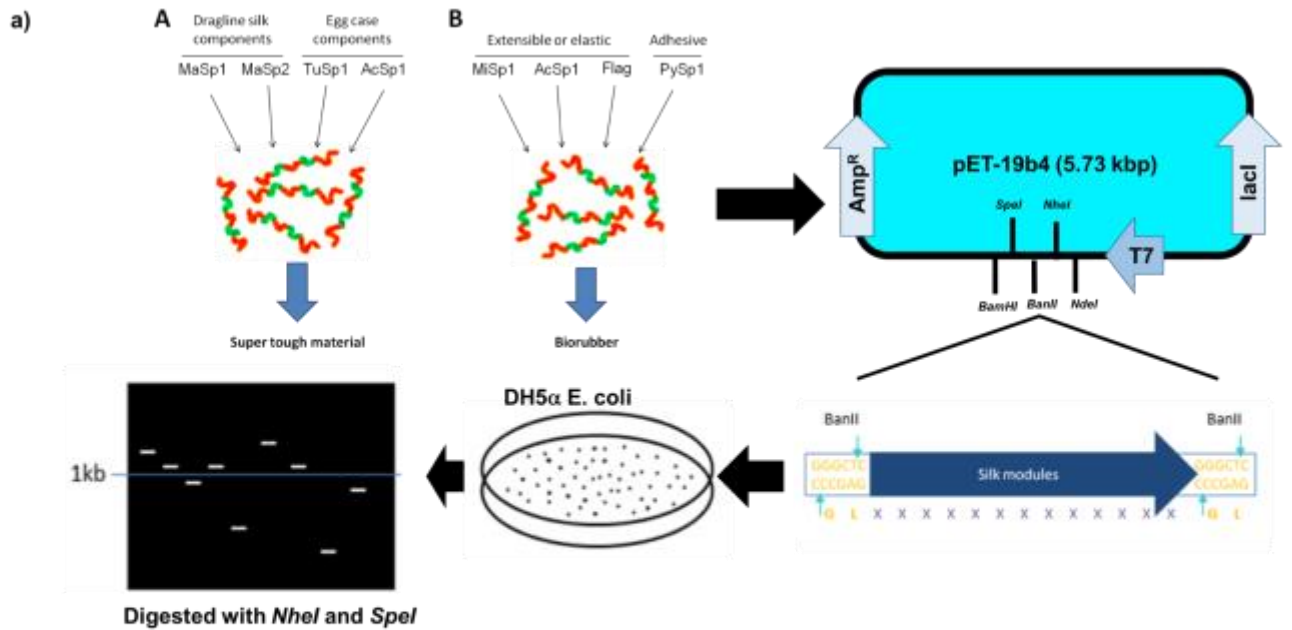
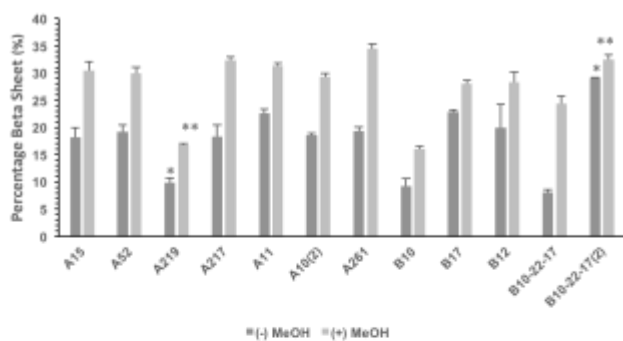


Figure 3. Establishment of the recombinant DNA libraries. (a) Generalized BanII restriction enzyme mediated cloning strategy for establishment of DNA library. (b) SDS-Page of select purified library constructs (A52, A15, A217, A219, A11, A10, A261, A10₂, B12, B10-22-17, B10-22-17₂) (c) Exemplar scanning electron microscope images of select HFIP based protein films (A10₂, B10, B10-22-17₂, B12, B17) with control and methanol treatment.

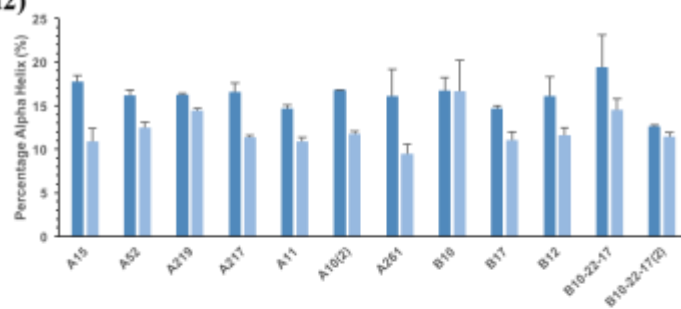
Table 2. Summary Table of relevant library colonies obtained from screening with molecular weights determined by primary sequence data and corresponding gene sequence.

