Araneoid Egg Case Silk: A Fibroin with Novel Ensemble Repeat Units from the Black Widow Spider, *Latrodectus hesperus*†

Xiaoyi Hu,‡ Barbara Lawrence,§ Kristin Kohler,|| Arnold M. Falick,⊥ Anne M. F. Moore,|| Erin McMullen,|| Patrick R. Jones,‡ and Craig Vierra*,||

Department of Chemistry and Department of Biology, University of the Pacific, Stockton, California 95211, Department of Chemistry, Eastern Illinois University, Charleston, Illinois 61920, and Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720

Received March 17, 2005; Revised Manuscript Received May 29, 2005

ABSTRACT: Araneoid spiders use specialized abdominal glands to manufacture up to seven different protein-based silks/glues that have diverse physical properties. The fibroin sequences that encode egg case fibers (cover silk for the egg case sac) and the secondary structure of these threads have not been previously determined. In this study, MALDI tandem TOF mass spectrometry (MS/MS) and reverse genetics were used to isolate the first egg case fibroin, named tubuliform spidroin 1 (TuSp1), from the black widow spider, *Latrodectus hesperus*. Real-time quantitative PCR analysis demonstrates TuSp1 is selectively expressed in the tubuliform gland. Analysis of the amino acid composition of raw egg case silk closely aligns with the predicted amino acid composition from the primary sequence of TuSp1, which supports the assertion that TuSp1 represents a major component of egg case fibers. TuSp1 is composed of highly homogeneous repeats that are 184 amino acids in length. The long stretches of polyalanine and glycine-alanine subrepeats, which account for the crystalline regions of minor ampullate and major ampullate fibers, are very poorly represented in TuSp1. However, polyserine blocks and short polyalanine stretches were highly iterated within the primary sequence, and ^{13}C NMR spectroscopy demonstrated that the majority of alanine was found in a β-sheet structure in post-spun egg case silk. The TuSp1 repeat unit does not display substantial sequence similarity to any previously described fibroin genes or proteins, suggesting that TuSp1 is a highly divergent member of the spider silk gene family.

Spider silk has extraordinary strength and toughness (*1*). The superior properties of the silk are derived from its chemical composition as a semicrystalline polymer that contains crystallite regions embedded in an amorphous matrix, which represent less organized regions. Many of the mechanical properties of the silk can be attributed to repetitive modules found within the primary sequence of the fibroins. NMR and X-ray diffraction experiments show that the crystalline regions consist of pleated β-sheet structures formed from polyalanine stretches that provide strength to the fiber, whereas the amorphous matrix, which provides elasticity, contains a glycine-rich 3_1-helix secondary structure (*2*−*4*). During fibroin manufacturing, silk proteins are secreted and stored as a liquid crystalline spinning dope (the material from which silk is spun). The dope ultimately passes through the spinning duct and is converted into a solid thread induced by extensional flow (*5*).

Orb-web-weaving spiders use specialized abdominal glands to synthesize up to seven types of fibroins and glues. On the basis of differences of amino acid composition of liquid collected from different silk-producing glands, it has been hypothesized that araneids generate silk diversity by the expression of different fibroin genes, which consist of an unknown number of spider silk gene family members. Of the seven different silk glands found in a typical araneid orb-weaver spider, fibroin cDNAs have been characterized from the major ampullate [MA; produces dragline and web frame silk (*6*−*11*)], minor ampullate [MI; makes temporary capture spiral silk (*9*, *12*)], flagelliform [assembled into the core fiber of the capture spiral (*11*, *13*, *14*)], and aciniform [prey wrapping silk (*15*)] glands. In addition to the highly internally repetitive modules, fibroins contain a conserved, nonrepetitive carboxy-terminal region. The divergence of the internal repetitive sequences that characterize each fibroin has been suggested to be responsible, in part, for the differing mechanical properties of the spun silk fibers produced by the different glands (*1*). Despite the divergence within the repetitive regions of the fibroin paralogues, some common molecular features are observed within fibroins, including four amino acid motifs that are found in various combinations and amounts (*11*, *16*). The four motifs include (1) polyalanine (A_n) stretches, (2) alternating glycine and alanine couplets [$(GA)_n$], (3) a stretch of three amino acids composed of two glycines followed by a variable amino acid [$(GGX)_n$], and (4) glycine-proline-glycine modules ($GPGX_n$). The modules described above are assembled in different numbers and combinations to form larger ensemble repeats that are iterated

† Supported by NSF CRUI Grant DBI-0112165 (Molecular Mechanisms of Mechanical Diversity in Spider Silks).

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

‡ Department of Chemistry, University of the Pacific.

§ Eastern Illinois University.

|| Department of Biology, University of the Pacific.

⊥ University of California.

10.1021/bi050494i CCC: $30.25
Published on Web 07/08/2005

many times throughout the internal region of the fibroins. To date, nucleic acid or protein sequences have not been described for aggregate (sticky glue) and pyriform (attachment disc) silks. Furthermore, only one partial cDNA sequence has been reported from *Araneus diadematus* that shows tubuliform-restricted patterns of expression named ADF-2[1] (*9*). Because morphological evidence supports the role of the tubuliform gland in egg case silk production, the tubuliform-restricted mRNA expression pattern of ADF-2 suggests this product may represent a component of araneoid egg case silk. However, published amino acid compositions of tubuliform silks do not correspond well with the predicted amino acid composition of the glycine-rich ADF-2 (*9*, *17*, *18*). The question of whether ADF-2 represents a main constituent of araneoid egg case core fibroins or whether, perhaps, the tubuliform silks contain other core fibroins that have not been described then arises. Our studies identify a novel cDNA that encodes a fibroin that shows tubuliform-restricted patterns of expression with predicted amino acid compositions that closely match reported amino acid compositions of araneoid tubuliform silks. These data suggest that this fibroin is a major constituent of egg case silk and reveal a novel type of consensus ensemble repeat unit within the primary sequence of the fibroin.

