Synthetic Spider Silk Fibers Spun from Pyriform Spidroin 2, A Glue Silk Protein Discovered in Orb-Weaving Spider Attachment Discs†

Paul Geurts,‡ Liang Zhao,§ Yang Hsia,‡ Eric Gnesa,‡ Simon Tang,|| Felicia Jeffery,‡ Coby La Mattina,‡ Andreas Franz,§ Leah Larkin,‡ and Craig Vierra*,‡

Departments of Biological Sciences, Chemistry, and Mechanical Engineering, University of the Pacific, Stockton, California 95211, United States

Received August 25, 2010; Revised Manuscript Received October 10, 2010

Spider attachment disc silk fibers are spun into a viscous liquid that rapidly solidifies, gluing dragline silk fibers to substrates for locomotion or web construction. Here we report the identification and artificial spinning of a novel attachment disc glue silk fibroin, Pyriform Spidroin 2 (PySp2), from the golden orb weaver *Nephila clavipes*. MS studies support PySp2 is a constituent of the pyriform gland that is spun into attachment discs. Analysis of the PySp2 protein architecture reveals sequence divergence relative to the other silk family members, including the cob weaver glue silk fibroin PySp1. PySp2 contains internal block repeats that consist of two subrepeat units: one dominated by Ser, Gln, and Ala and the other Pro-rich. Artificial spinning of recombinant PySp2 truncations shows that the Ser-Gln-Ala-rich subrepeat is sufficient for the assembly of polymeric subunits and subsequent fiber formation. These studies support that both orb- and cob-weaving spiders have evolved highly polar block-repeat sequences with the ability to self-assemble into fibers, suggesting a strategy to allow fiber fabrication in the liquid environment of the attachment discs.

Introduction

Spiders are capable of spinning high performance silk fibers that have extraordinary properties. Over 37000 different spider species have been identified and more than half of these spiders spin webs designed to capture prey.[1] Orb-weaving spiders are capable of spinning at least six different fiber types with a broad range of molecular properties, which have evolved to serve specialized functions including locomotion, prey swathing, web construction, and protection of eggs.[2,3] Biochemical studies to understand the relationship between the proteins that make up spider silk and their contribution to the mechanical properties of the fibers have revealed a diverse range of molecular modules within the protein sequences that are important for spider survival.[4,5] One characteristic feature of spider silk proteins is the presence of block amino acid repeats (called ensemble repeats or repetitive regions) that reoccur many times within the interior of the silk protein. Typically, silk proteins only contain one type of ensemble repeat within their amino acid sequence. Within each ensemble repeat (e.g., 200 amino acids) there are often shorter peptide motifs. Sequence analyses of the ensemble repeats from many silk proteins have revealed three major categories of conserved amino acid motifs: (1) Gly-Pro-Gly modules (GPGXX, with X often representing Q); (2) Ala-rich stretches that are 6−9 residues long or Gly-Ala couplets (GA repeated many times); and (3) two Gly residues followed by a variable amino acid (GGX). One of the extraordinary properties of silks has been attributed to the presence of the Ala-rich motifs or iterations of Gly-Ala couplets found within the ensemble repeats, which form crystalline β-sheet stacks that lead to a material with high tensile strength.[6,7] In addition, silk fibers are balanced with a high degree of extensibility, resulting in some of the toughest materials in the world that can absorb large amounts of energy before breaking. Extensibility has been attributed to GPGXX motifs, a module proposed to be involved in a β-turn spiral, likely providing elasticity.[8,9] GGX repeat motifs have been postulated to form 3_{10} helices, which form an amorphous matrix that connects crystalline regions, also providing elasticity.[10,11]

In addition to spinning fibers that serve an obvious architectural role, spiders also produce silk proteins for accessory functions. More ancient spiders, including tarantulas, have been reported to secrete glue silk into a viscous fluid that rapidly solidifies to help tarantulas cling to vertical surfaces.[12,13] Araneoid spiders secrete an analogous liquid substance that hardens in order to cement spider silk threads to physical structures during web construction.[14] They also anchor dragline silk fibers to facilitate locomotion during prey capture or predator evasion, therefore, serving a function similar to tarantula tarsal silk.[12,14] Although no sequence data exist to directly compare the proteins that comprise these different gluey materials, it is obvious that they play a critical ecological role required for both ancient and modern day spiders to survive in their respective environments.

Recently, we identified the first glue silk fibroin, called pyriform spidroin 1 (PySp1), as the major constituent of *Latrodectus hesperus* (black widow spider) attachment discs.[15] Scanning electron microscopy studies revealed that these structures consist of small diameter fibers embedded in a gelatinous substance, which dries to form a strong chemical adhesive.[15] Because anchoring silk fibers to substrates is important for locomotion and web construction in both orb and cob weavers, we hypothesized that orb weavers use similar molecular mechanisms to accomplish these tasks. To determine

* To whom correspondence should be addressed. Tel.: 209-946-3024. Fax: 209-946-3022. E-mail: cvierra@pacific.edu.

† This sequence has been deposited in GenBank and has the accession number HM020705

‡ Department of Biological Sciences.

§ Department of Chemistry.

|| School of Engineering.

10.1021/bm101002w
Published on Web 11/05/2010

whether glue silk fibroins were present within the proteome of orb weavers, we screened a cDNA library constructed from silk-producing glands of *N. clavipes*. Here we report the identification of a new member of the spider silk gene family expressed in the pyriform gland and spun into attachment discs, which we have dubbed pyriform spidroin 2 (PySp2). Phylogenetic analysis demonstrates that this fibroin is most closely related to the cob weaver glue silk fibroin, PySp1. Analysis of the PySp2 protein sequence reveals substantial deviation from the conventional motifs found in the spider silk fibroin family, a characteristic that is also common to PySp1. Significant strides have recently been made for spinning synthetic dragline silk fibers. Here we demonstrate for the first time that glue silk fibroins can be spun into artificial fibers using truncated PySp2 proteins purified from bacteria. These fibers are morphologically similar to reported synthetic fibers spun from major ampullate silk proteins at the whole fiber scale.[16,17] As a whole, our data suggest that PySp2 is the main fiber woven into the attachment discs of the orb weaver *N. clavipes*. Current silk biology research highlights the potential use of synthetic dragline silk fibers because of its high tensile strength. Fibroins such as PySp2 may represent an additional pathway, creating new biopolymers with distinct engineering properties.

Material and Methods

cDNA Library Construction. Fifteen golden orb weavers obtained from Hatari Invertebrates were dissected to obtain the silk-producing glands, including the major ampullate, minor ampullate, tubuliform, aciniform, flagelliform, pyriform, and aggregate glands. A composite cDNA library was constructed using all collected materials. Total RNA was extracted using the Trizol purification method according to manufacturer's instructions (Invitrogen). mRNA was isolated from total RNA according to manufacturer's instructions using the PolyATract mRNA Isolation System I kit (Promega). cDNA was constructed from the mRNA using the ZAP-cDNA Synthesis Kit, and the library was packaged using the ZAP-cDNA Gigapack III Gold Cloning Kit according to manufacturer's instructions (Stratagene).

