Biomimetic manufacturing of fibers

F. Teulé¹, C. Aubé¹, M. Ellison² & A. Abbott¹
¹Dept. of Genetics and Biochemistry, Clemson University, USA
²School of Materials Science and Engineering, Clemson University, USA

Abstract

Engineering any fiber protein requires understanding of their structure function relationships. We have chosen natural protein fibers such as spider silk and collagen as models to investigate the role that various proteins primary structural components play in fiber production. Spider dragline silk is essentially composed of two highly repetitive proteins called spidroins (1 and 2) that are made of alternating amorphous glycine-rich amino acid repeats and crystalline alanine-rich motifs.

We have focused our research on testing the role of alanine motifs found in these silk proteins in the mechanical properties of the resulting fiber. Three synthetic spidroin 1-like genes were engineered to determine the role of such alanine-rich motifs in spidroin 1 proteins. Each of these constructs was encoded for a synthetic spidroin 1 containing variable amounts of alanine motifs (normal, low, and none). We have also engineered three collagen-spidroin 1 copolymer genes for each of the synthetic spidroin 1 genes. Such collagen-silk copolymers are found in natural fiber proteins composing the byssus thread of marine mussels. The designed synthetic genes were introduced into yeast (Pichia pastoris) for protein production and characterization and blend fibers have been spun using electrospinning technologies.

Keywords: spider silk, spidroin, collagen, gene engineering, pichia pastoris, fiber proteins.

1 Introduction

Nature offers a variety of environmentally friendly high performance biomaterials that are produced by several organisms to achieve specific tasks. One of the most amazing is the dragline silk made in the major ampullate gland
of the golden orb-weaver spider named *Nephila clavipes*. This natural protein fiber is mostly composed of two highly repetitive proteins called spidroin 1 and 2 or MaSp1 and MaSp2 (Xu and Lewis [1] and Hinman and Lewis [2]). The primary structures of these spidroins consist of alanine motifs sandwiched between glycine-rich (MaSp1) or proline containing glycine-rich regions (MaSp2). As a high performance fiber, *Nephila clavipes* dragline silk combines high strength and high elasticity resulting in an exceptionally tough material whose mechanical properties have yet to be matched by man-made materials (Gosline et al. [3]). These mechanical and physical properties are believed to lie in the primary structure of these spidroin components. They also depend on how the liquid crystalline phase in which these proteins are found in the silk gland lumen is processed by the spider's spinning apparatus. Physical data (mostly X-ray) on dragline silk fibers and partial knowledge of the spidroin protein sequences allowed the formulation of hypotheses regarding the secondary structures, as well as the role of alanine or glycine rich regions in the fiber self-assembly processes (Hayashi et al. [4]). One model suggests that the alanine regions adopt anti parallel β-sheet structures leading to the formation of crystals arranged parallel to the fiber axis that would be responsible for the fiber strength. Thus, spider silk is an excellent model to investigate the relationships between the structure of the individual repetitive motifs present in the fiber proteins and their mechanical function in the resulting fiber. Moreover, the repetitive nature of these proteins is suitable for genetic engineering of novel custom designed protein based biomaterials.

Efforts to artificially produce spider silk proteins using genetically engineered bacteria or yeast (see Heslot [5] for review), or plants (Scheller et al. [6]) or mammalian systems (Lazaris et al. [7]) have been reported. Although in all cases the recombinant silk proteins were successfully produced, only one group succeeded to spin the silk-like proteins into fibers (Lazaris et al. [7]). However, the mechanical properties of the artificially spun fibers did not match those of the natural dragline silk fiber indicating that the handling of the proteins, in order to control the fiber assembly process, as well as the spinning process were also critical steps in the making of the fiber.