EXPERIMENTAL PROCEDURES

De Novo Sequencing of Egg Case Silk Peptides. Egg case silk was collected from adult female black widows spiders captured in the wild in San Joaquin County, CA. Egg cases were boiled in 8 M guanidinium chloride (GdnHCl) (3 mg/mL of solution) for 10 min to remove the coating proteins. Removal of the coating proteins was monitored by SDS−PAGE analysis and silver staining and was determined to be complete when fresh GdnHCl failed to solubilize additional proteins. After removal of the coating proteins, the insoluble core fibroin material was washed with 10 mL of water three times and then dissolved in hexafluoro-2-propanol (HFIP). Two hundred microliters of this solution was dried in a vacuum centrifuge, which left a protein thin film on the sides of the tube. This thin film was boiled in an 8 M GdnHCl/10 mM DTT solution for 30−60 min to facilitate fibroin solubilization and hydrolysis. After boiling, the viscous solution was diluted with 50 mM NH_4HCO_3 (pH 7.5) to reduce the GdnHCl salt concentration to less than 1 M. Ten micrograms of trypsin gold (Promega) was added and the digestion carried out at 37 °C overnight. The peptides that were generated were desalted using Omix C18 tips (Varian), and the monoisotopic masses of the protonated peptides were measured using a MALDI/TOF mass spectrometer (4700 Proteomics Analyzer, Applied Biosystems). The most abundant peptides were selected to undergo MS/MS after fragmentation by high-energy collision-induced dissociation (CID) to determine amino acid sequences. CID was performed with a (lab frame) collision energy of 1 kV. MS/MS spectra were analyzed manually to determine the peptide primary sequences.

[1] Abbreviations: GdnHCl, guanidinium hydrochloride; HFIP, hexafluoro-2-propanol; MaSp1, major ampullate spidroin 1; TuSp1, tubuliform spidroin 1; ADF-2, *Araneus diadematus* factor 2; MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; MS, mass spectrometry; HCl, hydrochloric acid; BLAST, Basic Local Alignment Search Tool; CID, collision-induced dissociation.

Amino Acid Analysis. The coating proteins were removed as described above, and the insoluble core silk fibroin was washed with water, dried, and then subjected to amino acid analysis. The core silk fibers were subjected to amino acid analysis at the Protein Chemistry Laboratory of Texas A&M University (College Station, TX). Briefly, vapor-phase hydrolysis of the core proteins by 6 M HCl was employed to generate the constituent amino acids. Amino acids that were obtained were derivatized with *o*-phthalaldehyde and 9-fluoromethylchloroformate. The derivatized amino acids were separated by reversed-phase high-performance liquid chromatography (HPLC) with UV detection.

Retrieval of the TuSp-1 cDNA. The peptide sequence APAFADAVSQAVR was used to synthesize degenerate oligonucleotides corresponding to the underlined region. PCR mixtures containing the forward primer pGAL4-AD (anchor library primer), 5′-AGGGATGTTTAATACCACTAC-3′, and the reverse primer, 5′-WGCRTCWGCRAAWGCWGGWGC-3′ (R = A or G; W = A or T), resulted in the amplification of a 1.6 kb DNA fragment from a cDNA library produced from black widow silk glands. This fragment was used to screen the cDNA library for larger TuSp1 inserts by conventional nucleic acid−nucleic acid hybridization. Positive clones were verified by Southern blot analysis. The largest TuSp1 cDNA (~3.3 kb) was sequenced using the GPS-1 system according to the manufacturer's instructions (New England Biolabs).

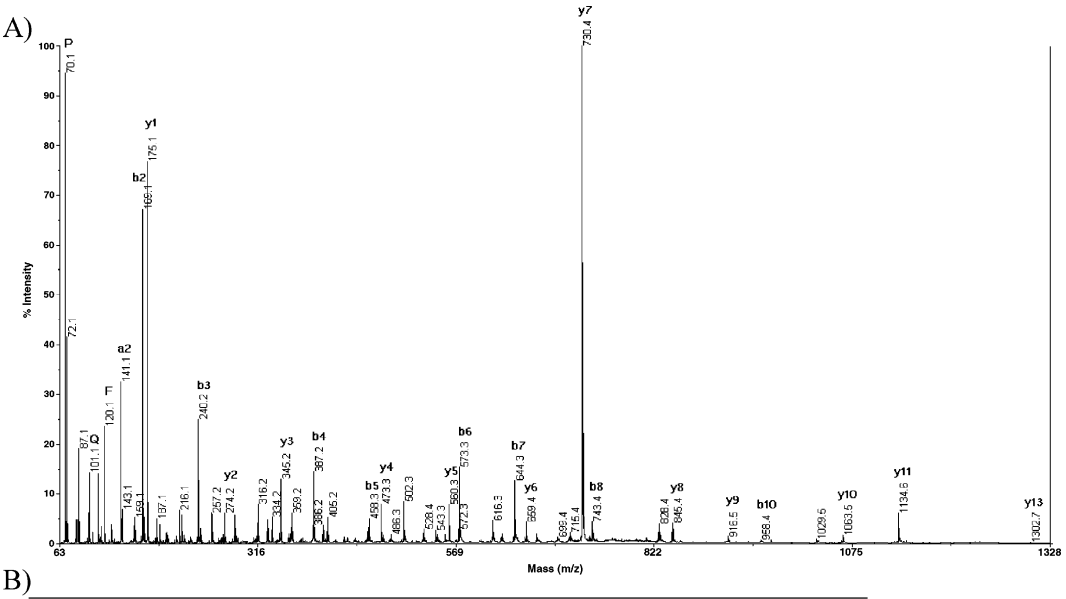
Real-Time PCR Analysis. Reverse transcription reactions were used for real-time PCR analysis using the DyNAmo SYBR Green qPCR kit according to the manufacturer's instructions. Real-time PCR fluorescence detection was monitored using an Opticon II instrument (MJ Research Inc.). Amplification products were monitored by SYBR Green detection and routinely checked using dissociation curve software and agarose gel electrophoresis. Oligonucleotides used for the analysis of TuSp1 were the forward and reverse primers 5′-TGTTGGTGTCGGAGCAAGTC-3′ and 5′-GCAGCGGAAGAAATTGCACTAG-3′, respectively.

