Spider Egg Case Core Fibers: Trimeric Complexes Assembled from TuSp1, ECP-1, and ECP-2†

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

Department of Chemistry and Department of Biology, University of the Pacific, Stockton, California 95211, and Howard Hughes Medical Institute Mass Spectrometry Laboratory, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720

Received October 15, 2005; Revised Manuscript Received January 24, 2006

ABSTRACT: Spider silk proteins are well-known for their extraordinary mechanical properties, displaying remarkable strength and toughness. In this study, matrix-assisted laser desorption ionization (MALDI) tandem time-of-flight (TOF) mass spectrometry (MS/MS) and reverse genetics were used to isolate a new cDNA sequence that encodes for a protein assembled into egg case silk from the black widow spider, *Latrodectus hesperus*. Analysis of the primary sequence of this protein reveals ∼52% identity to the egg case protein 1 (ECP-1) fibroin-like family member. On the basis of the similarity in the primary sequence and expression pattern, we have named this factor egg case protein 2 (ECP-2). Alignments of ECP-1 and ECP-2 demonstrate highly conserved N termini, with 16 Cys residues found within the first 153 amino acids. Traditional ensemble repeats found within reported fibroins were poorly represented in the primary sequence of ECP-2, but scattered blocks of polyalanine were present, along with a C terminus rich in GA repeats. Reverse transcription quantitative PCR analysis showed that ECP-2 is predominantly expressed in the tubuliform gland. Relative to ECP-1, ECP-2 mRNA levels were determined to be >2-fold higher. MALDI MS/MS analysis of peptide fragments generated from the large-diameter core fiber after enzymatic digestion and acid hydrolysis demonstrated the presence of a fiber that is trimeric in nature, containing tubuliform spidroin 1 (TuSp1), ECP-1, and ECP-2. We also report an additional primary sequence for TuSp1, demonstrating that TuSp1 contains two Cys residues within a nonrepetitive N-terminal region. In combination with the distinctive protein architectures of ECP-1 and ECP-2, along with their co-localization with TuSp1 in the core fiber, our findings suggest that ECP-1 and ECP-2 play important structural roles in the egg case silk fiber.

Spiders use different combinations of silk proteins to manufacture structures for reproduction, hunting prey, and locomotion (*1*, *2*). Silk fibers are assembled from one or more silk fibroins that are synthesized in the abdominal glands, which are connected to the spinnerets of the spider (*3*). Araneoid spiders (ecribellate orb-weaver) have been shown to contain seven morphologically distinct silk-producing glands (*2*). Each distinct gland synthesizes different silk types designed for diverse biological functions (*3*, *4*). The superior properties of silks are derived from their chemical composition as a semicrystalline polymer that contains crystallite regions embedded in an amorphous matrix, which represent less organized regions. Much of the mechanical properties of the silk can be attributed to repetitive modules found within the primary sequence of the spidroins (contraction of spider and fibroin). During spidroin production, silk fibroins are secreted and stored as a liquid crystalline spinning dope (the material from which silk is spun). The spinning dope ultimately passes through the spinning duct and is converted into a solid thread, which is induced by extensional flow (*5*).

The primary sequences of spider fibroins share a number of distinguishing features. Repeats of four motifs have been observed in the majority of sequenced spider silk proteins: (1) polyalanine (A_n) stretches, (2) alternating glycine and alanine couplets $(GA)_n$, (3) three amino acid motifs 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 many times throughout the internal region of the fibroins (tandem arrayed throughout the sequence) (*4*, *6*). Published studies have shown that A_n and $(GA)_n$ repeats form crystalline β sheets and $GPGX_n$ iterations adopt β-spiral structures; different proportions of these elements have been proposed to control the mechanics of the fibers (*6*). In addition to their repetitive architecture (tandem arrays of iterated repeats), all spidroins have nonrepetitive C-terminal regions, which exhibit sequence conservation across silk paralogues (*3*, *4*, *7*, *8*).

† 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.

§ Current address: Guidant Corporation, Santa Clara, CA 95054.

|| Department of Biology, University of the Pacific.

⊥ University of California.

10.1021/bi052105q CCC: $33.50
Published on Web 02/21/2006

Spiders enclose their eggs inside protective cocoons to facilitate development. These silk fibers are manufactured primarily by the tubuliform (often referred to as the cylindrical) silk glands. Tubuliform silks have been proposed to have specialized roles, participating exclusively in egg case construction (*9*, *10*). Production of egg case silk is limited to female spiders, with the tubuliform silk gene expression being initiated in the glands at sexual maturity (*9*, *11*). Egg case silk has been shown to contain a complex mixture of different proteins. Two components of spider egg cases have recently been characterized at the molecular level. One main element, tubuliform spidroin 1 (TuSp1),[1] has been identified in egg cases across a wide range of different spider species, with a theoretical amino acid composition that aligns closely to the compositions of raw egg case silk (*12*−*14*). Mass spectrometric experiments have demonstrated that TuSp1 is a component of black widow spider egg cases, while reverse transcription quantitative PCR analysis (RT-qPCR) has firmly established a tubuliform-restricted expression pattern for TuSp1 (*14*). Another constituent that has been identified in egg cases is the egg case protein 1 (ECP-1) (*15*). TuSp1 and ECP-1 are both abundant proteins found in black widow egg cases. Previously, we have demonstrated that black widow egg cases contain two differently sized diameter fibers, a large (4−5 μm) and small (500 nm) diameter filament (*15*). The larger diameter filaments have been shown to constitute the bulk of the egg case. Despite the recent identification of TuSp1 and ECP-1, little is known regarding the molecular constituents that assemble to form the core fibroin of the large-diameter filament of egg case silk. Moreover, the discovery of the major ampullate spidroins MaSp1 (ADF-2 orthologue in *Araneus diadematus*) and MaSp2 mRNAs in the tubuliform gland, which are dragline silk components, suggest major ampullate spidroins could be constituents of tubuliform silks (*3*, *13*).