Using *N. clavipes*’ major ampullate silk cDNA sequences (MaSp1) as a model, we wanted to determine the secondary structure adopted by the alanine motifs present in the transgenically produced silk-like proteins and their putative role in the formation of crystals that would account for the fiber’s strength. For fiber protein production, we are employing simple systems such as yeast (*Pichia pastoris*) and higher organisms such as plants as bioreactors. We have generated several yeast clones able to produce three customized spidroin 1-like proteins containing variable amounts of alanine motifs (regular, less and none). Moreover, we have also generated yeast clones producing copolymeric collagen-silk proteins inspired from the natural fibrous proteins forming the mussel byssal threads. These collagen-spidroin 1 copolymer gene constructs were engineered using sequences encoding the helical part (GXY)	extsubscript{n} of a root-knot nematode cuticle collagen (*Meloidogyne incognita*) (Wang et al. [8]) that were juxtaposed to each of our customized synthetic spidroin 1 genes.
2 Materials and methods

2.1 Gene engineering and cloning in yeast

2.1.1 Synthetic spidroin 1-like genes

2.1.1.1 Spidroin 1-like homopolymers (Figure 1) Sets of 5’ phosphorylated complementary oligonucleotides (IDT technology) spanning the region encoding the consensus repeat of *N. clavipes* spidroin 1 protein (Masp1) were used to build three types of spidroin 1-like genes encoding for spidroin 1-like proteins containing normal amounts of alanine motifs, lesser and no alanine motifs (Figure 1). Each complementary set of oligonucleotides was annealed at 65°C for 5 hours then slowly cooled to room temperature. Appropriate annealed reactions (Figure 1) were mixed and after ethanol precipitation, the recovered DNA fragments were ligated in a head to tail fashion overnight at 4°C using T4 DNA ligase (Promega). The resulting multimer sequences were electrophoretically separated on 0.8% agarose gels and the fragments between 500bp and 1000bp were isolated by electroelution followed by ethanol precipitation (Sambrook [9]). Each type of size-selected spidroin 1-like sequence was "filled" and "A-tailed" at 72°C for 1 hour in individual reactions using Taq DNA polymerase (Fisher) and dNTPs (Promega). The tailed products recovered by ethanol precipitation were ligated to a pGEMT-Easy vector (Promega) and the ligation reaction was used to transform competent *E. coli* strain JM109 (Promega). Clones containing plasmids with inserts between 500 and 600bp were selected and the plasmids DNAs were isolated using an alkaline lysis method (Birnboim [10]). The selected clones were sequenced using an ABI PRISM™ Dye terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer) and M13 (forward and reverse) universal primers. The sequences were resolved on an ABI 373 Stretch sequencer (Applied Biosystems Inc.). Three clones, each containing one of the spidroin 1-like candidate multimer sequences (normal, less or no alanine motif encoding sequences) ranging between 500 and 600bp, were selected.

2.1.1.2 Customization of the engineered sequences by PCR reaction All the PCR reactions were performed using a high fidelity DNA polymerase (Pfu DNA polymerase, Promega) and specific primers designed to add appropriate restrictions sites in the 5' and 3' of the genes to allow the cloning in the yeast expression vector.

For the spidroin 1-like homopolymer sequence amplification, we used specific primers designed to add an Eco RI and a Sac II restriction sites in the 5’ and 3’ of the three types of sequences (normal, less and no alanine motif encoding genes).

For the collagen-spidroin 1-like copolymer sequence engineering, we performed a PCR reaction on the *mi-col-2* collagen cDNA clone (Wang et al. [8]) using specific primers to amplify the collagen-like sequence we needed. The primers were specific to the region encoding for the helical part of this collagen and were designed to add an Eco RI and a Bam HI sites in the 5’ and 3’ of the 426...
bp amplified product respectively. In this case, the three types of spidroin 1-like sequences (normal, less and no alanine motif encoding genes) were individually customized by PCR amplification using specific primers that allowed the addition of a *Bam* HI and a *Sac* II sites in the 5' and the 3' end of the gene respectively.