Acquisition of NMR Spectra. ^{13}C CP-MAS spectra of 25 mg of solid silk were obtained on a CMX400 Chemagnetics spectrometer with a solid sample probe operating at 100.624 MHz. The sample was spun at the magic angle at a rate of 5.5 kHz. High-power decoupling was used, and a contact time of 2.5 s was chosen to match the Hartmann−Hahn condition. The number of acquisitions was 40 000. Chemical shifts were set using an external adamantane standard and referenced to TMS. All spectra were analyzed using an NMR analysis program (NUTS) from AcornNMR. Relative peak areas were estimated using the linefitting module in NUTS.

RESULTS

Isolation of Core Egg Case Silk Peptides and Sequence Analysis. No egg case silk fibroins with amino acid compositions matching published amino acid compositions of araneoid tubuliform silks have previously been described. Because *Latrodectus* and *Araneus* have been shown to have similar amino acid compositions for egg case silk and black widow spiders are indigenous to California, we selected this species for egg case silk studies (*18*, *19*). Egg cases from *Latrodectus* were initially treated with the strong denaturant GdnHCl to remove the coating proteins. Complete removal

A)



B)

M+H$^+$	*De novo* sequence
1302.7	APAFADAVSQAVR*
1384.7	LAALPYSTEYTR
2158.1	AAQTLGISNAPAFADAVSQAVR*
2467.2	TVGVGASPFQYANAVSNAFGQLLR*
4022.6	SVSNADQAANLAYNLGLNLAPSLGLNPSEALASSLR

FIGURE 1: Mass spectrometric analysis of peptides from egg case silk fibroins. (A) High-energy collision-activated dissociation (CID) MS/MS spectrum of fragments of the precursor ion at *m*/*z* 1302.7. On the basis of this spectrum, the sequence of this peptide was found to be APAFADAVSQAVR. (B) Peptide sequences obtained from hydrolysis and/or tryptic digestion of the core egg case silk components, followed by MALDI tandem TOF mass spectrometric analysis of the peptides. The asterisks denote peptides found within the retrieved cDNA sequence of TuSp1 after translation (Figure 2). The sequence with *m*/*z* 4022.6 represents a partial sequence because of the loss of large fragments during CID.

of the coating proteins was verified by SDS−PAGE analysis and silver staining (data not shown). After the coating proteins had been stripped off, the insoluble egg case silk fibroins were dissolved in HFIP to disrupt fibroin secondary structure and facilitate digestion with trypsin. Peptides with precursor ion masses (MH$^+$, monoisotopic) of 1302.7, 1384.7, 2158.1, 2467.2, and 4022.6 were sequenced, and the product ion mass spectrum for the peptide with a mass of 1302.7 (MH$^+$, monoisotopic) is shown (Figure 1A,B). Protein BLAST searches of the NCBInr database with the deduced peptide sequences revealed no significant similarities to any reported protein.

Cloning of a Novel Egg Case Fibroin cDNA. Reverse genetics was used to retrieve cDNA(s) encoding egg case silk fibroin(s) using the peptide sequences obtained by CID. Because silk fibroin mRNAs have been reported to have codon biases at the wobble position, degenerate oligonucleotides were designed to accommodate this feature (*20*). Using a composite cDNA library prepared from black widow silk glands as a template, anchored PCR led to the amplification of a DNA ladder. Sequence analysis of the largest amplified cDNA product (∼1.6 kb) identified an open reading frame with sequences matching peptide sequences obtained from the MS/MS analysis of the tryptic fragments from the core silk (compare Figure 1B with Figure 2). Ensemble repeats were apparent within the primary sequence of the protein, which led to the naming of the gene product tubuliform spidroin 1 (TuSp1). This nomenclature follows the convention of MaSp1 and MaSp2, where spidroin is a contraction of "spider fibroin" (*6*, *7*) and is consistent with its expression pattern (Figure 3). Because of the repetitive nature of silk genes, larger cDNAs are difficult to acquire using PCR. Thus, to obtain a longer cDNA, we used the 1.6 kb cDNA fragment to screen the cDNA library by conventional nucleic acid−nucleic acid hybridization. Many positive clones resulted from the screen, and 14 candidates were further characterized by restriction digestion and Southern blot analysis, leading to the identification of the clone carrying the largest cDNA of ∼3.3 kb (GenBank accession number DQ109035). Translation of the cDNA sequence revealed the presence of nearly homogeneous 184-amino acid repeats, which terminated with a 171-amino acid, nonrepetitive carboxy terminus (Figure 2). Protein−protein BLAST searches (BLASTp) of the NCBInr databases with the sequence showed similarity to fibroin 1 from *Euagrus chisoseus* and fibroin 3 from *Plectreurys tristis* (GenBank accession numbers AAK30600.1 and AAK30612.1, respectively). The *E* values for *E. chisoseus* fibroin 1 and *P. tristis* fibroin 3 were 2×10^{-8} and 9×10^{-4}, respectively. Ensemble repeats were observed within the primary sequence of TuSp1; however, the crystalline regions formed by long polyalanine stretches and GA couplet repeats were not found. Within the ensemble repeats, a complex mixture of serine and short polyalanine blocks that included a stretch of threonine, an amino acid that is reportedly rare in araneoid fibroins (*11*), was observed. Although the translated TuSp1 cDNA predicts a high serine content, no significant similarity was found relative to sericin or sericin-related proteins from the silkmoth *Bombyx mori*, as determined by BLASTp searches of the NCBInr protein database.