Here, we report the cDNA sequence of a spider silk gene that encodes for a protein product that represents a constituent of tubuliform core fiber silk. The egg case protein 2 (ECP-2) shares similar architecture to the distinctive ECP-1 fibroin-like molecule. We demonstrate that ECP-2 is highly expressed in the tubuliform gland. Analysis of the primary sequence of ECP-2 reveals two of the motifs characteristic to the majority of sequenced spider silks (repetitive GA couplets and scattered polyalanine blocks). Removal of the coating proteins and small-diameter fibers from egg cases using chemical treatment, followed by the analysis of the large-diameter core fibroin material after enzymatic digestion and acid hydrolysis using matrix-assisted laser desorption ionization (MALDI) tandem time-of-flight (TOF) analysis, shows that TuSp1, ECP-1, and ECP-2 are the main constituents of the large-diameter filament of egg cases. Furthermore, the cloning of the genomic DNA sequence corresponding to the N terminus of TuSp1 reveals the presence of two cysteine residues within a nonrepetitive architecture, which was not previously discovered. In combination with the Cys-rich N termini of ECP-1 and ECP-2 and their assembly with TuSp1 in the core fiber, this result is supportive of a functional role of ECP members as molecular cross-linkers interweaved into a trimeric egg case core fiber complex.

[1] Abbreviations: GdnHCl, guanidinium hydrochloride; HFIP, 1,1,1,3,3,3-hexafluoro-2-propanol; MaSp1, major ampullate spidroin 1; MiSp1, minor ampullate spidroin 1; ECP-1, egg case protein 1; ECP-2, egg case protein 2; TuSp1, tubuliform spidroin 1; ADF-2, *Araneus diadematus* factor 2; MALDI−TOF, matrix-assisted laser desorption/ionization time-of-flight; MS, mass spectrometry; BLAST, basic local alignment search tool; CAD, collision-activated dissociation.

EXPERIMENTAL PROCEDURES

Collection of Spiders, Egg Cases, and Gland Isolation. Approximately 15 adult female black widows (*Latrodectus hesperus*) were collected in San Joaquin County, CA. Spiders were kept separately in 22 × 40 × 30 cm aquarium tanks with a 14 × 18 × 25 cm wooden frame inserted to aid web construction and deposition of the egg cases. Spiders were fed one large cricket weekly. Fresh egg cases were collected within 24 h of being deposited, and individual egg cases were cut open to remove the eggs. Egg cases were discarded if any eggs were broken during the isolation procedure. Silk from each egg case was extracted with 8 M guanidinium hydrochloride (GdnHCl) (3 mL of solution/mg of silk) for 10 min with agitation. The supernatant, which had a viscous nature, was removed and subjected to dialysis. The solid residue, which failed to dissolve, was processed as described in the analysis of large-diameter core fiber material in the Experimental Procedures. Spider glands used for the quantification of mRNA levels were obtained by euthanizing the spiders with carbon dioxide, followed by the removal of the glands by microdissection with the aid of a dissecting microscope. During our isolation procedures, we routinely observed two different colors of the tubuliform gland. Some tubuliform glands were uniformly white, whereas others had a white-orange appearance. We classified those glands that contained no visible color as white and those with a combination of colors (white and orange) as orange.

In-Gel Trypsin Digestion of Egg Case Silk Proteins from the Black Widow Spider. Trypsin (20 μg/mL) was prepared according to the instructions of the manufacturer (Tryspin gold, Promega). Proteins dissolved in 8 M GdnHCl from egg cases were size-fractionated on a 4−20% polyacrylamide-gradient gel. Polypeptides sized at ~100 kDa were excised from the gel using a sterile scalpel. Isolated proteins were digested in-gel following a previously published protocol (*16*). Briefly, the gel pieces were destained using the solutions provided in the ProteoSilver plus Silver Staining Kit (Sigma). After destaining, the gel slices were cut into ~1 mm^3 cubes using a precleaned scalpel. Gel pieces from two samples were transferred into prewashed 0.6 mL centrifuge tubes and dehydrated with 25 mM NH_4HCO_3/acetonitrile (ACN). The dried gel particles were rehydrated and reduced with 10 mM DTT, alkylated with 55 mM iodoactamide, dried again, and rehydrated and digested overnight in 250 ng of trypsin. After digestion, peptides were extracted twice with 30 μL of 50% acetonitrile/5% formic acid solution. The combined extracts were concentrated to ~10 μL, and the peptides were desalted using C18 Zip-Tips. One piece of blank gel was processed using the same procedure to serve as a negative control.

Analysis of Insoluble Material from the Large-Diameter Core Fiber. For the core fiber analysis, the egg cases were treated with 8 M GdnHCl as described above to remove the coating proteins and small-diameter fibers. Material that failed to dissolve after initial treatment was defined as egg

case core material (large diameter). To ensure no residual coating proteins survived the initial 8 M GdnHCl wash, as well as smaller diameter constituents, we returned the insoluble fiber to fresh 8 M GdnHCl, boiled the material, and washed the silk with water. After the removal of the water, the sample was dehydrated and dissolved in 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP). After HFIP treatment, the solvent was evaporated and the sample was returned to 8 M GdnHCl, boiled, and then subjected to tryptic digestion. HFIP is routinely used to dissolve core fibroins and can efficiently disrupt secondary-structure β-sheet structure. Trypsin digestions were performed after the 8 M GdnHCl was diluted to 1 M by using 50 mM NH_4HCO_3 (pH 7.8). Peptides were desalted using C18 OMIX tip (Varian).