Cloning of the PySp2 cDNA. A degenerate oligonucleotide was designed against *L. hesperus* PySp1 for initial *N. clavipes* library screens. This primer was designed against the repeating peptide unit AQAQAEAAA found in PySp1 and contained the degenerate nucleotide sequence 5′-WGC WGC WGC YTC WGC YTG WGC YTG WGC-3′ (W = A or T, Y = C or T). PCR was performed using a reverse anchor primer designed against the packaging vector, pBluescript SK (-). Resulting DNA bands were gel extracted, ligated into the plasmid pBAD-TOPO and transformed via heat shock into a chemically competent One Shot TOP10 cells (Invitrogen). Plasmid DNA was retrieved from 81 different transformants, digested with the restriction enzymes *PmeI* and *Nco*I, and analyzed by Southern blot using a PySp1 cDNA probe. A total of 19 positive clones from the Southern blot analysis were subjected to DNA sequencing as previously described.[18] After sequencing, the translated cDNAs were analyzed against the Non-Redundant National Center for Biotechnology Information (nrNCBI) database using the algorithm basic local alignment search tool (protein) (BLASTp). One positive clone that carried a 519 bp cDNA insert was selected for further studies. Translation of this cDNA revealed the presence of novel block repeats, a characteristic of silk-fibroin family members. This cDNA was used to screen the composite cDNA library prepared from *N. clavipes* using conventional nucleic acid-nucleic acid hybridization. Approximately 50000 plaques were screened using this methodology, resulting in 39 positive plaques. These primary plaques were put through a secondary round of plaque purification, followed by excision of the plasmid DNA from the viral chromosome using an in vivo excision protocol as described by the manufacturer (Stratagene). After plasmid DNA recovery, the cDNA inserts from the 39 different clones were released from pBluescript SK(-) by double digestion with the restriction enzymes *Eco*RI and *Xho*I. These products were then analyzed by Southern blot using the original 519 bp cDNA fragment as a radiolabeled probe. Over 20 positive clones from the Southern blot analysis were sequenced. The largest cDNA insert recovered was 3.2 kb. Four overlapping inserts ranging in size from 2 to 3.2 kb were used to assemble the final 3.2 kb PySp2 cDNA sequence.

Collection of Attachment Disc Materials from Spiders and MS Analysis. Spiders were placed into glass aquariums containing a wooden stick frame and housed for 3−7 days to allow for web construction and attachment disc formation. The dragline silk fibers were carefully severed from attachment discs with scissors. When a sterile surgical blade was used, over 25 attachment discs were scraped off the glass walls of the aquarium. Materials on the blade were collectively dissolved in less than 50 μL of 9 M LiBr and pooled in a single microcentrifuge tube. After the solubilization of the attachment discs, the solution was brought up to 100 μL of final reaction volume. To facilitate solubilization, the sample was heated for 20 min at 95 °C and then vortexed for an additional 20 min. Following vortexing, the sample was diluted from 9 to 1 M LiBr using 50 mM ammonium bicarbonate. A total of 10 μL of 1 mg/mL trypsin solution (Trypsin Gold, Promega) was added to the solution which was then incubated at 37 °C for 5 h. Samples were desalted and concentrated using C18 OMIX pipet tips according to the manufacturer's instructions (Varian, Inc.) and then analyzed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) or MALDI tandem TOF analysis. The spectrum was recorded using a α-cyano-4-hydroxycinnamic acid matrix and the instrument was operated in reflectron mode.

For the mass spectrometric analysis of the pyriform luminal contents, the proteins were extracted from the tissue using a similar methodology as previously described.[18] Following protein extraction, the sample was resuspended in buffer, digested with trypsin, and analyzed by MALDI-TOF, as outlined above.

Recombinant Expression of Pysp2 Truncation Constructs. Primers were designed to amplify three PySp2 truncations. The forward primer used to create all three constructs was 5′-CTC GAG GTA GTG TCT CAA GTA CAG CAG GCA TC-3′. The reverse primer for the largest construct containing the conserved C-terminus was 5′-GGA TCC TGC AAG TGC GGC AAG CAT AGC-3′ and led to an amplified product of 1431 bp (referred to as $PySp2_{230}$+CT). The sequence for the 690 bp construct ($PySp2_{230}$) composed of one full block repeat used the reverse primer 5′-GGA TCC GGG TGC TGG GCG GGG AAA AG-3′. The smallest PySp2 truncation ($PySp2_{190}$) corresponded to a 576 bp fragment and required the reverse primer 5′-GGA TCC AAC TAC TTG ACT AAG AGA AGA CAA GAT TGA ACT GG-3′. Amplified products were gel extracted and cloned into the pBAD/Thio-TOPO expression vector (Invitrogen) and transformed by heat shock into BL21 cells (Promega). Transformants were grown to saturation overnight with shaking at 37 °C in 150 mL of LB with 0.05 mg/mL ampicillin, then diluted 1:5 with 500 mL of LB, and supplemented with 625 μL of 20% arabinose (final concentration of 0.02%) for induction and allowed to shake at 37 °C for 4 h. The entire culture was triple pelleted in one container at 8000*g*, and the cells were resuspended with 15 mL of native cell lysis buffer according to manufacturer's instructions (Qiagen) and lysed by sonication (duration was ∼20−30 s). Insoluble cellular debris was pelleted at 10000*g*.

The supernatant containing soluble protein was removed and incubated with 1 mL of a 50% slurry of Ni-NTA Agarose (Qiagen) and affinity chromatography performed using a gravity drip column. The resin was washed with buffer, and the protein eluted with 11 mL of elution buffer according to manufacturer's instructions. Purification of $PySp2_{230}$+CT and $PySp2_{190}$ truncations was performed using denaturation buffers whereas the $PySp2_{230}$ construct was purified with nondenaturing solutions. Protein purity was confirmed by SDS-PAGE followed by silver staining using a ProteoSilver Silver Stain Kit according to manufacturer's instructions (Sigma). The elution fraction was dialyzed against deionized water for two days at 4 °C to remove

salts and imidazole and subject to protein concentration analysis, and the sample was dried completely using a Labconco FreeZone 12 Plus freeze drier. The sample was resuspended in hexafluoroisopropanol to a concentration of approximately 50 mg/mL and spun into a solid fiber in 90% isopropanol, as previously described.[13] Fibers were mounted on stubs and observed using a Hitachi S-2600 scanning electron microscopy (SEM). The artificial fibers were also mounted on a glass slide and imaged by atomic force microscopy (AFM) along the longitudinal direction using a 10 nm diameter spherical silicon oxynitrite cantilever tip on an Asylum MFP-3DBioscope (Asylum Research, CA).