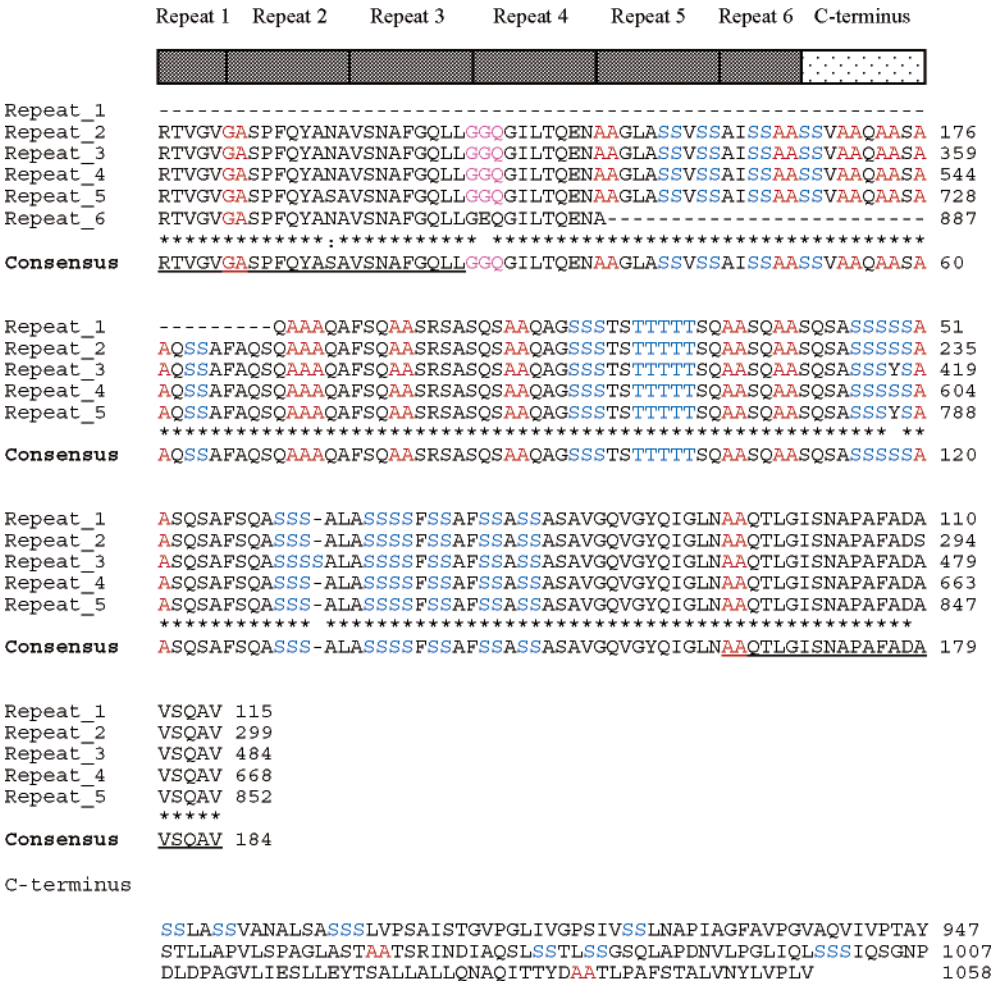


FIGURE 2: Molecular properties of TuSp1. Primary sequence of TuSp1 and the derived consensus repeat for egg case silk fibroin. Translation of the TuSp1 nucleotide sequence predicts a protein with 1058 amino acids. Amino acids are indicated by one-letter abbreviations, and gaps are represented by dashes. The conceptual translation product predicts a 100.5 kDa protein with an estimated pI of 4.45. No potential start codons were found within the TuSp1 ORF, which indicates the retrieved nucleic acid represents a partial cDNA sequence. The amino acid sequences of the six repeat units are aligned with each other. The first and sixth repeats are not complete. The asterisks signify matches to the majority-rule consensus sequence. The nonrepetitive carboxy terminus is shown below the repeat units. Peptide sequences determined by MALDI tandem TOF mass spectrometry are underlined in the derived consensus sequence. The subrepeat sequence motifs of GGX and A_n, which are abundant in other characterized araneoid fibroins (9, 11, and 15), are colored magenta and red, respectively. Other subrepeats which have been rarely reported in other fibroins include S_n and T_n and are colored blue.

Egg Case Silk Amino Acid Composition. The raw egg case silk amino acid compositions from *L. hesperus* corresponded well with the predicted values of TuSp1 (Table 1). Removal of the egg case silk coating proteins with GdnHCl also yielded similar percent amino acid compositions relative to those of raw silk (Table 1). These correspondences were consistent with TuSp1 representing a major constituent of egg case silk. Subtle differences in certain residues were noted; in particular, glycine and alanine levels were slightly higher in the raw egg case silk material relative to the predicted amino acid compositions from the TuSp1 primary sequence.

TuSp1 Shows Tubuliform-Restricted Patterns of Expression. The tubuliform glands have been reported to be present in only females, and the development of these glands parallels maturation of the ovaries (*21*). Thus, morphological and developmental studies suggest the tubuliform gland is chiefly responsible for egg case silk production. However, to prove this hypothesis, isolated fibroins must be established as constituents of raw egg case silk and their corresponding mRNAs (or fibroin proteins) shown to have tubuliform-restricted patterns of expression. Because our mass spectrometric data confirm that TuSp1 is present within the raw egg case silk, we next investigated whether TuSp1 mRNA was found within the tubuliform gland. We performed real-time quantitative PCR analysis to examine the mRNA expression profile of TuSp1. TuSp1 mRNA levels were the highest in the tubuliform gland, with virtually no detection in the major, minor, flagelliform, and aggregate glands or the ovaries (Figure 3). Lower levels were detected in the fat tissue, but were ~100 fold lower (Figure 3). Detection of TuSp1 mRNA in the fat likely reflects a low-level tubuliform gland contamination, as during fat tissue isolation we often pierce the tubuliform gland because of the anatomical positioning of the tissues. Northern blot analysis demonstrated the TuSp1 mRNA was >6 kb in size, which is consistent with reported fibroin mRNA sizes (data not shown).