De Novo Peptide Sequencing Using MALDI Tandem TOF Mass Spectrometry (MS/MS). Two digests obtained from the first and second GdnHCl treatments were analyzed using a MALDI tandem TOF mass spectrometer (4700 Proteomics Analyzer, Applied Biosystems). The matrix used was α-cyano-4-hydroxycinnamic acid in 50% (v/v) ACN at a concentration of 10 mg/mL. A total of 0.5 μL of the digests was mixed with an equal volume of matrix solution and spotted onto the MALDI target plate. The monoisotopic masses of the peptides generated by trypsin digestion were examined in positive full scan mode. Target peptides were selected to undergo high-energy collision-activated dissociation (CAD) to determine their primary amino acid sequences. The resulting MS/MS spectra were interpreted manually to obtain the *de novo* peptide sequences.

Cloning of the ecp-2 Gene. The peptide sequence DGQLDFQGLR was used to design reverse degenerate oligonucleotides (the underlined region corresponds to the primer design area). These degenerate primers where used in combination with the forward primer (anchor) from the pGAL4-AD library vector. PCR reactions containing the forward primer (5′-CGATGATGAAGATACCCCACC-3′), together with the reverse primer 5′-CCTTGRAAGTCTAATTGWCC-3′ (encodes for GQLDFQG; R = A or G; W = A or T), successfully amplified a 540-bp ECP-2 from a cDNA library produced from the silk glands of black widow spiders (corresponding to ECP-2 amino acid sequence 1−180, Figure 3A). The reverse primer (anchor) from the pGAL4-AD library vector 5′-GCACAGTTGAAGTGAACTTGC-3′ was used with the gene-specific forward primer 5′-CGATGCACAAAAGTAAGAATACC-3′ (corresponding to amino acids 152−158) to amplify the 3′ end of the ECP-2 cDNA. This primer sequence is not found within the ECP-1 sequence. One cDNA fragment of size ∼1.8 kb was successfully obtained and sequenced. The two overlapping cDNA pieces were joined together to give the unified cDNA sequence of ECP-2. The primary ECP-2 sequence was obtained by translation.

RT-qPCR Analysis. Reverse transcription reactions were used for RT-qPCR analysis using the DyNAmo SYBR Green qPCR kit according to the instructions of the manufacturer. RT-qPCR 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 comparative analysis of the ECPs and TuSp1 were as follows: ECP-1 forward and reverse primers GAATCCAGTAGTGCCTCCCAATT (nucleotides 1110−1132) and TTGTGAACTCTCCTCCTTGACT (nucleotides 1293−1314), respectively; ECP-2 forward and reverse primers 5′-CGAAGTGGCAGAATTTCAACATCTG-3′ (nucleotides 1558−1582) and 5′-GAATTGATTCCACCGCCTTGAGTG-3′ (nucleotides 1743−1766), respectively; TuSp1 forward and reverse primers 5′-TGTTGGTGTCGGAGCAAGTC-3′ and 5′-GCAGCGGAAGAAATTGCACTAG-3′, respectively. Because there is a weak similarity with respect to certain regions in the nucleic acid sequences of *ecp-1* and *ecp-2*, we designed primer sets to unique regions to ensure the exclusive amplification of either *ecp-1* or *ecp-2*. For TuSp1, the primer set was designed to amplify a segment of a single module found within TuSp1 molecules. On the basis of the analysis of TuSp1 cDNA sequences, the modules are ∼552 bp (*14*). Because the average transcript size for TuSp1 has been shown to be ∼13 kb for *Nephila clavipes* (*12*), the number of modules repeated within a single TuSp1 molecule is estimated to be ∼23. Thus, for accurate expression values, the RT-qPCR data require an adjustment to reflect this property because measurements in this assay normally assume a single target site for each primer set/molecule. Design of the primer set to the repeat modules for traditional silks is currently the most efficient method for RT-qPCR analysis, because other silk paralogues have been shown to have conserved nonrepetitive C termini and presumably N termini.

Genomic DNA Walking Kit. Two reverse primers were constructed that contained the nucleotide sequences TuSp-1 R1 5′-AACAGCATTTGCATATTGGAA-3′ (represents the outer reverse oligonucleotide corresponding to the amino acid sequence FQYANAVA) and TuSp-1 R2 5′-AGCTGGGGCATTGCTGATTCC-3′ (inner reverse primer corresponding to the sequence GINSNAPA). Genomic DNA was isolated from the black widow spider using the DNeasy tissue kit according to the instructions of the supplier (Qiagen). The genomic DNA library used for the cloning of the 5′ end of the TuSp-1 was constructed according to the instructions of the manufacturer (Clontech). The secondary reaction from the *Eco*RV library successfully led to the amplification of a ∼2.2-kb TuSp-1 promoter fragment. This fragment was subject to DNA sequencing as previously described (*17*).

RESULTS

Identification of Abundant Proteins in Egg Cases. As previously reported, the analysis of the egg case extracts by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE) followed by silver staining showed a broad distribution of molecular weights, ranging from approximately 10 to >300 kDa (lane 2 in Figure 1) (*15*, *18*). The broad distribution of protein molecular weights likely results from protein degradation and/or the presence of glue proteins. One of the most prominent protein bands observed in dissolved egg case silk is a 100-kDa protein doublet. The primary sequence of one of the protein doublet species was recently determined, which led to the identification of a new class of fibroins called ECP-1 (*15*). However, the identity of the second band has yet to be elucidated. To determine the primary sequence of the other protein in this doublet, we excised the doublet, digested (both parts) with trypsin, and analyzed the peptides by MALDI MS. Many peptide ions were observed after tryptic digestion (Figure 2A).