Results

PySp2 Represents a New Member of the Spider Silk Gene Family. To search for novel cDNAs coding for glue silk fibroins in orb weavers, we performed PCR using a cDNA library prepared from the silk-producing glands of *N. clavipes* as a template, along with a forward primer that contained degeneracy corresponding to the peptide repeat unit AQAQAE-AAA from *L. hesperus* PySp1. This gene-specific primer was used in combination with a reverse anchor primer designed against the cDNA library cloning vector. Amplification with this primer set led to cDNA fragments ranging from 400−1800 bp, which were subsequently "shot gun" cloned into the cloning vector pBAD-TOPO. Plasmids from individual transformants were retrieved, digested and analyzed using Southern blot analysis using a radiolabeled single-stranded cDNA probe specific for the black widow PySp1. Analysis of 81 different samples resulted in 19 positive clones (data not shown). These 19 clones were further subject to DNA sequencing, which led to the identification of a 519 bp insert, translation of which revealed molecular characteristics similar to PySp1. The sequence of this clone was then used to rescreen the cDNA library by conventional nucleic acid−nucleic acid hybridization. Several positive clones with cDNAs ranging from 2 to 3.2 kb were used to reconstruct the 3,204 bp sequence of the PySp2 cDNA (GenBank Accession No. HM020705). Translation of the reconstructed cDNA sequence revealed an open-reading frame that predicted a protein containing 1012 amino acids, with a molecular weight of ~101 kDa and a pI = 11.69. The translation represents a partial sequence, as no initiator methionine was found in the N-terminal region of the protein sequence. Partial cDNA sequences are commonly reported for spider silk genes, given the length (8−12 kb) and redundancy of the transcripts.[19,20] Analysis of the translated sequence revealed ensemble repeats that were ~230 amino acids (Figure 1A). Within each ensemble repeat, there were two subrepeat units, one ~40 amino acids, which was Pro and Ala rich, and the other ~190 amino acids rich in Ser, Gln, and Ala. The smaller subrepeat contained the consensus sequence PAPRPXPAPX, with X representing a subset of amino acids that mostly had hydrophobic R groups. This element was often found repeated consecutively within the subrepeat. Another consensus sequence, (L/V)(A/S)QSQQ(A/S)S, was iterated in the large subrepeat (Figure 1A). These two different subrepeat blocks (Figure 1B) may form distinct types of secondary conformations, potentially representing important structural transitions within the fibroin.

Protein−protein BLAST searches of the nrNCBI database using the entire translated sequence revealed similarity to a single protein, *L. hesperus* PySp1 (GenBank Accession FJ973621.1). The Expect value (E-value), which is a parameter that describes the number of hits one can "expect" to see by chance when searching a database was found to be 3e−15. The similarity was confined to the last 105 amino acid residues, showing 54% identities and 73% positives (percent positives include both amino acid residues that are identical and amino acid residues that have similar side-chain properties, e.g., Arg and Lys). This region corresponds to the nonrepetitive C-terminal region that is conserved in spider silk family members. When only these last 105 amino acid residues were used to perform BLASTp searches of the nrNCBI database, it again revealed the strongest similarity to *L. hesperus* PySp1 (E-value = 1e−14), as well as other fibroin family members with E-values ranging from 3e−04 to 8.5. Manual alignment of the C-terminal regions of fibroins identified in *N. clavipes* shows a high degree of conservation within the sequence, which demonstrates that PySp2 is a new member of the silk fibroin family (Figure 1C). Based upon our phylogenetic analysis of PySp2 (data not shown), its expression pattern (Figure 2A) and its fiber localization (Figure 2B), we have named this new fibroin Pyriform Spidroin 2 (PySp2).

PySp2 Shows a High Degree of Polarity. Theoretical amino acid composition analysis of several members of the *N. clavipes* fibroin family was performed using the longest versions of the cDNAs deposited in GenBank. This comparison demonstrates that PySp2 contains the highest percentage of polar amino acids, comprising Ser + Thr (S+T), Gln (Q), Glu (E), and Arg + Lys (R+K). These residues account for 49% of the protein sequence, whereas the second most polar fibroin is TuSp1 with 37.8% (Figure 1D). We reported an identical trend when comparing PySp1 to other fibroin family members from *L. hesperus* (15). Ser levels for PySp2 are highest at 25.7%. Gln is 16.9% and Ala 16.4%. Strikingly, very little Gly (2.9%) is found within the PySp2 translated sequence, representing the lowest Gly content in all reported *N. clavipes* fibroins (Figure 1D). When comparing the predicted amino acid profile of *N. clavipes* PySp2 to *L. hesperus* PySp1, both display strikingly low levels of Gly along with high degrees of polarity (S+T, Q, E, R+K), 49 and 46%, respectively. Relative to PySp1, PySp2 contains lower levels of basic and acidic amino acid residues (E and R+K) but higher amounts of Ser and Gln. PySp2 also contains considerably lower Ala content than PySp1 but with a concurrent increase in the levels of other hydrophobic amino acids, in particular, Pro and Leu (Figure 1D). Collectively, these data show that the overall percent polarity of PySp2 and PySp1 is very similar.

PySp2 Expression is Restricted to the Pyriform Tissue. To determine which silk-producing gland(s) is responsible for the synthesis of PySp2, we dissected the following tissues from *N. clavipes*: the major and minor ampullate, tubuliform, aciniform, flagelliform, and pyriform glands. Luminal contents were independently extracted from these different glands and in-solution trypsin digests were performed on the protein lysates. MS analysis of pyriform glandular lysate revealed over 16 peptide ions with distinct masses. These peptide masses (MH^+, monoisotopic) included 1032.3, 1060.5, 1254.7, 1270.4, 1515.3, 1562.4, 1603.8, 1790.4, 1820.6, 1847.4, 2124.1, 2214.5, 2524.8, 2542.8, 2583.9, and 2702.4 (Figure 2A). Three of these peptide ion masses (1603.8, 2124.1, and 2702.4) match predicted products from the translated PySp2 sequence after theoretical tryptic digest (Figure 1A and Table 1). Two of the peptide ions, those with *m/z* values of 1603.8 and 2124.1, correspond to regions found within the C-terminal nonrepetitive, conserved region. These peptides and their corresponding masses are unique to PySp2, despite the high conservation within the C-terminal region of the spider silk fibroin family. Additionally, the peptide ion with *m/z* 2702.4, which is also exclusively found within the PySp2 sequence, localizes to the block repeats of PySp2 (Figure 1A). Overall, our MS analysis detected 3 out of