Secondary Structure of Postspun Egg Case Silk. To understand the relationship between the repetitive units found within the primary sequence of TuSp1 and their secondary structure, a ^{13}C CP-MAS spectrum was acquired using raw

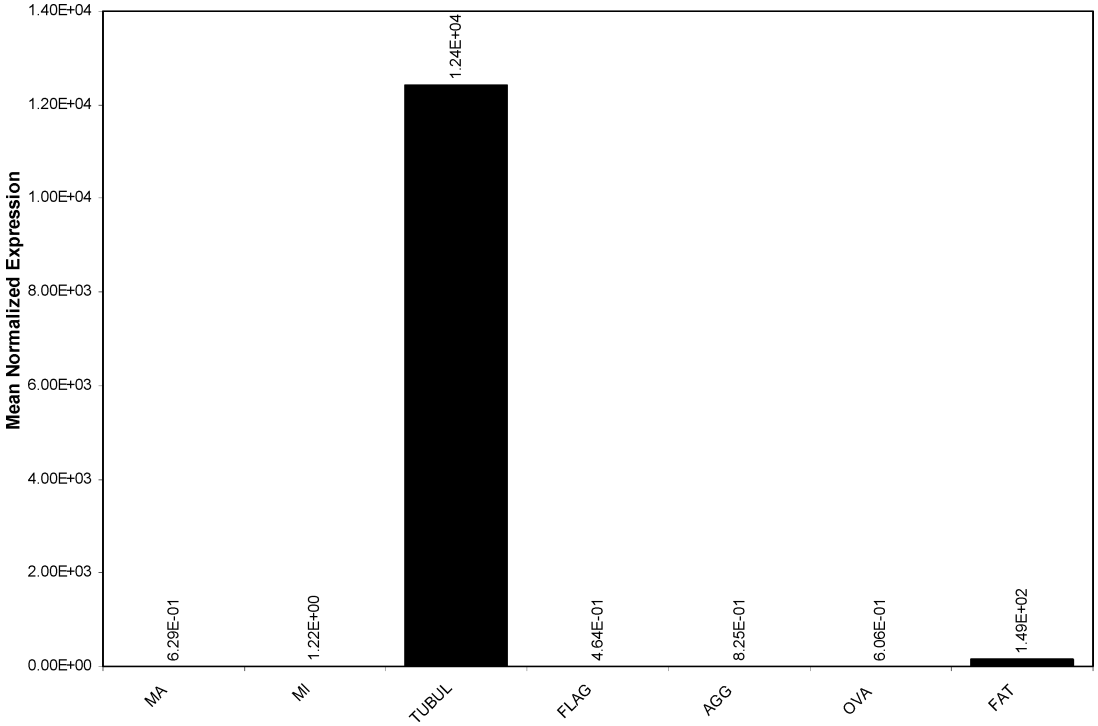


FIGURE 3: TuSp1 mRNA is tubuliform gland specific. Real-time quantitative PCR (RT-qPCR) was used to determine the expression pattern of TuSp1 in a variety of different tissues. Total RNA was isolated from the major ampullate gland, minor ampullate gland (MI), tubuliform (TUBUL), flagelliform (FLAG), aggregate (AGG), ovaries (OVA), and fat (FAT). Equivalent amounts of total RNA were reversed transcribed using MMLV and aliquots used for RT-qPCR. Samples were performed in triplicate and normalized internally using 18S rRNA. Data are representative of experimental results obtained from two independent trials.

L. hesperus egg case silk (Figure 4A). Peak assignments were made using published values for polypeptides (*22*, *23*) and spider silk (*24*, *25*). The relative area under the peaks (determined by deconvolution) corresponds roughly to the mole percent amino acids for the silk and for the predicted amino acid sequence. However, long ^{13}C relaxation times and differences in cross-polarization efficiency mean that integrated ^{13}C peak areas can be considered only as estimates.

It has been established that the ^{13}C chemical shifts of the α-carbons, β-carbons, and carbonyl carbons are dependent not on amino acid sequence but on the conformation of the protein backbone (*22*, *23*). This chemical shift dependence has allowed the use of NMR in determining the secondary structure of amino acids in silkworm silk (*26*), *Nephila* MA silk (*2*), and *L. hesperus* MA silk (*25*). If an amino acid is found in more than one specific conformation, the resulting chemical shift dispersion will cause line broadening and in some cases more than one peak will be observed. Alanine exhibits an especially large change in chemical shift with secondary structure. Figure 4B compares the chemical shifts for polyalanine in four different secondary structures with our results for *L. hesperus* egg case silk along with our previously published values for MA silk. In the egg case spectrum, the major components for alanine α- and β-carbons (peaks 1 and 5 in Figure 4A) have chemical shifts that are within experimental error of the chemical shifts of the polypeptide β-sheet. These results indicate the alanine residues in the silk are primarily in a β-sheet environment. However, the results of the spectral deconvolution reveal that shoulders on peaks 1 and 6 (labeled 1′ and 5′, respectively) make up roughly 20−30% of the total areas of peaks 1 and 1′ and peaks 5 and 5′, respectively. The chemical shifts of these shoulders lie between the shifts for α-helix and random coil of alanine and roughly match the chemical shifts for prespun MA silk inside the gland (*25*). We therefore conclude that although the alanine residues in egg case silk are primarily in a β-sheet environment, approximately 20−30% of the alanines are in a non-β-sheet, likely coiled, structure.