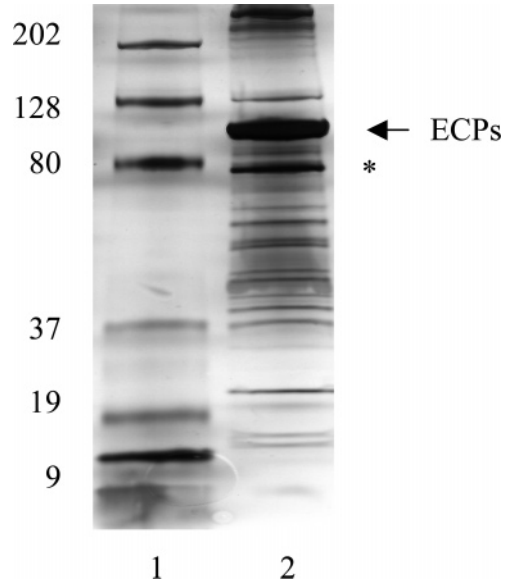


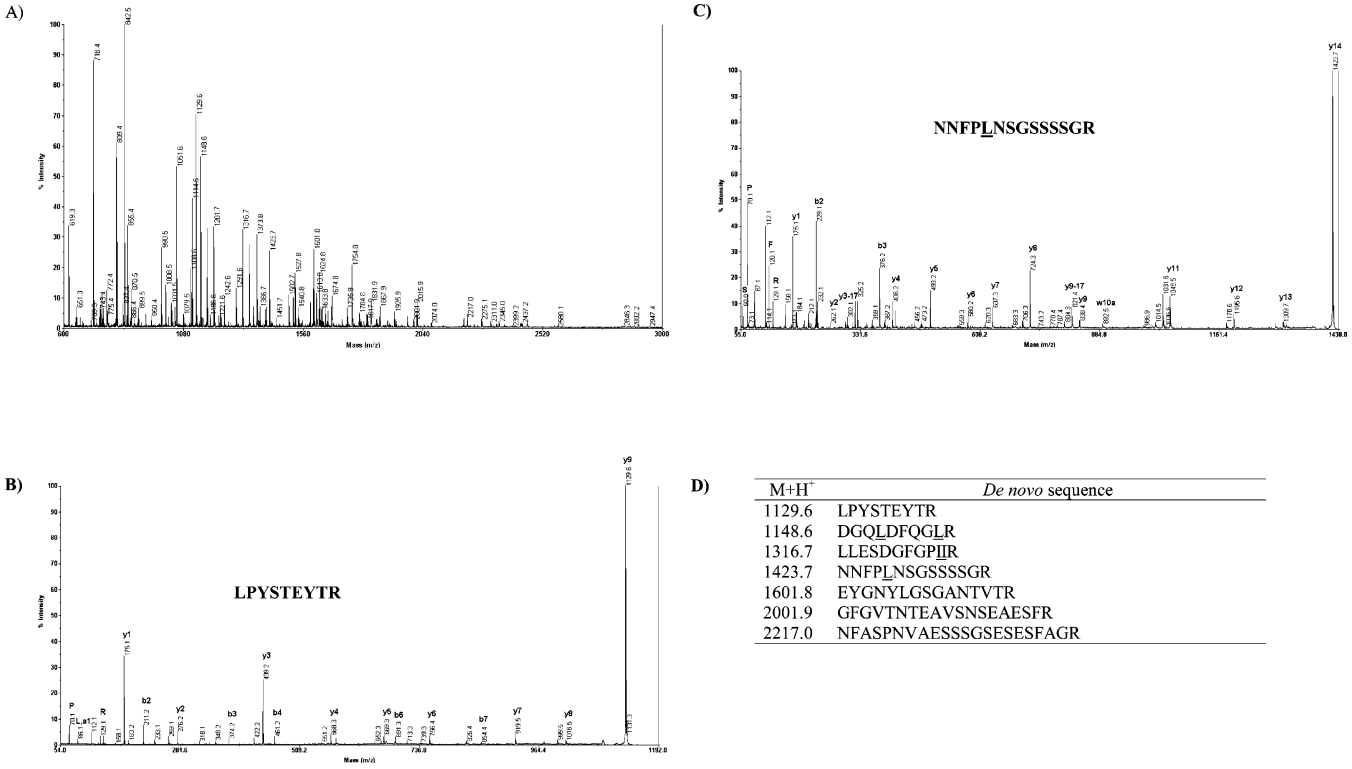
FIGURE 1: Analysis of proteins isolated from the egg case silk fiber of black widow spiders. Lane 1, molecular markers with sizes indicated in kilodaltons; lane 2, egg cases were treated with 8 M GdnHCl, and the solubilized proteins were resolved using SDS−PAGE analysis (4−20% gradient gels), followed by protein detection using silver staining. Extraction of TuSp1 by GdnHCl treatment is largely inefficient (relative to HFIP treatments), and the larger molecular-weight protein bands migrating slower than 200 kDa do not represent TuSp1 molecules or ECPs, as determined by in-gel tryptic digestion and MALDI tandem TOF analysis (data not shown). Most likely, these bands are derived from the small-diameter fiber of egg case silk. Excision of another major protein band migrating at ~80 kDa (*), followed by in-gel trypsin digestion and MALDI MS analysis, produced a peptide fingerprint that overlapped with peptide ions found by MALDI MS analysis of the 100-kDa protein doublet (data not shown).