![Spidroin 1-like multimer construction](image)

**Figure 1:** Spidroin 1-like multimer construction.

A. Major ampullate spidroin 1 amino acid consensus repeat [4]; B. Sets of oligonucleotides spanning the region encoding the spidroin 1 consensus repeat (sets B1, B2 and B3) that were used to engineer sequences encoding spidroin 1 proteins with regular, less and no alanine motifs respectively; C. Annealing of each complementary oligonucleotides; D. Generation of spidroin 1-like sequences by head to tail ligation (multimerization). D1, D2 and D3 are sequences encoding for spidroin 1-like proteins containing the regular amount of alanine motifs, a lesser amount and no alanine motifs respectively.

All the spidroin 1-like and collagen PCR products were first cloned in pGEMT-Easy (Promega) in *E. coli* strain JM109 (Promega) and the inserts of the candidate recombinant plasmids were sequenced. The two sets of spidroin 1-like clones and the collagen clone containing the correct sequence flanked by the appropriate 5' and 3' restrictions sites were identified and selected.

### 2.1.2 Cloning in *Pichia pastoris*

All the customized homopolymer and copolymer sequences were cloned in the yeast expression vector pPICZαA (Invitrogen) at the *Eco* RI/Sac II sites. This vector contains features allowing the addition of a histidine tag for easy purification of the secreted recombinant proteins.

For spidroin 1 homopolymer gene construction, the pGEMT-Easy recombinant plasmids containing the customized spidroin 1-like sequences were subjected to an *Eco* RI/Sac II restriction digestion while the set of cloned customized spidroin 1-like sequences used for the copolymer construction were subjected to a *Bam* HI/Sac II restriction digestion. The pGEMT-Easy recombinant clone containing the customized collagen-like sequences was subjected to an *Eco* RI/Bam HI restriction digestion.

For the homopolymer gene construction, the pPICZαA vector was ligated to each of the *Eco* RI/Sac II types of spidroin 1 sequence using a 1:3 vector: insert ratio.
For the copolymer construction, three separate ligation reactions were set up and each one contained a 1:3:3 molar end ratio of the prepared Eco RI/Sac II pPICZαA vector, the Eco RI/Bam HI collagen gene and one of the three Bam HI/Sac II spidroin 1-like sequences respectively.

Each ligation reaction was used to transform E. coli TOP10F' cells. Candidate plasmids containing homopolymer or copolymer genes of the correct size were identified. These six plasmid clones were individually isolated and sequenced using vector specific primers: 5’AOX1, α Factor, and 3’AOX1 primers (Invitrogen).

2.2 Pichia pastoris transformation and expression studies

Each of the three homopolymer or the copolymer constructs were individually introduced in Pichia pastoris strain GS115 (Invitrogen). Transformations and expression studies (using BMGH/BMMH media) were performed as suggested by the manufacturer (Easy select™ Pichia Expression Kit, Invitrogen). Yeast transformants were obtained by plating the transformation reaction on a media containing five times the suggested antibiotic (zeocin) concentration. Clones showing higher protein production levels were characterized as "super expressers" and were selected for our study.

2.3 Protein extraction and purification

All protein extractions from yeast cells were performed following the instructions specified by the manufacturer (Easy select™ Pichia Expression Kit, Invitrogen). Proteins were purified by affinity chromatography (Talon Metal Affinity Resins, Clontech) or by precipitation with 30% ammonium sulphate.

2.4 Western blot and Dot blot analysis

All intracellular or secreted recombinant proteins were subjected to SDS-PAGE analysis of 4%/10% polyacrylamide gels followed by Coomassie blue staining or Western-blot analysis using antibodies directed against a myc epitope present in the fusion protein (ECL™ Western Blotting Analysis Systems, Amersham Pharmacia Biotech). Dot blot analyses of crude cellular protein extracts or correspondent culture medium were also used as a method to quickly screen and select yeast transformants producing the recombinant proteins. Each time, Western-blot analysis was used to confirm the presence, the size and the production level of each engineered protein.