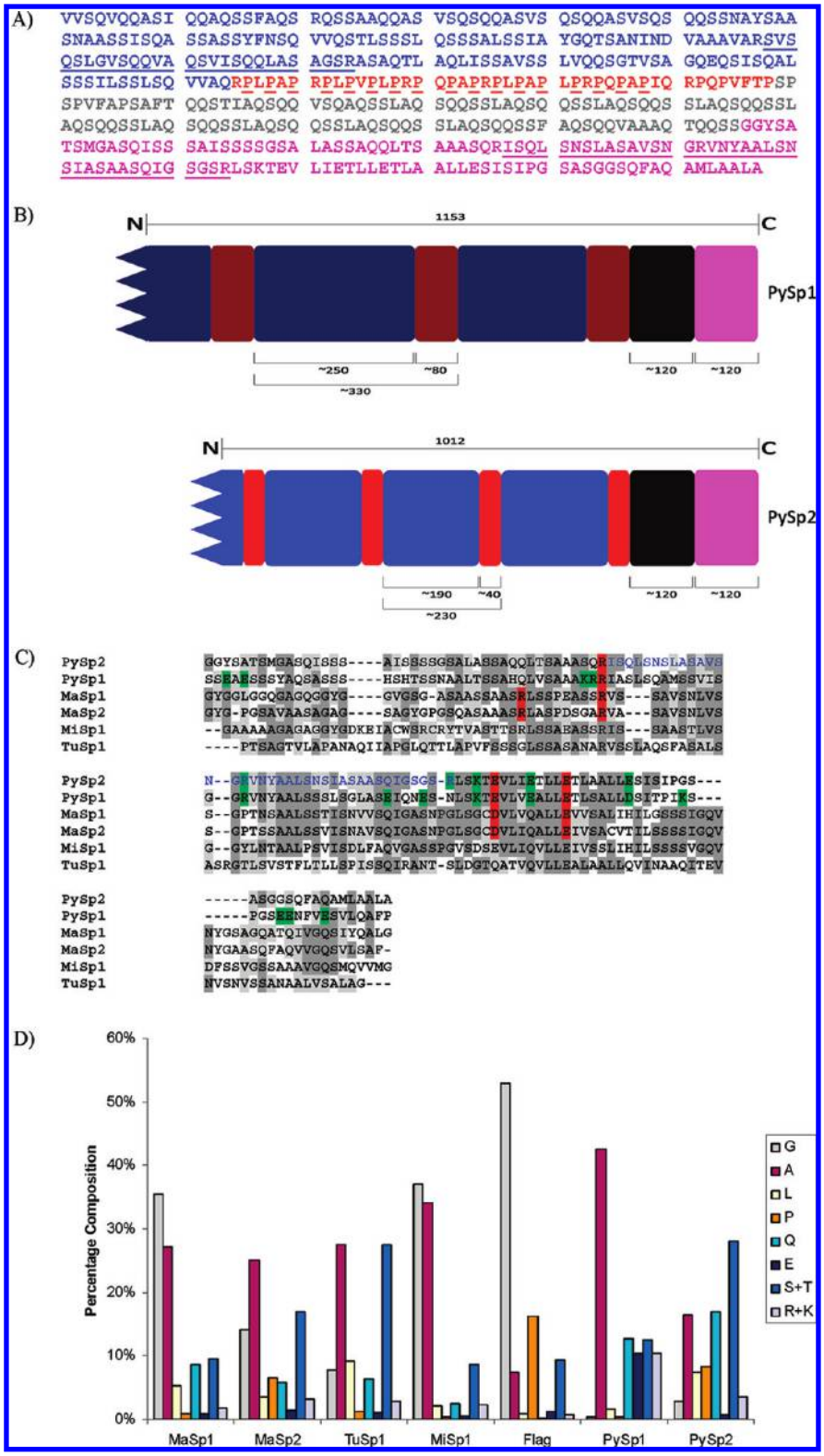

Figure 1. Amino acid sequence of *N. clavipes* PySp2 shows similarity to spider fibroin family members. (A) Translation of the PySp2 cDNA sequence. Peptide sequences determined by MS or MS/MS analysis that are found within the open reading frame are indicated with a solid underline. Single character underlined regions reflect Pro-rich segments. Red and blue coloration denotes the 2 sub-block repeats found within one ensemble repeat. The last 122 amino acids (pink) represent the nonrepetitive C-terminus. (B) Schematic comparison of *L. hesperus* PySp1 with *N. clavipes* PySp2. (C) Alignment of PySp2 carboxy-terminal sequence to other fibroin family members. Alignments were performed with the last C-terminal 122 amino acids of the fibroins using the computer algorithm CLUSTALW. Amino acids are represented by one-letter abbreviations, gaps are indicated by dashes. Dark-shaded gray areas represent identical residues that are found in the highest frequency. Light-shaded patterning indicates residues that have common polarity or side chain size. The blue underlined regions represent two segments that correspond to peptide ion masses 1603.8 and 2124.1. Red-shaded letters represent residues shown to be involved in salt-bridge formation in MA silks[22] and functionally conserved residues in glue silk fibroins. Green-shaded regions indicate amino acid residues with charged R-groups. (D) Predicted amino acid composition profile from the translated *N. clavipes* cDNAs of major ampullate spidroin 1 (MaSp1; AY654292.1), major ampullate spidroin 1 (MaSp2; EU599240.1), tubuliform spidroin 1 (TuSp1; AAX45295.1), minor spidroin 1 (AF027735.1), flagelliform silk (AF218621 and AF218622), PySp2 (HM020705), and *L. hesperus* PySp1 (FJ973621).