Library colony	Estimated MW (kDa)	Sequence
A10	28.8	MaSp1-MaSp2-MaSp1-AcSp1
A11	29.3	MaSp2-MaSp1-MaSp2-AcSp1
A15	21.3	MaSp1-TuSp1
A46	21.9	MaSp2-TuSp1
A52	23.0	MaSp2-AcSp1
A217	27.4	(MaSp1) ₃ -AcSp1
A219	22.0	(MaSp1) ₄ -(MaSp2) ₃
A261	56.0	(TuSp1) ₂ -AcSp1-(MaSp1) ₂
A10₂	56.7	(MaSp1-MaSp2-MaSp1-AcSp1) ₂
B10	15.2	(PySp1) ₃ - Flag
B12	32.7	PySp1-MiSp1-MaSp1-Flag-AcSp1
B17	27.6	(PySp1) ₂ -AcSp1
B22	11.5	PySp1-MiSp1-PySp1
B10-22-B17	55.3	(PySp1) ₃ -Flag-PySp1-MiSp1-PySp1 ₃ -AcSp1
B10-22-B17₂	81.4	(PySp1) ₃ -Flag-PySp1-MiSp1-PySp1-(PySp1 ₂ -AcSp1) ₂

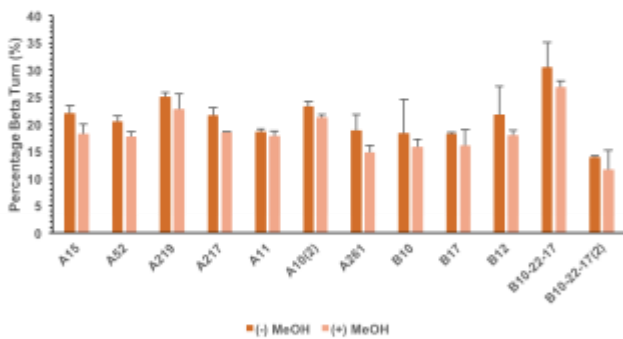
a1)



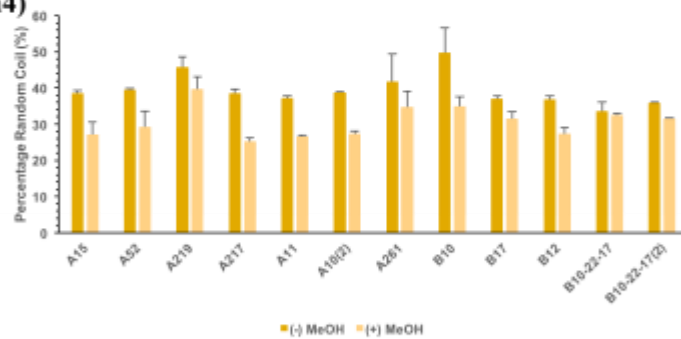
a2)



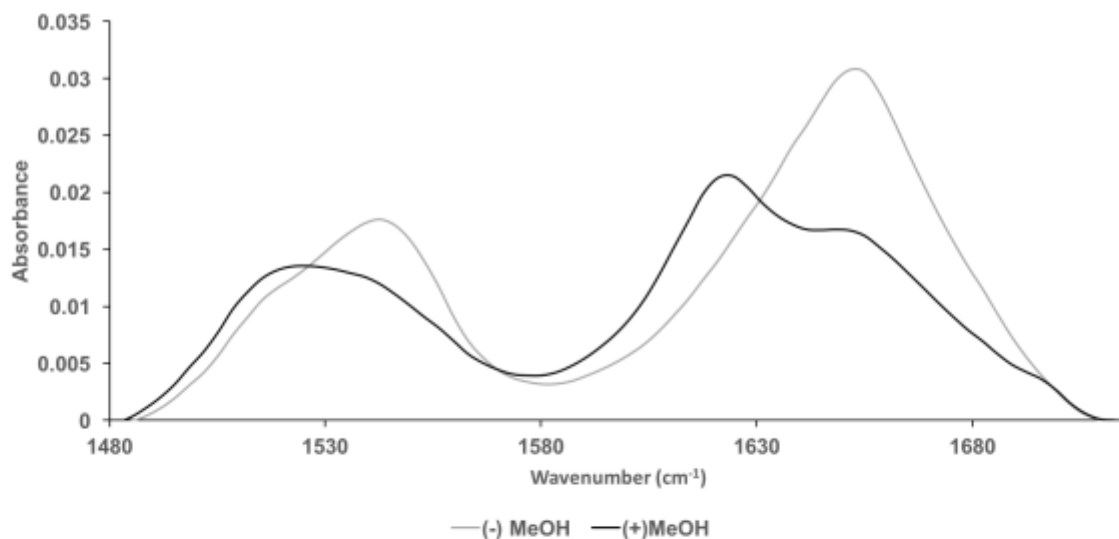
a3)



a4)



b)



c)