DISCUSSION

Thus far, little is known about the primary sequences and secondary structure of fibroins assembled into egg case silk from spiders. We have used a proteomic approach coupled with reverse genetics to isolate one major constituent of egg case silk. MALDI tandem TOF-MS analysis of peptides obtained after tryptic digestion of egg case fibroins revealed novel peptide sequences from araneoid spider egg case silk (Figure 1B). Peptides with protonated ion masses of 1302.7 and 2158.1 are likely generated through the combination of tryptic digestion and hydrolysis because analysis of the primary sequence of TuSpl failed to reveal sequences that were flanked by both arginine and lysine residues. In the translated cDNA, both peptides are predicted to contain the N-terminal residue asparagine. One could speculate that the asparagines promote in-chain peptide cleavage by losing a primary amine. This reaction is likely facilitated when the sample is treated with GdnHCl and boiled. Both aspartic acid and asparagine residues have been reported to represent hot spots for nonenzymatic degradation of proteins, which may provide one explanation for the generation of peptides that have resulted from both tryptic and nonenzymatic cleavage (*27*). Moreover, the production of peptides by this process is not entirely surprising given that raw egg case silk contains only ∼1% arginine, with little, if any, lysine detected by the amino acid composition analysis. The peptide with *m*/*z* 2467.2 nearly matched a stretch found within the consensus

Table 1: Predicted Amino Acid Composition of TuSp1, as Well as Experimentally Determined Compositions of Raw Egg Case Silk (untreated), GdnHCl-Washed Egg Case Silk, Dragline Silk of *L. hesperus*, Predicted ADF-2 (GenBank accession number U47854), and Raw Egg Case Silk from *A. diadematus* (*19*)[a]

	Gly	Ala	Glx	Tyr	Pro	Arg	Asx	Ser	Leu	Ile	Thr	Val	Phe
predicted TuSp1	6.4	24.7	11.1	1.7	2.3	1.0	3.8	25.7	5.8	2.8	5.3	5.5	3.9
core egg case silk of *L. hesperus* ($n = 3$)	7.3 ± 0.6	26.0 ± 0.2	10.8 ± 0.2	1.5 ± 0.1	1.6 ± 0.0	1.4 ± 0.2	4.4 ± 0.1	25.6 ± 0.7	5.0 ± 0.0	2.5 ± 0.0	5.1 ± 0.2	5.0 ± 0.1	4.1 ± 0.1
raw egg case silk of *L. hesperus* ($n = 3$)[b]	7.7 ± 0.6	23.3 ± 0.6	11.0 ± 0.0	1.7 ± 0.6	2.0 ± 0.6	1.3 ± 0.6	6.3 ± 1.1	25.7 ± 0.6	5.0 ± 0.0	2.0 ± 0.0	5.0 ± 0.0	na[c]	na[c]
dragline of *L. hesperus* ($n = 5$)[b]	42.0 ± 1.8	27.4 ± 0.2	11.8 ± 0.4	4.6 ± 0.5	2.8 ± 0.4	2.0 ± 0.0	1.6 ± 0.5	3.8 ± 0.4	1.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	na[c]	na[c]
predicted ADF-2	34.4	22.4	9.9	3.1	2.0	1.0	2.0	11.6	6.1	2.7	0.3	3.7	0.0
waw egg case silk of *A. diadematus*	8.6	24.4	8.2	1.0	0.6	1.5	6.3	27.6	5.7	1.7	3.4	na[c]	na[c]
% similarity of predicted TuSp1 to core egg case silk	87.7	95.0	97.3	88.2	69.6	71.4	86.4	99.6	86.2	89.3	96.2	90.9	95.1

[a] The data from *Latrodectus* core egg case silk are expressed as an average of multiple samples ($n = 3$). The percent amino acid composition for Cys and Trp could not be determined because of their destruction in the liquid HCl hydrolysis assay. Lys, Met, and His levels were found to be less than 0.2%. Published values for dragline and egg case silk of *L. hesperus* are presented for comparison. The values are expressed per 100 residues (mole percent). [b] Data from ref *18*. [c] Not available.

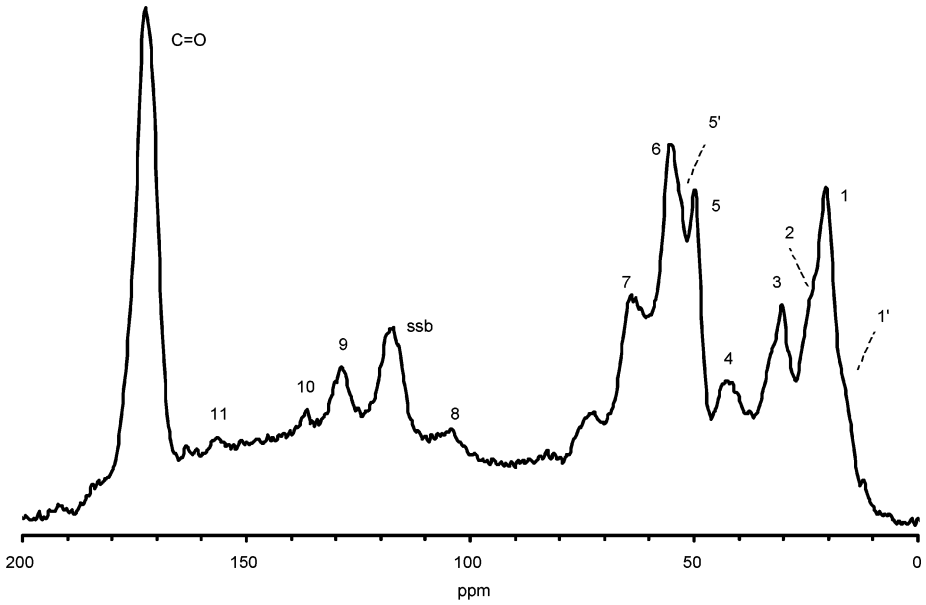
sequence of TuSp1, except for the arginine at the C-terminus (compare Figures 1B and 2). The lack of the arginine at this position may be explained by an allelic difference of TuSp1 in the spider population or, perhaps, an upstream module that has not been characterized. Two peptides (*m*/*z* 1384.7 and 4022.6) were not identified within the TuSp1 predicted sequence. The inability to find these peptide sequences might be because these sequences reside in the uncharacterized N-terminal region of TuSp1, or because other core fibroin protein(s) are present within egg case silk.