Peptide ions with experimental masses that matched predicted peptide masses obtained from ECP-1 were excluded from further MS/MS analysis. Seven peptides were sequenced using high-energy CAD, those with precursor ion masses (MH^+, monisotopic) of 1129.6, 1148.6, 1316.7, 1423.7, 1601.8, 2001.9, and 2217.0 (Figure 2D). The product ion spectra from the peptides of MH^+ 1129.6 and 1423.7 are shown in parts B and C of Figure 2, respectively. Analysis of the derived peptide sequences using the algorithm basic local alignment search tool (BLAST) revealed no significant similarity to any polypeptides in the NCBInr protein database for peptide ions 1129.6, 1148.6, 1316.7, 1423.7, and 1601.8. Peptide ions 2001.9 and 2217.0 showed similarity at the primary sequence level to ECP-1, with *E* values of 7×10^{-6} (identities 15/18 = 83%) and 0.011 (identities 15/21 = 71%), respectively.

Isolation of the ecp-2 Gene. A cDNA library prepared from black widow spider silk glands was screened by anchored PCR. Reverse degenerate oligonucleotides that corresponded to a region of the peptide sequence DGQLDFQGLR (underlined region) obtained by MS/MS analysis (Figure 2D) were designed and used in PCR along with the forward primer (anchor) from the pGAL4-AD library vector. This primer combination led to the successful amplification of a single 540-bp cDNA fragment [amino acid residues corresponded to 1−180 (Figure 3A)]. Using the nucleotide sequence data obtained from the 540-bp gene piece, we amplified the remaining end of the cDNA using a modified form of rapid amplification of cDNA ends (3′ RACE). One overlapping cDNA fragment was obtained from the 3′ RACE, which led to the reconstruction of a ~2.6-kb cDNA fragment that contained a long open-reading frame (ORF; Figure 3A; GenBank accession number DQ341220). In protein−protein BLAST searches against the translated ORF, the top match corresponded to ECP-1 (52% identities with an *E* value = 7.0×10^{-128}; GenBank accession number AY994149). No other significant similarity was observed to other proteins in the databases using this algorithm. On the basis of the similarity to ECP-1, we have named this gene *ecp-2*. The region of ECP-2 that showed the strongest similarity to ECP-1 corresponded to the N-terminal 153 amino acids. A total of 16 Cys residues were found to be conserved in the N termini of the ECP products (Figure 3B). As expected for secreted proteins, the primary sequence of ECP-2 contains an N-terminal signal sequence, with the predicted cleavage site between residues 29 and 30 (*19*). The likely function of the putative signal sequence is to allow for the secretion of ECP-2 into the glandular lumen of tubuliform tissue. Because signal sequences for proteins directed to the endoplasmic reticulum are located on the N termini of secreted proteins, this observation supports our assertion that the primary sequence of ECP-2 is complete (Figure 3A). Members of the spider silk protein family have been shown to contain distinct, nonrepetitive C-terminal regions. The C-terminal region is conserved in sequence and length across phylogenetically distant species and among diverse silk types. Analysis of the C-terminal region of ECP-2 (last 170 residues) showed no similarity to the nonrepetitive C terminus of other spider fibroins. However, despite the lack of conservation to the C-terminal nonrepetitive regions of other silk paralogues, the C-terminal region of ECP-2 contained fibroin-like GA repeats (Figure 3A); these repeats are commonly found within internal iterated repeats of fibroins, including the fibroin heavy chain (Fib-H) of the silkmoth *Bombyx mori* (20), minor ampullate fibroin 1 (MiSp1) (*21*), and MaSp1 (*22*).

Six peptides sequenced by MS/MS (*m*/*z* 1129.6, 1148.6, 1423.7, 1601.8, 2001.9, and 2216.9) showed 100% identity to translated regions of the ECP-2 cDNA, with the exception that, in some cases, the *de novo* sequences showed I instead of L or vice versa. It is sometimes difficult to distinguish these two isobaric amino acid residues by MS/MS because they have identical masses (compare Figure 2D and Figure 3A). One peptide ion (*m*/*z* 1316.7) with the sequence LLESDGFGPIIR was nearly identical to a region found within the translated ECP-2 sequence at location 509−520 (compare Figure 2D with Figure 3A). The only difference is the substitution of Asp for Asn. The conversion of N to D could have resulted from acid-catalyzed deamidation during the sample preparation process, because this event is particularly favored when Gly residues are C-terminal to the Asn that is observed to be deamidated (*23*). Further evidence to buttress this assertion comes from the analysis of an expanded view of the peptide MS near *m*/*z* 1316.7, which shows a lower intensity peptide ion with *m*/*z* 1315.7 (data not shown). This latter ion likely represents the original peptide LLESNGFGPIIR that was not deamidated. Asn to Asp transitions have often been observed in other proteins as a result of acid treatment (*24*).