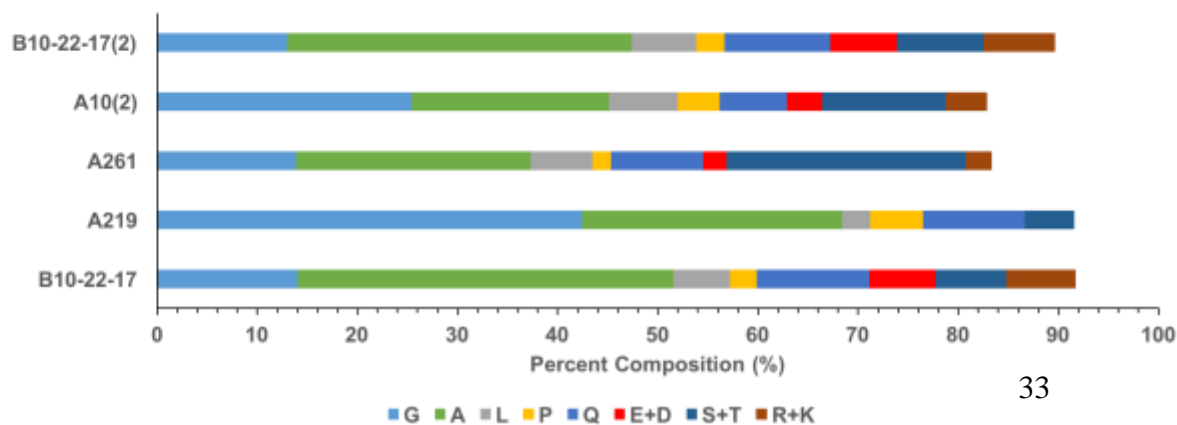


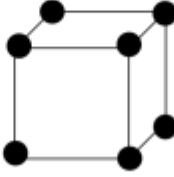
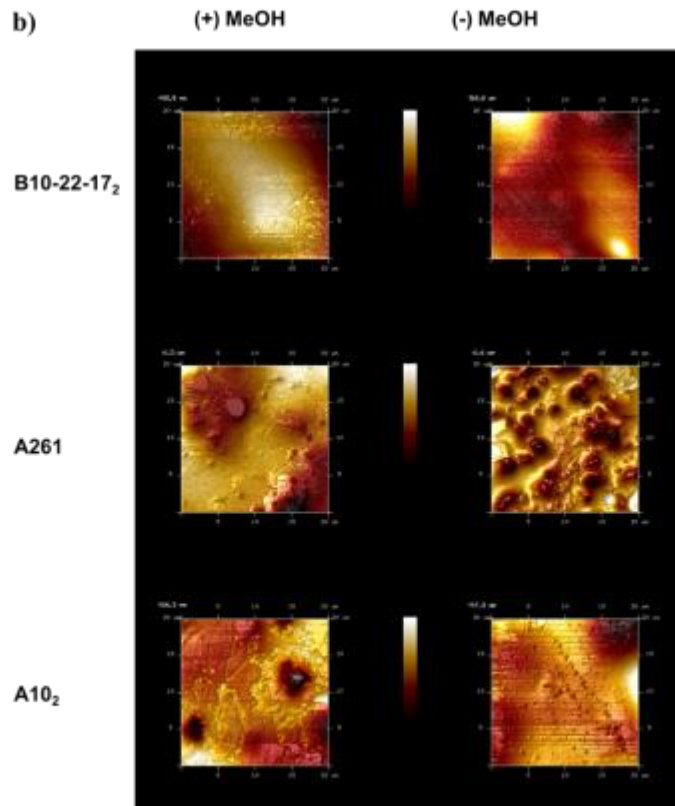
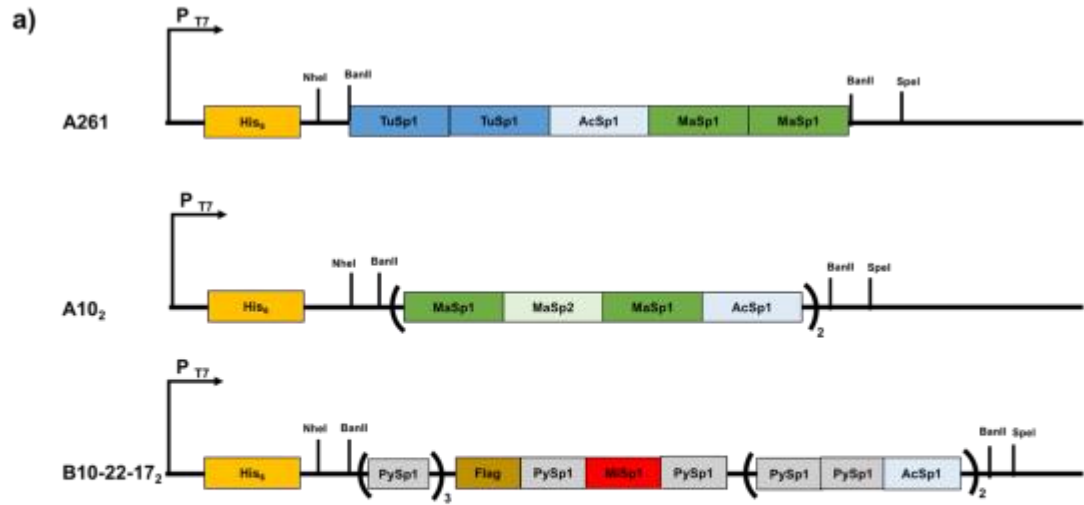