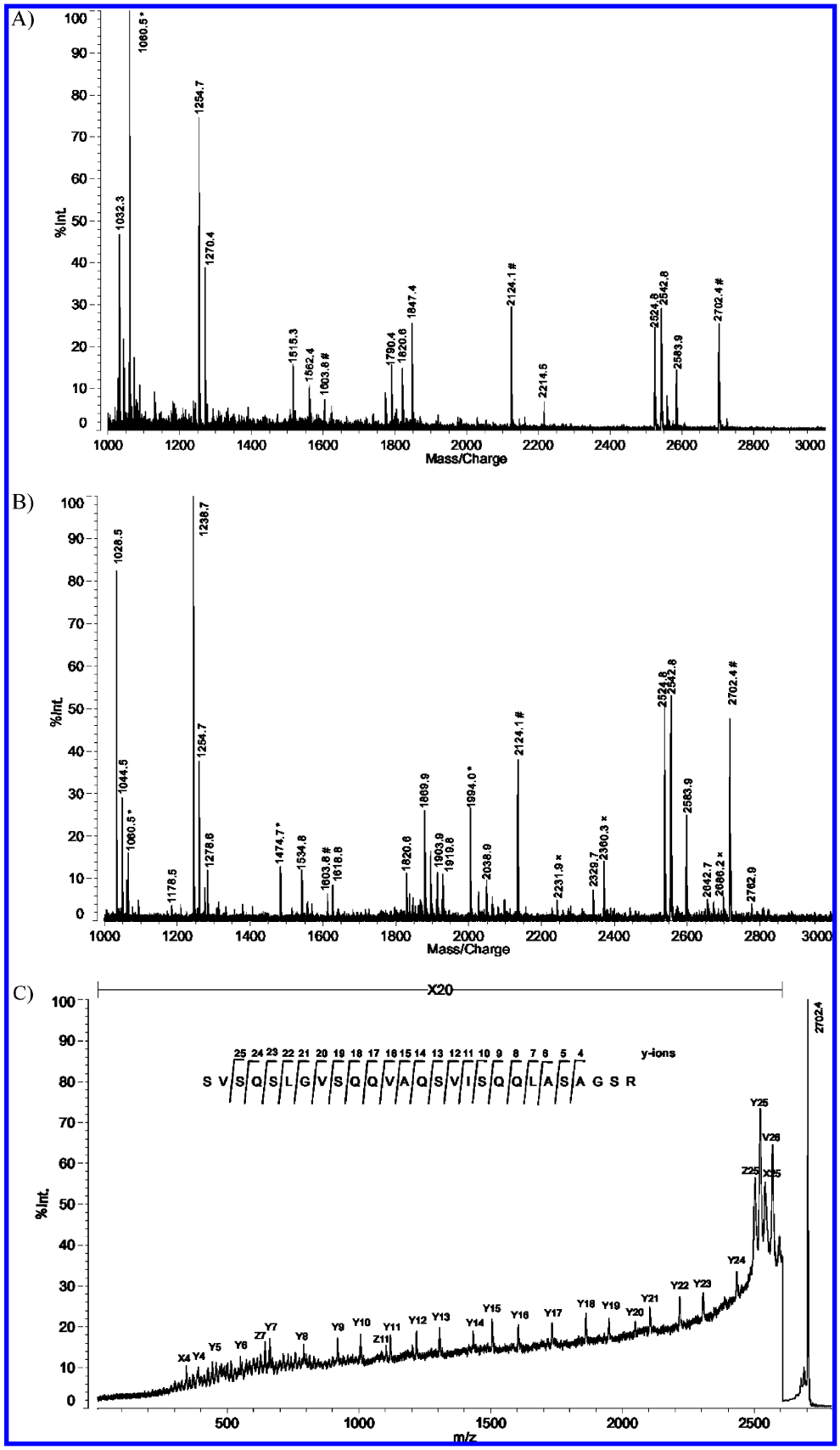

Figure 2. MS and MS/MS analysis shows the presence of PySp2 in the luminal contents of the pyriform gland and attachment disc fibers. (A) MALDI-TOF mass spectrum of an in-solution trypsin digestion of the luminal contents extracted from the pyriform gland of *N. clavipes*; peptide ions from PySp2 (1603.8, 2124.1, and 2702.4) are labeled with "#". (B) MS spectrum following in-solution tryptic digestion of the attachment discs from *N. clavipes*. Keratin or tryptic autolysis peptide signals are labeled with "*". Peptide ion masses labeled with "×" match products from theoretical digestion of MaSp1. (C) MS/MS spectrum of precursor ion with *m*/*z* 2702.4 after high energy collision-activated dissociation (MH^+, monoisotopic) reveals the sequence SVSQSLGVSQQVAQSVISQQLASAGSR. The internal leucine or isoleucine residues could be any combination of leucine and isoleucine, but exact amino acids were determined using the codons found within the PySp2 cDNA sequence (Figure 1A).

Table 1. Observed *m/z* Values [MH+, Monoisotopic] of PySp2 Peptides from In-Solution Tryptic Digestion of the *N. clavipes* Pyriform Glandular Extract and Solubilized Attachment Discs

	PySp2 peptides	$[M + H]^+_{theor}$ (*m/z*)	$[M + H]^+_{obs}$ (*m/z*)
1	I_{932}SQLSNSLASAVSNGR$_{947}$	1603.8	1603.8
2	V_{948}NYAALSNSIASAASQIGSGSR$_{969}$	2124.1	2124.1
3	S_{415}VSQSLGVSQQVAQSVISQQLASAGSR$_{441}$	2702.4	2702.4
4	S_{653}VSQSLGVSQQVAQSVISQQLASAGSR$_{679}$	2702.4	2702.4

4 peptide masses expected from theoretical digestion of the translated PySp2 cDNA sequence. An ion mass of 3987.1 was predicted but not observed, perhaps due to ion suppression. Several additional peptide ion masses not corresponding to PySp2 fingerprints were detected in our analysis. These could either be derived from other proteins stored within the pyriform tissue or, alternatively, may reflect distinct peptides derived from the missing N-terminal sequence of our PySp2 translated product. Attempts to detect PySp2 peptide ion masses after tryptic digestion of protein extracts collected from all of the other silk-producing glands were unsuccessful (data not shown). Therefore, PySp2 is one of the constituents stored within the pyriform gland of *N. clavipes*, and protein expression is restricted to the pyriform gland.

PySp2 is Found in the Attachment Discs. One of the chief functions proposed for the pyriform tissue has been the production of attachment disc fibers.[14] Because our MS data supported the presence of PySp2 in the luminal contents of the pyriform gland, we examined the attachment discs from *N. clavipes* to determine whether PySp2 was spun into this matrix. Attachment discs were dissolved in LiBr and the solubilized proteins were subject to in-solution tryptic digestion. The tryptic products were then examined using MS analysis. Over 23 different peptide ion masses were detected. Three of the 23 peptide ion masses match products from the theoretical tryptic digestion of the translated PySp2 sequence; these include *m/z* values of 1603.8, 2124.1, and 2702.4. These three ion masses are identical to tryptic products detected from the luminal contents of the pyriform gland after digestion (Figure 2A,B). Additional analysis of precursor ion *m/z* 2702.4 by MS/MS revealed a 100% match to the block repeat within the translated PySp2 sequence (Figure 2C), corresponding to the peptide SVSQSLGVSQQVAQSVISQQLASAGSR. (Figure 1A and Table 1). Overall, these data demonstrate that PySp2 is a constituent of the attachment discs of orb weavers. Additionally, four peptide ion masses that were not accounted for within the PySp2 sequence after theoretical digestion were observed in both the pyriform luminal contents and attachment disc extracts following trypsin treatment. These peptide ion masses corresponded to *m/z* values 1254.7, 1820.6, 2542.8, and 2583.9 (Figure 2A,B). Their presence in both samples after in-solution tryptic digestion suggests that they are derived from the missing N-terminal sequence of PySp2 and/or represent other proteins secreted by the pyriform during attachment disc formation. Several ion masses unique to the solubilized attachment disc materials after trypsin digestion were detected, including *m/z* values of 1028.5, 1044.5, 1278.6, 1534.8, 1618.8, 1869.9, 1903.9, 1919.8, 2038.9, 2329.7, and 2360.7 (Figure 2B). Because these peptide ion masses were exclusively found in the attachment disc materials and not within the pyriform gland luminal contents, it implies that other proteins from different silk-producing glands are also spun into attachment discs. Aligning these data with theoretical tryptic digestion of other *N. clavipes* silk proteins deposited in the protein database reveals the presence of peptide ion masses matching the major ampullate silk fibroin MaSp1 (2231.9, 2360.3, 2686.2; Figure 2B). The presence of major ampullate silk fibroins is consistent with the fact that dragline silk fibers are embedded within the attachment disc structure.[15]

PySp2 Truncations Self-Assemble into Solid Fibers. Previous studies have revealed that black widow spider attachment disc fibers are embedded in a glue-like matrix.[15] To determine the ultrastructure of the attachment disc fibers from orb-weaving spiders, we used scanning electron microscopy to visualize the fibers. SEM analysis revealed a complex network of small glue silk fibers, serving to anchor the larger diameter dragline silk fibers similar to attachment disc morphology seen in the black widow spider (Figure 3A,B).[15] Seeking to mimic the natural glue silk fibers, we expressed and purified three truncated versions of the PySp2 protein from bacteria (Figure 3C,D). The largest construct ($PySp2_{230}$+CT) contained the conserved C-terminus and one full block repeat, as well as the intervening sequence between them. The second construct ($PySp2_{230}$) contained one full block repeat, and the third ($PySp2_{190}$) comprised of the S, Q, and A rich subrepeat of the block region, while lacking the proline rich subrepeat (Figure 3C). The purified proteins (Figure 3D) were forced through a syringe into an isopropanol bath. All three constructs successfully resulted in fiber formation (Figure 4B−D). These results demonstrate that the S+Q+A-rich region is sufficient for self-assembly and fiber formation (Figure 4A,B). Analysis of the $PySp2_{230}$ spun fibers using AFM revealed a periodic banding that was consistent throughout the fiber, suggesting that there are polymeric subunits that make up the whole fiber. The diameters of the single polymer chains are on the order of 150 nm (Figure 4E,F).