3 Results and discussion

3.1 Spidroin 1 homopolymer and collagen-spidroin 1 copolymer engineering in Pichia pastoris

We have successfully generated spidroin 1-like sequences encoding for proteins containing a variable amount of alanine motifs (normal= ALA, less= LA, and
none= NA) using synthetic oligonucleotides that designed to span the region encoding for the amino acid consensus repeat of the major ampullate spidroin 1 protein (Figure 1). We used a head to tail ligation strategy of synthetic variants of a spidroin 1-like monomer repeat to reproduce the repetitive nature of the spidroin 1 gene sequence. Although this multimerization process could generate spidroin 1-like sequences of sizes up to 6kbp, we deliberately selected spidroin 1-like sequences ranging between 500 bp and 1000 bp for several reasons: 1) To confirm the sequence of the clones, 2) To create individual cassettes containing sequences encoding for different fiber protein amino acid motifs that could later be used in different combinations for fiber protein engineering. We wanted to determine the role and secondary structure of the alanine motifs present in spidroins produced "in vivo". To this end, we have engineered yeast clones containing three types of spidroin 1-like sequence each able to produce spidroin 1 like protein variants homogenous in size that would only differ by the number of alanine motifs present in their sequence. In our case, we were also interested to test the ability of a simple eukaryote such as yeast to manufacture, handle and secrete such fiber proteins. We first cloned the three types of spidroin 1-like multimer sequences ranging between 500 and 600bp in E. coli for sequence verification before introducing them in Pichia pastoris for production of the three types of homopolymer proteins (spidroin 1-ALA, -LA, or -NA).

We also generated collagen/spidroin1-like copolymer sequences (COALA, COLA, and CONA) juxtaposing a customized collagen sequence encoding the helical part of a cuticle collagen from a parasitic nematode in combination to each of our customized synthetic spidroin 1-like sequences.

3.2 Yeast expression studies

After transformation in yeast and selection on media containing higher antibiotic levels, we screened up to 50 independent yeast transformants for each of the three homopolymer or copolymer clones looking at their recombinant protein production. In each case, yeast cellular extracts and culture media were analyzed using a Dot blot method to first check for protein production and then Western blot to confirm the identity of the expected recombinant proteins by molecular weight. Both of these methods were also useful to compare the level of protein production and secretion between different clones within a category of homopolymer or copolymer clones. The expected molecular weights of the homopolymer proteins were 16.75 kDa, 19.9 kDa and 18.5 kDa for SPIDROIN 1-ALA (ALA), SPIDROIN 1-LA (LA) and SPIDROIN 1-NA (NA) respectively. The expected molecular weight for the copolymers were 28.9 kDa, 29.8 kDa, and 30.6 kDa for the COLLAGEN/SPIDROIN 1-ALA (COALA), COLLAGEN/SPIDROIN 1-LA (COLA), and COLLAGEN/SPIDROIN 1-NA (CONA) respectively.

We identified and selected six yeast clones each producing the spidroin 1-ALA, LA, or NA proteins and the copolymers collagen spidroin 1-like COALA, COLA, or CONA at higher levels ("super expresser") (Figure 2).
Panel A, Lanes 1 to 3; Molecular weight marker, β-Galactosidase, 119 KDa (positive control for intracellular expression), protein extract of wild P. pastoris strain GS115 (negative control) respectively; Lanes 4 and 5: spidroin 1-LA homopolymer and copolymer respectively. Panel B, Lanes 1 and 2: spidroin 1-NA homopolymer and copolymer respectively. Panel C, Lanes 1 and 2: Spidroin 1-ALA homopolymer and copolymer respectively. The predominant homopolymer proteins detected correspond to dimer forms and their sizes are 39.4, 36.6 and 33.1KDa for spidroin 1-LA, -NA and -ALA respectively (fine arrows panels A, B and C). The copolymer proteins detected also correspond to dimer forms of the collagen-spidroin 1-like proteins and their sizes are 59.2, 60.9, and 57.4 KDa for COLA, CONA, and COALA respectively (bold arrows in panels A, B and C).