Figure 4. Amino acid composition and secondary structures of recombinant protein films. (a) Summary graph of percent (a1) β -sheet (gray and black), (a2) α -helix (dark blue and light blue), a3) β -turns (dark orange and light orange), (a4) Random coil (dark yellow and light yellow) for recombinant films. Percentage of the peak present at wavenumbers corresponding to a particular secondary structure across the amide I absorption band from 1720-8.1) 1580 cm^{-1} (β -turns 1668-1696 cm^{-1} , α -helix 1658-1667 cm^{-1} , random coil 1630-1657 cm^{-1} and β -sheet 1618-1629 cm^{-1} and 1697-1703 cm^{-1}) relative to the total peak area was used to determine percentage secondary structure composition. with mean and standard deviation shown (n=3 per sample). (*, **) indicate statistical significance when compared to recombinants within the same library for both methanol treated and untreated samples. Significance was determined between recombinants within the same library via one way ANOVA followed by post hoc Tukey-Kramer test and within the same sample by a paired t-test ($p > 0.05$). (b) Representative FTIR spectra for recombinant A10 across amide I and amide II absorption spectra with methanol induced (black) and uninduced (gray) films. (c) Percentage amino acid composition of five recombinant proteins (B10-22-17₂, A10₂, A261, A219, B10-22-17) with emphasis on glycine (light blue), alanine (green), leucine (gray), proline (yellow), glutamine (navy blue), glutamic acid and aspartic acid (red), serine and threonine (dark blue), arginine and lysine (brown). Amino acid composition calculated from translated primary sequences derived from sequenced DNA.

Table 3. Summary table of secondary structural motif counts as determined from primary protein sequence for five select recombinant protein films (B10-22-17₂, A10₂, A261, A219, B10-22-17) along with their percentage secondary structure after methanol induction as determined from FTIR analysis.

Sample	Sequence	GPGXX Motifs	GGX Motifs (X = Y, L, Q)	A _n /GA _n Motifs	AAQAA /AASQSA Motifs	Percentage Beta Sheet (%)	Percentage Beta Turn (%)	Percentage Alpha Helix (%)
B10-22-17 ₂	[PySp1 ₃ -Flag-PySp1-MiSp1-PySp1-(PySp1 ₃ AcSp1 ₂) ₂]	8	4	0/4	0/0	32.5	11.6	15.5
A10 ₂	[MaSp1-MaSp2-MaSp1-AcSp1 ₂]	4	16	6/4	0/0	29.3	21.2	11.7
A261	[[TuSp1 ₂ -AcSp1-(MaSp1 ₂) ₂]	0	9	2/2	2/4	34.4	14.9	9.5
A219	[[MaSp1 ₂ -(MaSp1 ₃) ₂]	6	11	7/4	0/0	16.7	22.8	14.4
B10-22-17	[PySp1 ₃ -Flag-PySp1-MiSp1-PySp1 ₃ -AcSp1]	8	3	0/4	0/0	24.3	26.9	14.5
Structural Role								
		Elastic β-spiral	Amorphous 3_{10} helix	Crystalline β-sheet				