Peaks found within the mass range 600−3000 Da of the mass spectrum of the tryptic digest of the 100-kDa protein



D)

M+H$^+$	*De novo* sequence
1129.6	LPYSTEYTR
1148.6	DGQLDFQGLR
1316.7	LLESDGFGPIIR
1423.7	NNFPLNSGSSSSGR
1601.8	EYGNYLGSGANTVTR
2001.9	GFGVTNTEAVSNSEAESFR
2217.0	NFASPNVAESSSGSESESFAGR

FIGURE 2: MALDI−TOF and MALDI tandem TOF analyses of tryptic fragments generated from the 100-kDa protein species. (A) Peptide map obtained by in-gel tryptic digestion of the 100-kDa protein doublet extracted from the egg case silk of *L. hesperus* using 8 M GdnHCl. (B) High-energy CAD (MS/MS) spectrum of the precursor ion with *m*/*z* 1129.6. The sequence of this peptide was found to be LPYSTEYTR. (C) High-energy CAD spectrum of the precursor ion with *m*/*z* 1423.7. This sequence was found to be NNFPLNSGSSSSGR. (D) Peptide sequences obtained from in-gel tryptic digestion of the core egg case silk components, followed by MALDI tandem TOF mass spectrometric analysis of the peptides. Underlined I or L residues indicate that these amino acids were confirmed at this position by their corresponding *w* ions. All peptides were found within the retrieved cDNA sequence of ECP-2 after translation (Figure 3A).

doublet (Figures 1, 2D, and 3A) account for 43% of the theoretical primary sequence of ECP-2. Peptides that were components of ECP-1 were excluded from this analysis. Because ECP peptides smaller than 600 and larger than 3000 Da, which were not detected by MS and not included in the peptide map, occupy 44% of the theoretical sequence, we consider 43% sequence coverage as high. The predicted mass of ECP-2 is 79.6 kDa (translation initiated from the M), with a theoretical pI of 7.38. Interestingly, excision of another major protein band migrating at ~80 kDa (Figure 1; see the asterisk), followed by in-gel trypsin digestion and MALDI MS analysis, produced a peptide fingerprint that overlapped with peptide ions found by MALDI MS analysis of the 100-kDa protein doublet (data not shown). Because the peptide sequences obtained by in-gel tryptic digestion of the 100-kDa protein doublet were used to retrieve the ECP-2 cDNA, the ~80-kDa protein could be a degradation product of ECP-2. Alternatively, the experimentally observed ~80-kDa band could represent the monomeric form of ECP-2, suggesting that the 100-kDa protein doublet contains two or more silk proteins that have failed to denature.

Expression of ECP-2 Is Predominantly Restricted to the Tubuliform Gland. To examine ECP-2 mRNA levels in the black widow, we collected total RNA from a variety of different spider tissues for RT-qPCR analysis. The orange tubuliform gland showed the highest level of ECP-2 mRNA expression, with lower levels detected in the white tubuliform and minor ampullate glands (Figure 4A). The orange tubuliform gland was observed to synthesize ~8- and ~29-fold more ECP-2 mRNA relative to the white tubuliform and minor ampullate glands, respectively (Figure 4A). The detection of the ECP-2 mRNA in the tubuliform gland supports the role of this tissue in egg case silk production. The low levels of ECP-2 in the ovaries and major ampullate tissues likely reflect baseline levels of gene transcription. A comparison of ECP-1 and ECP-2 mRNA expression levels demonstrated that ECP-2 is ~2.6-fold more abundant in the orange tubuliform gland (Figure 4B). To compare the relative expression patterns for the ECPs and TuSp1, we also examined the TuSp1 transcript levels in the tubuliform glands (Figure 4B). TuSp1 mRNA levels were found to be substantially higher, with ~1110- and 429-fold more mRNA relative to ECP-1 and ECP-2, respectively. However, because individual TuSp1 molecules contain iterated motifs (estimated at ~23 motifs, each ~552 bp in size) and our primer set targeted a single motif, the RT-qPCR raw data require normalization to account for the repetitive nature of the TuSp1 template. Thus, after adjustment, TuSp1 mRNA levels were determined to be ~48- and ~19-fold higher relative to ECP-1 and ECP-2, respectively.

Black Widow Egg Case Core Fiber Is a Trimeric Complex Containing TuSp1, ECP-1, and ECP-2. To investigate the constituents of the core fiber of egg case silk, we removed the coating proteins along with elements that comprise the small-diameter fibers with guanidine, dissolved the insoluble large-diameter core fibroin (~70% of the silk mass) in HFIP, and then digested the fiber solution with trypsin. The efficiency of stripping off coating proteins and the elimination of the smaller diameter fibers were monitored by scanning electron microscopy and SDS−PAGE analysis

A)

```
  1  GTRLSNCKRM  FTLVGLLSLL  GVQIGIALGD  DVCFNKCLSK  ISGECQSLVY
 51  TQINPCSFEC  NCDGVYSYHV  EETFTRCGSH  KLCYQGECLT  EVPRQCQRRY
101  GYGYIGLLNA  YNPCAFSCYN  ADVPCELYEE  NLVDGTACTT  SNSVIGQCLL
151  GRCTKVRIPY  STEYTRGIYP  IRDGQLDFQG  LRIPSASSTV  NAVRSAATEG
201  EVGGSESNAA  AAASSEAYNR  GEGNANSRAV  GNARTSVGQN  SAARAEAAAA
251  ASSETYNRGE  SIASSRAAGN  ARTSVGQNSA  ARAEAAAAAR  SRGSQGGSSA
301  TSSASADAAA  ATGRRGQYGV  ATGSATGVSQ  GYGATISSRR  QQQGFSEALS
351  QSDSRNNFPL  NSGSSSSGRA  TSRGIARNFA  SPNVAESSSG  SESESFAGRN
401  SAGSRGSSVG  RAVNQGAGRL  TNSLSESESG  AYSSNYGSGS  RGISRGSAAS
451  FGGRGFGVTN  TEAVSNSEAE  SFRDSSNVAG  LPEVSSSSGS  RSTQISYESD
501  GRDGINYKIL  ESNGFGPIIR  SGRISTSESD  GIEADSYGKG  SNAIVAGNSF
551  VTSSGGQNDN  IFAATNNFVN  SGVNGITNSI  FTQGGGINSA  SSSVSNSQAG
601  RREYGNYLGS  GANTVTRFLN  GEDNIWSIPS  RSAVGTGIFS  GRGNSVYSGS
651  SGESGANART  YRNGGAGATA  GAEAGAASGA  AAGAGASSGA  GAGAGASSGA
701  GAGAGAGASS  GAGARAGAGA  GAGAGAGASS  GADANAGAVA  SSGAGANAGA
751  GASSGAGANA  GVGAGAGAGA  NAGAGAGAGA  NAGAGAGAGA  NAGAGAGAGA
801  GASSGAGAGA  GAAASAGAEA  AAGARGRAAH  FANL
```