Analysis of the TuSp1 primary sequence reveals four nearly identical 184-amino acid ensemble repeats. Comparison of the TuSp1 primary sequence with other published fibroins demonstrates the highest degree of similarity to fibroin proteins from *Plectreurys* (Haplogynae) and *Euagrus* (Mygalomorphae), which are two non-araneoid spiders. Fibroins from *Plectreurys* and *Euagrus* have ensemble repeats unlike analogous units for other described silks, containing runs of serine and small numbers of GGX and $GPG(X)_n$ elements (*11*). TuSp1 units consist of subrepeats similar to these silks, with several short polyalanine runs, scattered polyserine blocks (ranging from 2 to 5 amino acid clusters), and one long polythreonine stretch (Figure 2). Within the consensus repeat, serine accounted for more than 28% of the residues. One could speculate that the similarity of TuSp1 to fibroins from *Euagrus* (fibroin 1; GenBank accession number AAK30600.1) and from *Plectreury* (fibroin 3; GenBank accession number AAK30612.1) could suggest that these non-araneoid fibroins have similar biological functions. On the basis of the hypothesized phylogenetic relationships of Araneae, which has been established using morphological evidence, *Euagrus* represents an evolutionarily ancient spider (*11*, *28*, *29*). The similarity between *Latrodectus* TuSp1 and *Euagrus* fibroin 1 may imply that the conservation of repetitive sequences from egg case silk fibroins of cob-weaving spiders has been maintained over millions of years of evolutionary history.

The predicted amino acid composition of TuSp1 closely resembles the residue composition of raw egg case silk and washed cases (after removal of coating proteins). Slightly smaller amounts of glycine and alanine were predicted for the TuSp1 cDNA sequence relative to values obtained from raw egg case silk. This is most likely due to an unidentified sequence toward the amino terminus of TuSp1 and/or additional proteins in the core silk fiber. Consistent with the predicted amino acid percentages for TuSp1, these values were also supported by the amino acid percentages determined by the NMR studies. Collectively, these results support the fact that TuSp1 represents a major element of egg case silk, and in postspun silk, alanine residues are predominantly involved in β-sheet structures. Thus, it would appear that the glycine-rich ADF-2, which has been detected in the tubuliform gland of *Araneus*, is unlikely to represent the major constituent of egg case silk in araenoid spiders because, if present in large amounts, predicted ADF-2 values would correspond more closely to the experimentally determined amino acid compositions of tubuliform silks (Table 1).

Because NMR spectroscopy shows that approximately 70% of alanines are found in β-sheet secondary structures, the question of which short polyalanine blocks or other scattered alanines within the primary sequence of TuSp1 contribute to this structure arises. Solving this question will

A)



B)

	α-helix[a]	random coil[a]	β-sheet[a]	3_1-helix[a]	MA silk[b]	EC silk[c]
Ala C_β	15.1	17.1	20.1	17.4	20.3	20.1/ 16.4
Ala C_α	52.5	50.6	48.7	48.7	48.7	49.0/ 52.3
Ala $C_{C=O}$	176.0	175.2	171.9	172.1	171.8	175.5

FIGURE 4: Solid-state NMR analysis of egg case silk from *L. hesperus*. (A) ^{13}C NMR spectrum of egg case silk from *L. hesperus*. Assignments are as follows: (1) Ala C_β, (2) Gln C_β, (3) Gln C_γ, (4) Gly C_α, (5) Ala C_α, (6) Ser C_α, (7) Ser C_β, (8) Tyr C_ϵ, (9) Tyr C_δ, (10) Tyr C_γ, and (11) Tyr C_ζ. The carbonyl carbons for all amino acids appear under one peak (labeled C═O). A spinning sideband in the solid-state spectrum is denoted by ssb. The broad background, which has a maximum around 125 ppm, is an artifact of the instrument. (B) Chemical shifts (in parts per million) for polyalanine carbons in different secondary structures and for *L. hesperus* egg case (EC) and major ampullate silk. Data from Wishart (*23*, *26*) for MA silk and Lawrence (*25*) for EC silk. Where two chemical shifts are given, the major component is given first.

require a comprehensive identification of all major components assembled into egg case silk as well as the N-terminal sequence of TuSp1. Such knowledge is likely to be required for delineation of the relationship between the lengths of polyalanine stretches, their assembly into crystallite structures, and their relevance to the material properties of different silk types.

The largest TuSp1 cDNA that was retrieved from our cDNA library screen was ~3.2 kb. Given the detection of >6 kb TuSp1 transcripts using Northern blot analysis, we estimate that our cDNA represents less than half of the nucleic acid sequence. However, because silk fibroins contain ensemble repeats, the predicted primary sequence and amino acid composition from the TuSp1 cDNA are likely to be representative of the full-length protein sequence. Further sequence information will await the retrieval of the genomic DNA sequence of the TuSp1 gene.

In this study, we have applied a new approach to retrieving spider fibroins. Mass spectrometry combined with reverse genetics was used to elucidate a major constituent of egg case silk. This product demonstrates tubuliform-restricted patterns of expression, and analysis of the predicted primary sequence of TuSp1 supports its classification as a spider fibroin. After comparing TuSp1 with reported spider silk protein sequences from different Araneae species, we conclude TuSp1 represents a highly divergent member of the araneid spidroin gene family, which is supported on the basis of the similarity to the more phylogenetically distant non-araenoid spider fibroins. Our approach should allow for the rapid identification of other silk fibroins, their confirmation to particular silk types, and the elucidation of minor constituents assembled into silks. This information will allow for a more in-depth understanding of the relationship between the primary sequences of these proteins and their contributions to the mechanical properties of silk.