Discussion

Here, we document the discovery of PySp2, a new glue silk fibroin that is the major structural component of attachment discs produced by the orb weaver *N. clavipes*. We also report the first glue silk fibroin to be spun into an artificial fiber. Phylogenetic analysis indicates that the gene is closely related to *L. hesperus* PySp1, the only other spider glue silk protein discovered to date.[15] Analysis of its conserved C-terminus puts PySp2 squarely into the well-characterized, diverse spider silk fibroin family.[21] Alignment of the nonrepetitive (NR) C-terminus of PySp2 to other members of the silk family shows strong sequence conservation at specific positions, implying this region may potentially serve a similar biological function. Inspection of the PySp2 protein sequence, as well as the only other identified glue silk protein, PySp1, reveals the presence of three out of four residues necessary for salt bridge formation in the C-termini of MA fibroins (R52, D93 [E in glue silks] and E101); however, there is a deviation at position 48 with the presence of an uncharged Q (Figure 1C; red coloration sites).[22] The lack of a positively charged residue at position 43 (R43 or K43) in the NR C-termini of glue silks could imply a different structural conformation relative to MA silks and/or the use of a different salt bridge within their polypeptide chains. Interestingly, the glue silks NR C-termini contain the highest percentage of amino

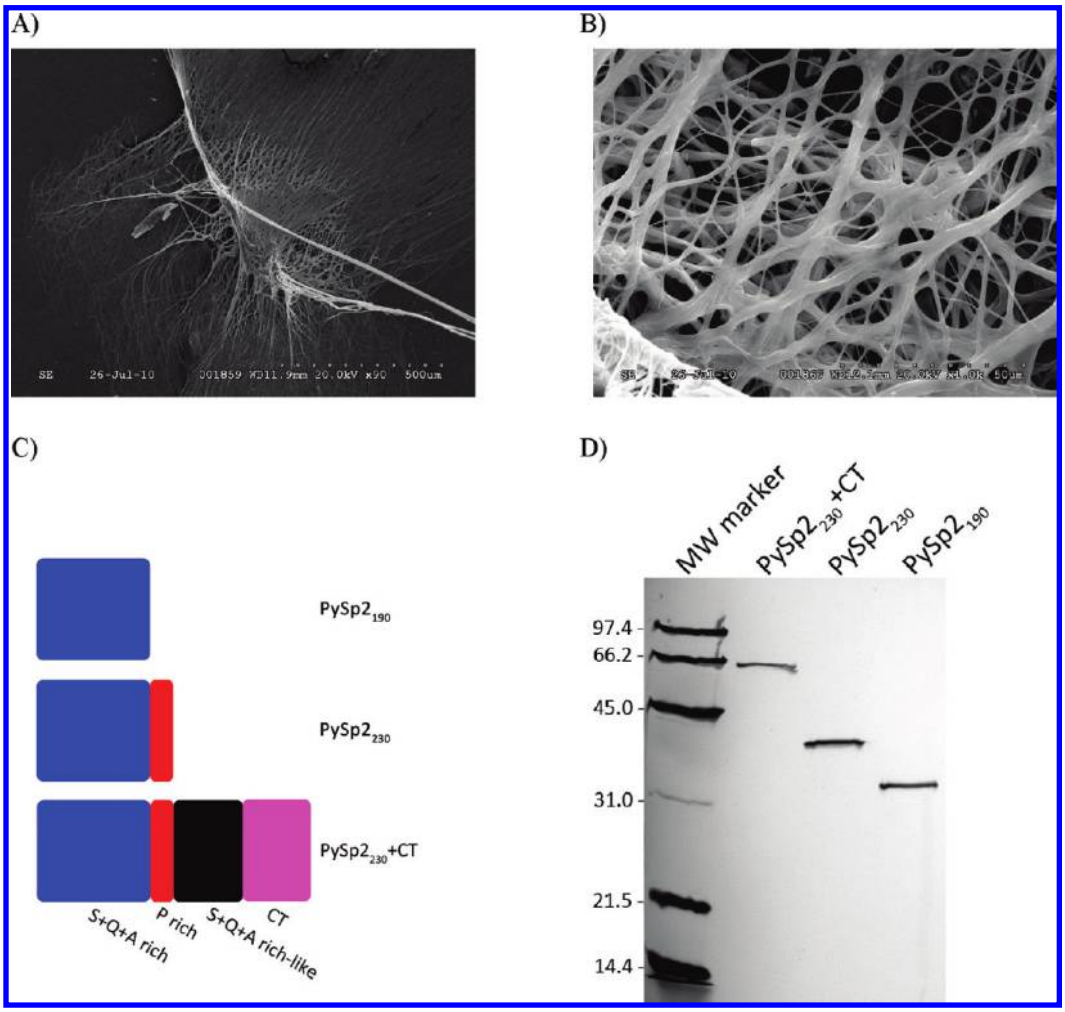

Figure 3. Scanning electron microscopy of natural attachment disc from *N. clavipes* and SDS-PAGE analysis of purified PySp2 constructs. (A) 90× magnification of attachments discs. (B) 1000× magnification of attachment discs. (C) Schematic representation of three heterologously expressed PySp2 proteins: $PySp2_{190}$, $PySp2_{230}$, and $PySp2_{230}$ + CT. (D) Size fractionation of the purified PySp2 truncations using SDS-PAGE analysis, followed by protein visualization via silver staining.

acid residues with basic or acidic side chain groups within the spider silk family, but whether this influences their secondary structure is unknown (Figure 1C; green coloration). Recent molecular details reveal the NR C-termini of major ampullate (MA) silks are essential for controlling the switching between the storage and assembly forms of MA proteins.[22] High resolution NMR studies involving recombinant MA silk proteins demonstrate their NR C-termini form a parallel-oriented dimeric five-helix bundle with helix 4 constituting the main dimerization site.[22] The formation of two highly conserved salt bridges (R43-D93 and R52-E101) that involve helices 1, 2, and 4 provide critical intermolecular interactions that serve to maintain the correctly folded NR domain for silk protein storage and the inhibition of undesired aggregation in MA silks.[22] Further biophysical studies will need to be performed to determine whether the NR C-termini of glue silks have similar protein folds relative to MA silks and serve as molecular switches to control fiber assembly.

Similar to the cob weaver glue silk PySp1, our MS data shows that the expression pattern of PySp2 is restricted to the pyriform gland, the tissue which has long been hypothesized to be responsible for attachment disc biosynthesis.[23,24] MS analysis proves that the gene is not only expressed in the pyriform gland, but also that the product is indeed incorporated into the attachment disc structure. Recombinant spinning of PySp2 leads to fibers equivalent in appearance to what has been observed for synthetic fibers spun from heterologous expression systems using purified major ampullate silk fibroins.[17] Our results show that the block repeats of PySp2 self-assemble into a fiber with or without the aid of the conserved C-terminus and the Pro-rich subrepeat. It will be interesting to observe what effects the lack of these regions has on the mechanical properties of the synthetic fiber. Of note is the fact that all of the fibers were spun from thioredoxin tagged fusion proteins. It has been reported, and our experience confirms, that thioredoxin aids in the expression/solubility of fibroins.[25] High expression and solubility will be necessary if the silk spinning process is to be upscaled for mass production; however, it is unknown what effects the thioredoxin tag may also have on the mechanical properties of the glue silk fibers and should be investigated. Although our PySp2 constructs were designed with an enterokinase protease cleavage site between the thioredoxin tag and the glue silk protein, removal of the thioredoxin tag after purification was not necessary for self-assembly of the proteins. Initial attempts to express the truncated PySp2 proteins in bacteria without the thioredoxin tag resulted in lower expression levels. However, if the thioredoxin tag is shown to influence the mechanical properties of the spun fibers, another method to enhance expression/solubility will need to be developed.