B)

```
ECP-1      ----------MFTFLGLISLLGVQIGIALGQGDVCFNKCLSRISGGCQSLIYTQVNPCA
ECP-2      --GTRLSNCKRMFTLVGLLSLLGVQIGIALG-DDVCFNKCLSKISGECQSLVYTQINPCS
                     ***::**:************  *********:*** ****:***:***

ECP-1      FQCTCDGVVTYHVEETFTKCGSRKLCYQGECLTEVPNRCERRYGYGYIGLLNPYNPCVFY
ECP-2      FECNCDGVYSYHVEETFTRCGSHKLCYQGECLTEVPRQCQRRYGYGYIGLLNAYNPCAFS
           *:* **** :********:***:************* :*:************ ****:*

ECP-1      CHNADVPCESFEENFVDGTTCYSSNSVIGQCLLGRCAEGGLTFSSGYIQGQ
ECP-2      CYNADVPCELYEENLVDGTACTTSNSVIGQCLLGRCTKVRI----------
           * ******* :***:****:* :************** :: :
```

C)

$[M+H]^+$	Location	$[M+H]^+$	Location
709.3	503-508	718.4	167-172
730.4	283-290	743.4	828-834
772.4	412-419	806.4	259-266
809.4	446-454	990.5	235-244 and 273-282
990.5	235-244	1051.6	632-642
1129.6*	158-166	1148.6*	173-182
1201.7	183-194	1242.6	492-502
1315.7*	509-520	1411.7	245-258
1423.7*	356-369	1599.7	643-659
1601.8*	603-617	1647.8	618-631
1658.8	524-539	1735.8	474-491
1867.9	293-314	2001.9*	455-473
2217.0*	378-399	2275.1	316-339
2457.1	195-220		

Sequence coverage 43%

FIGURE 3: Primary sequence of ECP-2 shows similarity to the spider fibroin-like member ECP-1. (A) Translation of the nucleotide sequence from the ECP-2 cDNA contains an ORF. The longest ORF encodes a protein 825 amino acids in length (starting from highlighted M). The conceptual translation product predicts a 79.6-kDa protein with an estimated pI of 7.38. One potential start codon with a good Kozak sequence was found within the ECP-2 ORF (*27*). Peptide sequences determined by MS/MS that are found within the ORF are indicated by underlines. Stretches of A_n tracts and fibroin GA repeats are depicted in blue and bold. Cysteine residues potentially involved in disulfide bond linkages are shown in red and bold. (B) Alignment of the N termini of ECP-1 and ECP-2. Dashes represent insertion of gaps to maximize alignment. Asterisks indicate identical residues, whereas colons represent R groups with similar side-chain properties. (C) Sequence coverage for peptide ions found in the translated ECP-2 cDNA sequence was found to be ~43%. Peptide masses marked with an asterisk indicate fragments sequenced by MS/MS.

(data not shown). The tryptic peptides generated from the large-diameter filament material were then analyzed by MALDI MS (Figure 5A). A total of 18 peptide ions were further subjected to MS/MS analysis to determine their primary sequences (Figure 5B). Peptides with *m/z* 619.4, 1502.7, and 1613.8 showed 100% matches to predicted sequences within the translated ECP-1 cDNA sequence (Figure 5B) (*15*). A peptide with *m/z* 1750.8 was also within ECP-1 but resulted from one missed cleavage event and contained an internal disulfide bond. The MS/MS-derived