Panels A and B show the Dot blot analyses of the intracellular protein fraction and culture media (secreted fraction) respectively. Panel C shows the Western blot analysis of the secreted proteins (same samples as in B; Lanes 1 to 11: T0 to T10). Sampling of both cells and media (without cells) were collected from the cultures at regular time intervals after the initial start of induction (T0= start of induction, T1 to T10 are collections from = 4 hours and up to 88 hours after induction).

We carried on the expression studies for each of the selected yeast clones were to determine the optimal time of protein production and secretion for each of the homopolymers (ALA, LA, and NA) and copolymers (COALA, COLA, COALA).
and CONA). We found that all the categories of spidroin 1-like proteins and corresponding copolymers are being produced intracellularly by the selected yeast clones. However, we found that only the COALA and COLA secreted proteins were detected by both Dot blot and Western blot analysis for the (Figure 3).

For the homopolymer protein production, Dot blot analyses suggested that the homopolymer proteins ALA and LA were not only being highly produced in the cell and that they were being secreted. However Western blot analysis confirmed these results for ALA only. In comparison to the other homopolymers, the NA proteins are briefly produced intracellularly and are not detected in the culture media by none of the techniques used. Moreover the very little amount of NA proteins detected in the cells suggests a stability issue regarding the spidroin 1-NA protein or transcript probably inherent to its sequence. The major forms of the homopolymer and copolymer proteins detected in the cell are dimers. Since Southern blot analysis showed the presence of the engineered homopolymer and copolymer genes as monomeric forms, the protein dimers detected by Western-blot analysis are not the product of an eventual gene recombination that would have led to a doubling of the gene size (data not shown). Thus, these results suggest that these proteins have a very high tendency to aggregate and we first thought that they might be crosslinked since the dimer forms were stable through treatment with SDS and heat. However additional analysis of cellular extracts using urea-PAGE technique (data not shown) suggested that stronger chaetropic agents could disrupt these aggregates thus showing the absence of physical crosslinks in these homopolymers. Furthermore, these results indicated that these proteins might aggregate due to the presence of the glycine rich region since only two of them (ALA and LA) contained extra alanine motifs while NA did not and still could aggregate. We also have to take into account that the solubility of these proteins in the cell extract buffer and the culture media might be different due to the difference in salt concentration and pH conditions of the two buffers. Thus the secreted proteins may precipitate out of the culture media preventing them from entering the polyacrylamide gel while the intracellular proteins are maintained in solution. This would provide a possible explanation to why Western blot analysis on secreted spidroin 1-like proteins does not give any evidence (or very little for ALA) of the presence of the protein in the culture media as expected although Dot blot analysis was capable of detecting their presence. The results obtained for the copolymer proteins suggest that the addition of a collagen sequence seems to enhance the overall solubility of these proteins. The trend is such that the proteins containing no alanine motifs (homopolymer spidroin 1-NA and copolymer CONA) seem to be less stable than the ones that do contain these motifs (ALA or LA homopolymers and copolymers) suggesting a “stabilizer” role for the alanine motifs.

We are currently perfecting the purification methods for all these spidroin 1-like homopolymer and copolymer proteins using metal ion affinity chromatography in combination with heat treatment, salt concentration and pH modifications to be more adequate for large scale production. We have also spun
fiber blends of purified COALA or COLA proteins using electro spinning methods and are characterizing the properties of the resultant materials.

Acknowledgements

We would like to thank the National Textile Center for the funding of this research and Dr R. S. Hussey (University of Athens, GA) for generously providing the mi-col-2 cDNA clone.

References