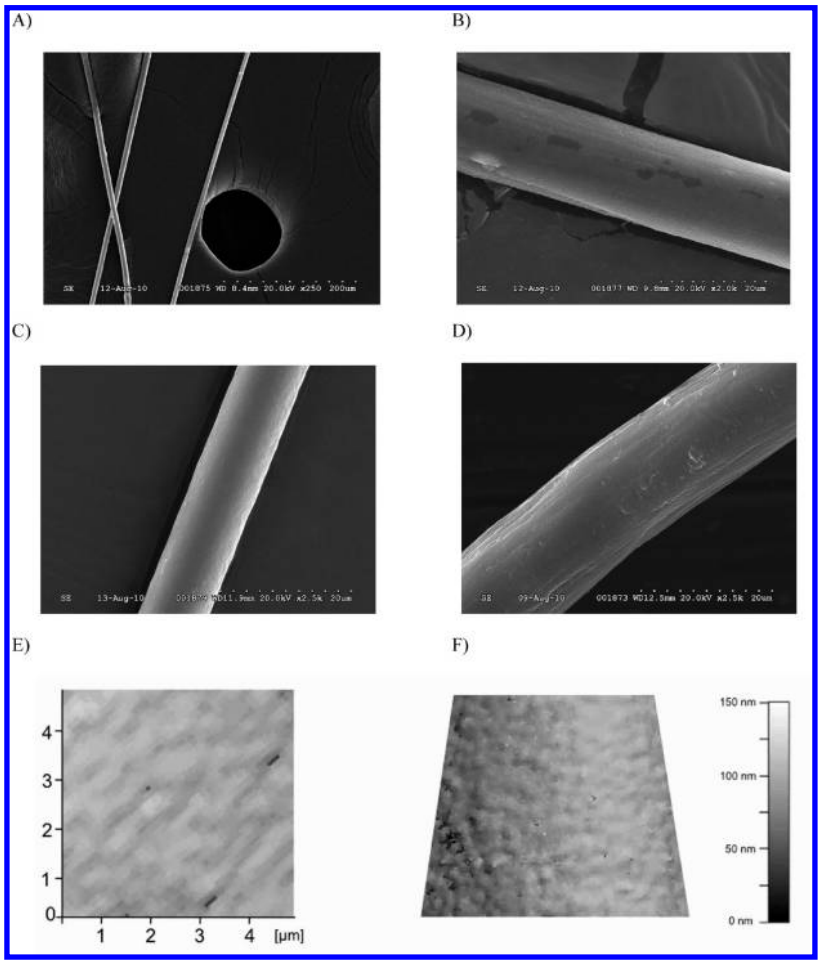

Figure 4. Analysis of glue silk fibers spun from purified PySp2 recombinant proteins using SEM (A−D) or AFM (E,F). (A) 250× magnification of $PySp2_{190}$ spun fibers. The large circular hole in the middle of the image is from the background of the SEM stub. (B) 2000× magnification of $PySp2_{190}$ spun fibers. (C) 2500× magnification of $PySp2_{230}$ spun fibers. (D) 2500× magnification of $PySp2_{230}$ + CT spun fibers. (E) Atomic force micrograph of the surface of the $PySp2_{230}$ spun fibers. The shown area approximates 5 × 5 μm, and the repeated banding pattern is typical throughout the fiber, suggesting that the spun fiber may be constituted by polymeric subunits in the ultrastructure. (F) A 3-dimensional plot of the same region from the AFM image that demonstrate the topographical variations that may be attributed to the polymeric subunits within the whole fiber.

The only two spidroins found to originate in the pyriform gland share a high degree of polarity within their protein sequences, which includes high glutamine content, a characteristic that sets them apart from all other fibroins, and may result from the peculiarity of being spun into the aqueous matrix of the attachment disk. In vertebrate proteins, high glutamine levels have been associated with protein aggregation.[26] Therefore, one possibility is that the high glutamine in PySp2 may facilitate protein aggregation or the self-assembly process. There are clear functional similarities between PySp2 and PySp1; however, surprisingly, these fibroins execute the same function using distinct internal block repeats. While other fibroin family members show significant sequence similarity in their block repeat units when analyzed across species (i.e., it is clear that they are orthologous proteins), the block repeats of PySp1 are unrecognizable in PySp2. Repeating units in PySp2 are smaller, approximately 230 amino acids in length compared to 316−376 in PySp1. Strikingly, the ensemble repeat unit of both pyriform spidroins is subdivided into two subrepeats. In PySp2, the smaller 44 amino acid subrepeat is dominated by Pro (spaced every other residue) and Ala, while the larger 186 amino acid subrepeat is rich in Ser, Gln, and Ala. In contrast, the smaller 78 amino acid subrepeat of PySp1 is dominated by equal amounts of Ser and Ala with lower but significant levels of His, Thr, and Gln, while the larger ∼298 amino acid subrepeat is more than half Ala, with Gln, Arg, and Glu accounting for the majority of the remaining sequence. If pyriform spidroins form fibers with crystalline and amorphous regions similar to other silks, perhaps the observed subrepeats are serving these distinct purposes. Qualitative observation of fibers spun from $Pysp2_{190}$ revealed fibers that were stiffer relative to the other two truncations, and aggregation appeared to occur more rapidly. This seems to indicate that the purpose of the Pro-rich region in the recombinant fibers is to function as an extensibility module. Further biophysical studies will be needed to pinpoint the precise function of the subrepeat blocks in PySp2. Additionally, structural analysis will be needed to determine whether the secondary and tertiary structures formed by the subrepeat regions of both attachment disc spidroins are similar, and whether their conformations mimic other spidroins. The homogeneous nature of the SEM and AFM imaging show that there are no clear defects in the microstructure of the whole fibers, suggesting that material processing techniques may be quite viable. AFM imaging also reveals that there may be polymeric subunits. This suggests the recombinant fibroins self-assemble into polymer chains, allowing fiber manufacturing processes, such as extrusion spinning.

Pyriform spidroin protein sequences differ substantially from other terrestrial silks; however, some insects have evolved mechanisms to spin fibers in an aquatic environment. Because

pyriform silks are spun into a liquid matrix, it raises the question whether pyriform spidroin sequences share similarities to fibroins generated by these aquatic silk-spinning organisms. Recently, it was reported that the larval stage of a caddisfly, an aquatic insect, spins silks underwater for protection and food gathering.[27] These silks are incredibly polar, containing large amounts of serine, similar to the pyriform silks. In the caddisfly silk proteins, the serine residues are heavily phosphorylated and associated with divalent cations. This neutralizes the charge imbalance of the silk protein, allowing the exclusion of water and aiding in polymerization.[27] One interesting future study will involve determining whether PySp2 is subject to serine phosphorylation and whether the attachment disc matrix contains inorganic divalent cations, which could help elucidate the mechanism of attachment disc adhesion.

Within the silk fibroin family, there are often examples of paralogues that are coexpressed in the same silk-producing gland (e.g., MaSp1 and MaSp2).[28,29] We attempted to determine whether PySp2 was simply a paralogue of PySp1 by searching for the PySp2 sequence within a black widow spider composite cDNA library, but were unable to find evidence of its presence. This is consistent with MS/MS data for 55 peptides obtained from in-solution tryptic digestion of the attachment discs of the black widow spider, which failed to produce sequences either identical or similar to PySp2.[15] Combined with the fact that we could not identify PySp1 through screening a cDNA library prepared from *N. clavipes* silk glands, these data support the assertion that attachment discs from orb and cob weavers are not composed of a combination of PySp1 and PySp2. Given this observation, a central issue that surfaces involves the evolutionary history of the pyriform spidroins. Phylogenetic data indicate that orb weavers gave rise to cob weavers.[5] Therefore, one hypothesis is that the cob weaver pyriform gene resulted from drastic mutations within the repeating blocks of the orb weaver gene. Alternatively, the orb weaver pyriform gene may have duplicated, giving rise to two distinct pyriform genes, with one copy mutating drastically, followed by a loss-of-function of the ancestral gene. Further analysis of the genomic DNA from the orb and cob weaver will be required to address which process most likely occurred during cob weaver evolution.