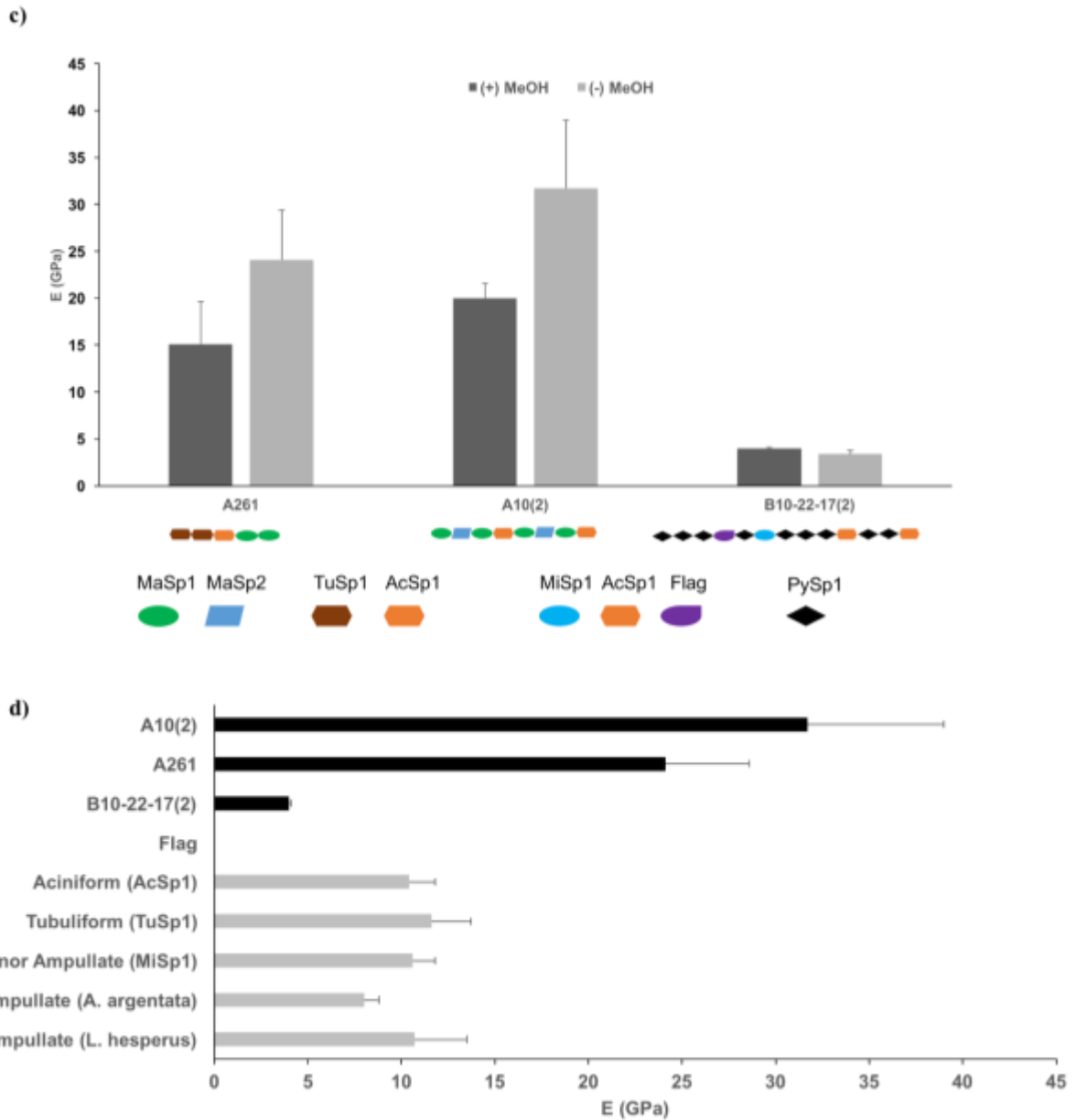


Figure 5. Mechanical characterization of selected protein films. (a) gene sequences with included restriction enzymes of the three Library colonies selected (A261, A10₂, B10-22-17₂). (b) Representative AFM topographic images of HFIP based protein films with and without methanol induction. (c) Summary graph of Young's modulus (E) for the three recombinant proteins. Bars in black are methanol induced samples while bars in gray are uninduced samples (mean and standard deviation shown). (d) summary graph showing Young's modulus for uninduced Library A samples and methanol induced library B samples along with Young's modulus for selected native proteins. Young's moduli for native proteins derived from Instron tensile testing measurements of full length fibers from Blackedge and Hayashi 2006. Mean and standard deviation shown.