ACKNOWLEDGMENT

We thank Brian L. Phillips for use of the W. M. Keck Solid-State NMR facility at the University of California, Davis, CA.

REFERENCES

1. Gosline, J. M., Guerette, P. A., Ortlepp, C. S., and Savage, K. N. (1999) The mechanical design of spider silks: From fibroin sequence to mechanical function, *J. Exp. Biol. 202* (Part 23), 3295−3303.
2. Simmons, A., Ray, E., and Jelinski, L. W. (1994) Solid-state ^{13}C NMR of *Nephila clavipes* dragline silk establishes structure and identity of crystalline regions, *Macromolecules 27*, 5235−5237.
3. Parkhe, A. D., Seeley, S. K., Gardner, K., Thompson, L., and Lewis, R. V. (1997) Structural studies of spider silk proteins in the fiber, *J. Mol. Recognit. 10*, 1−6.
4. van Beek, J. D., Hess, S., Vollrath, F., and Meier, B. H. (2002) The molecular structure of spider dragline silk: Folding and orientation of the protein backbone, *Proc. Natl. Acad. Sci. U.S.A. 99*, 10266−10271.

5. Knight, D. P., Knight, M. M., and Vollrath, F. (2000) β transition and stress-induced phase separation in the spinning of spider dragline silk, *Int. J. Biol. Macromol. 27*, 205−210.
6. Xu, M., and Lewis, R. V. (1990) Structure of a protein superfiber: Spider dragline silk, *Proc. Natl. Acad. Sci. U.S.A. 87*, 7120−7124.
7. Hinman, M. B., and Lewis, R. V. (1992) 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.
8. Beckwitt, R., and Arcidiacono, S. (1994) Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae), *J. Biol. Chem. 269*, 6661−6663.
9. Guerette, P. A., Ginzinger, D. G., Weber, B. H., and Gosline, J. M. (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family, *Science 272*, 112−115.
10. Beckwitt, R., Arcidiacono, S., and Stote, R. (1998) Evolution of repetitive proteins: Spider silks from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae), *Insect Biochem. Mol. Biol. 28*, 121−130.
11. Gatesy, J., Hayashi, C., Motriuk, D., Woods, J., and Lewis, R. (2001) Extreme diversity, conservation, and convergence of spider silk fibroin sequences, *Science 291*, 2603−2605.
12. Colgin, M. A., and Lewis, R. V. (1998) Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like "spacer regions", *Protein Sci. 7*, 667−672.
13. Hayashi, C. Y., and Lewis, R. V. (1998) Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks, *J. Mol. Biol. 275*, 773−784.
14. Becker, N., Oroudjev, E., Mutz, S., Cleveland, J. P., Hansma, P. K., Hayashi, C. Y., Makarov, D. E., and Hansma, H. G. (2003) Molecular nanosprings in spider capture-silk threads, *Nat. Mater. 2*, 278−283.
15. Hayashi, C. Y., Blackledge, T. A., and Lewis, R. V. (2004) 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.
16. Hayashi, C. Y., Shipley, N. H., and Lewis, R. V. (1999) Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins, *Int. J. Biol. Macromol. 24*, 271−275.
17. Vollrath, F., and Knight, D. P. (2001) Liquid crystalline spinning of spider silk, *Nature 410*, 541−548.
18. Casem, M. L., Turner, D., and Houchin, K. (1999) Protein and amino acid composition of silks from the cob weaver, *Latrodectus hesperus* (black widow), *Int. J. Biol. Macromol. 24*, 103−108.
19. Anderson, S. O. (1970) Amino Acid Composition of Spider Silks, *Comp. Biochem. Physiol. 35*, 705−711.
20. Hinman, M., Dong, Z., Xu, M., and Lewis, R. V. (1992) Spider silk: A mystery starting to unravel, *Results Probl. Cell Differ. 19*, 227−254.
21. Kovoor, J. (1987) Comparative structure and histochemistry of silk-producing organs in Arachnids, in *Ecobiology of Spiders* (Nentwig, W., Ed.) pp 159−186, Springer-Verlag, Berlin.
22. Saito, H. (1986) Conformation-dependent ^{13}C chemical shifts: A new means of conformational characterization as obtained by high-resolution solid-state ^{13}C NMR, *Magn. Reson. Chem. 24*, 835−852.
23. Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C chemical-shift index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data, *J. Biomol. NMR 4*, 171−180.
24. Yang, Z., Livvak, O., Seidel, A., LaVerde, G., Zax, D. B., and Jelinski, L. W. (2000) Supercontraction and backbone dynamics in spider silk: ^{13}C and ^{2}H NMR studies, *J. Am. Chem. Soc. 122*, 9019−9025.
25. Lawrence, B. A., Vierra, C. A., and Moore, A. M. (2004) Molecular and mechanical properties of major ampullate silk of the black widow spider, *Latrodectus hesperus*, *Biomacromolecules 5*, 689−695.
26. Saito, H., Iwanaga, Y., Tabeta, R., Narita, M., and Asakura, T. (1983) A High Resolution ^{13}C NMR Study of Silk Fibroin in Solid State by the Cross Polarization-Magic Angle Spinning Method: Conformational Characterization Utilizing Conformation-Dependent ^{13}C Chemical Shifts, *Chem. Lett.*, 427−430.
27. Geiger, T., and Clarke, S. (1987) Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation, *J. Biol. Chem. 262*, 785−794.
28. Coddington, J. A., and H. L. (1991) Systematics and evolution of spiders (Araneae), *Annu. Rev. Ecol. Syst. 22*, 565−592.
29. Hormiga, G., Scharff, N., and Coddington, J. A. (2000) The phylogenetic basis of sexual size dimorphism in orb-weaving spiders (Araneae, Orbiculariae), *Syst. Biol. 49*, 435−462.

BI050494I