Conclusion

From a biomaterials standpoint, PySp2 is an intriguing structural protein because it has evolved as a specialized fibroin whose biochemical characteristics are markedly distinct relative to other fibroin family members. First, its chemical properties allow it to be spun and function in a liquid environment. Most of the well-studied silks (e.g., dragline silk) are known to experience supercontraction when exposed to water, a process that affects mechanical properties and is due to the malleable nature of the amorphous regions. The crystalline region is mostly unaffected.[30,31] It will be interesting to observe whether the PySp2 spun fibers are susceptible to supercontraction given its distinct chemical makeup, block repeat architecture and the fact that they are naturally spun into a liquid material that readily dries. An interesting application of PySp2 could be the utilization of its unique repeat blocks for generating artificial silk proteins. Scientists are currently investigating the possibility of engineering chimeric fibroins by combining domains from different fibroins to optimize specific mechanical features.[32] The discovery of PySp2 now expands the repertoire of known self-assembling sequences that are available for integration, perhaps opening the door for the creation of next generation materials for a wide variety of applications.

Acknowledgment. This work was supported by a NSF RUI Grant MCB-0950372 entitled *Molecular Characterization of Black Widow Spider Silks*. We appreciate the critical reading of the manuscript by Xiaoyi Hu (Abbott Laboratories, Santa Clara).

Note Added after ASAP Publication. This article posted ASAP on November 5, 2010. Leah Larkin has been added as a co-author. The correct version posted on November 30, 2010.

References and Notes

(1) Coddington, J. A.; Levi, H. W. *Annu. Rev. Ecol. Syst.* **1991**, *22*, 565–592.
(2) Vollrath, F.; Knight, D. P. *Int. J. Biol. Macromol.* **1999**, *24* (2−3), 243–9.
(3) Gosline, J. M.; Guerette, P. A.; Ortlepp, C. S.; Savage, K. N. *J. Exp. Biol.* **1999**, *202*, 3295–303.
(4) Gosline, J. M.; DeMont, M. E.; Denny, M. W. *Endeavour* **1986**, *10*, 31–43.
(5) Gatesy, J.; Hayashi, C.; Motriuk, D.; Woods, J.; Lewis, R. *Science* **2001**, *291* (5513), 2603–5.
(6) Simmons, A.; Ray, E.; Jelinski, L. W. *Macromolecules* **1994**, *27*, 5235–5237.
(7) Hinman, M. B.; Jones, J. A.; Lewis, R. V. *Trends Biotechnol.* **2000**, *18* (9), 374–9.
(8) Hayashi, C.; Lewis, R. V. *J. Mol. Biol.* **1998**, *275* (5), 773–84.
(9) Zhou, Y.; Wu, S.; Conticello, V. P. *Biomacromolecules* **2001**, *2* (1), 111–25.
(10) van Beek, J. D.; Hess, S.; Vollrath, F.; Meier, B. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (16), 10266–71.
(11) Dong, Z.; Lewis, R. V.; Middaugh, C. R. *Arch. Biochem. Biophys.* **1991**, *284* (1), 53–7.
(12) Gorb, S. N.; Niederegger, S.; Hayashi, C. Y.; Summers, A. P.; Votsch, W.; Walther, P. *Nature* **2006**, *443* (7110), 407.
(13) Perez-Miles, F.; Panzera, A.; Ortiz-Villatoro, D.; Perdomo, C. *Nature* **2009**, *461* (7267), E9–10, discussion.
(14) Foelix, R. *Biology of Spiders*; Oxford University Press: New York, 1996.
(15) Blasingame, E.; Tuton-Blasingame, T.; Larkin, L.; Falick, A. M.; Zhao, L.; Fong, J.; Vaidyanathan, V.; Visperas, A.; Geurts, P.; Hu, X.; La Mattina, C.; Vierra, C. *J. Biol. Chem.* **2009**, *284* (42), 29097–108.
(16) Lazaris, A.; Arcidiacono, S.; Huang, Y.; Zhou, J. F.; Duguay, F.; Chretien, N.; Welsh, E. A.; Soares, J. W.; Karatzas, C. N. *Science* **2002**, *295* (5554), 472–6.
(17) Xia, X. X.; Qian, Z. G.; Ki, C. S.; Park, Y. H.; Kaplan, D. L.; Lee, S. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (32), 14059–63.
(18) Vasanthavada, K.; Hu, X.; Falick, A. M.; La Mattina, C.; Moore, A. M.; Jones, P. R.; Yee, R.; Reza, R.; Tuton, T.; Vierra, C. *J. Biol. Chem.* **2007**, *282* (48), 35088–97.
(19) Colgin, M. A.; Lewis, R. V. *Protein Sci.* **1998**, *7* (3), 667–72.
(20) Hayashi, C. Y.; Lewis, R. V. *Science* **2000**, *287* (5457), 1477–9.
(21) Challis, R. J.; Goodacre, S. L.; Hewitt, G. M. *Insect Mol. Biol.* **2006**, *15* (1), 45–56.
(22) Hagn, F.; Eisoldt, L.; Hardy, J. G.; Vendrely, C.; Coles, M.; Scheibel, T.; Kessler, H. *Nature* **2010**, *465* (7295), 239–42.
(23) Craig, C. L. *Annu. Rev. Entomol.* **1997**, *42*, 231–67.
(24) Vollrath, F.; Knight, D. P. *Nature* **2001**, *410* (6828), 541–8.
(25) Stark, M.; Grip, S.; Rising, A.; Hedhammar, M.; Engstrom, W.; Hjalm, G.; Johansson, J. *Biomacromolecules* **2007**, *8* (5), 1695–701.
(26) Kim, S.; Nollen, E. A.; Kitagawa, K.; Bindokas, V. P.; Morimoto, R. I. *Nat. Cell Biol.* **2002**, *4* (10), 826–31.
(27) Stewart, R. J.; Wang, C. S. *Biomacromolecules* **2010**, *11* (4), 969–74.
(28) Xu, M.; Lewis, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87* (18), 7120–4.
(29) Hinman, M. B.; Lewis, R. V. *J. Biol. Chem.* **1992**, *267* (27), 19320–4.
(30) van Beek, J. D.; Kummerlen, J.; Vollrath, F.; Meier, B. H. *Int. J. Biol. Macromol.* **1999**, *24* (2−3), 173–8.
(31) Yang, Z.; Livvak, O.; Seidel, A.; LaVerde, G.; Zax, D. B.; Jelinski, L. W. *J. Am. Chem. Soc.* **2000**, *12*, 9019–9025.
(32) Asakura, T.; Nitta, K.; Yang, M.; Yao, J.; Nakazawa, Y.; Kaplan, D. L. *Biomacromolecules* **2003**, *4* (3), 815–20.

BM101002W