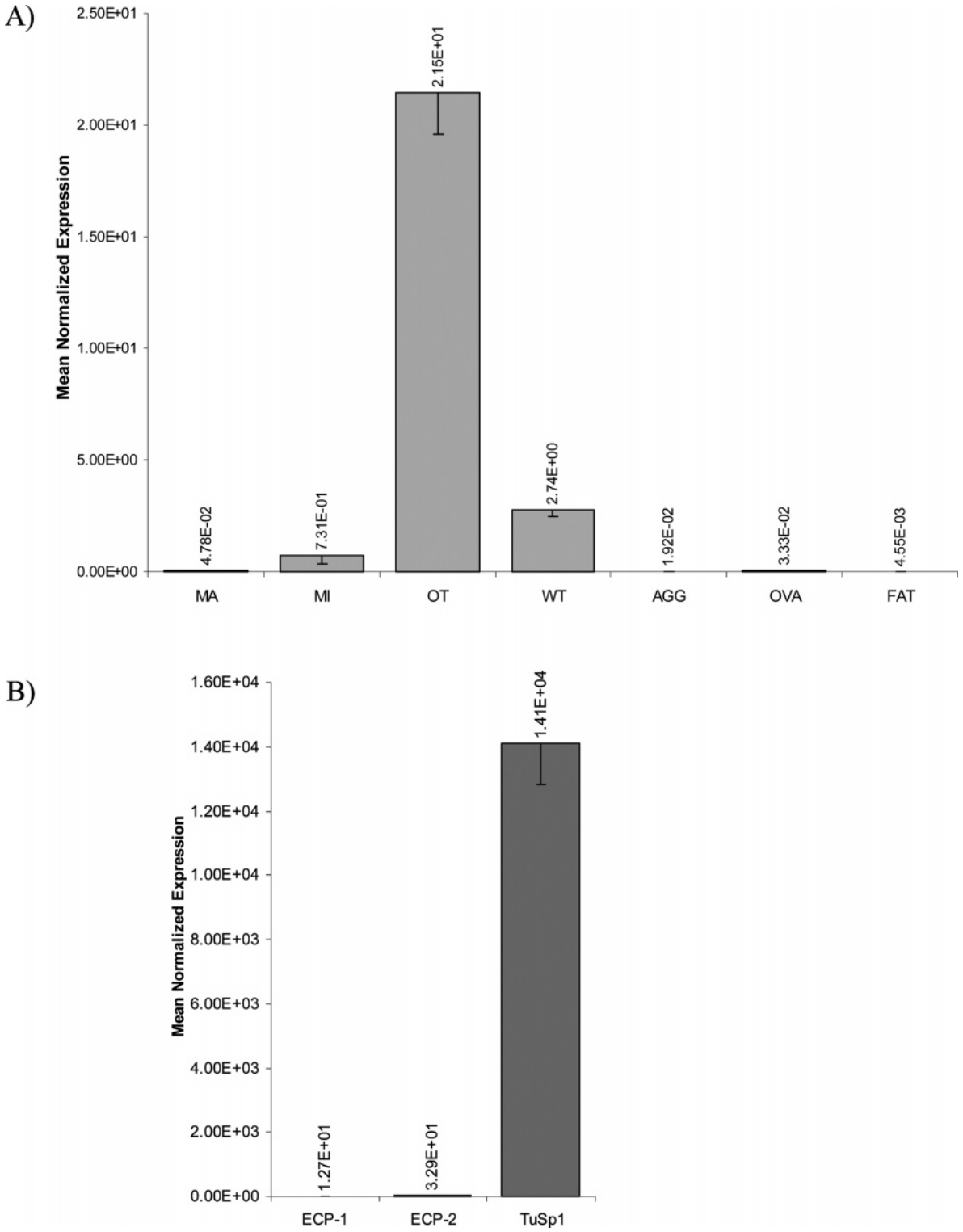
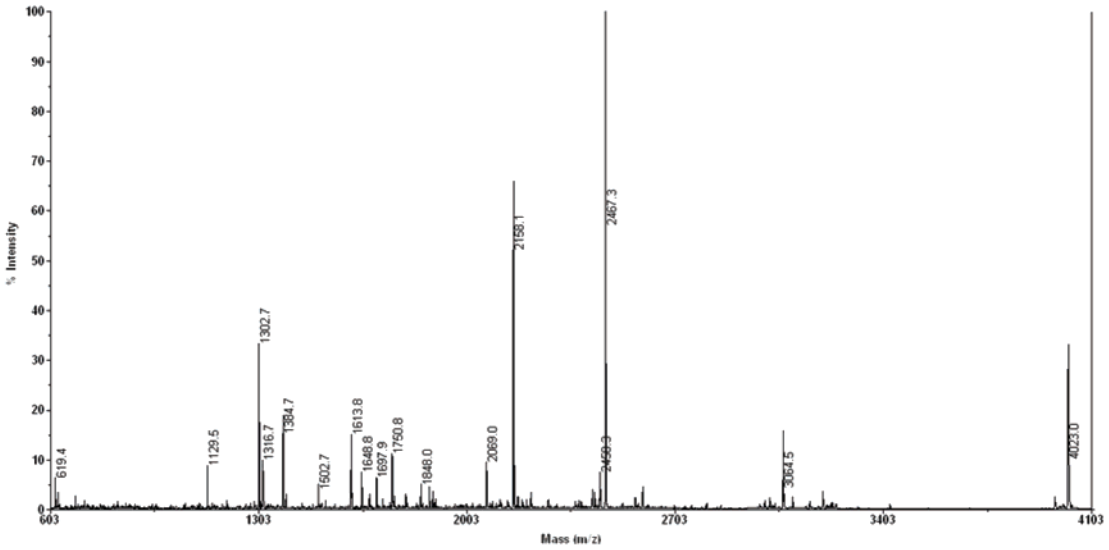


FIGURE 4: ECP-2 mRNA is tubuliform gland-specific and found in higher levels relative to ECP-1 transcripts. (A) RT-qPCR analysis was used to determine the expression pattern of ECP-2 in a variety of different tissues. The total RNA was isolated from the major ampullate gland, minor ampullate gland (MI), orange tubuliform (OT), white tubuliform (WT), aggregate (AGG), ovaries (OVA), and fat (FAT). (B) RT-qPCR analysis was used to determine the relative abundance of ECP-1, ECP-2, and TuSp1 mRNAs. For RT-qPCR experiments, equivalent amounts of total RNA were reversed-transcribed using MMLV and aliquots used for the RT-qPCR analysis. Orange tubuliform glands were used for the ECP-1 and ECP-2 analysis, whereas white tubuliform glands were used for the examination of TuSp1 expression. Samples were performed in triplicate and normalized internally using 18S rRNA as previously described (*14*). Data are representative of experimental results obtained from two independent trials.

sequences of peptides with *m*/*z* 1129.5, 1316.7, 1384.7, 1648.8, and 1848.0 demonstrated that these peptides were located within the predicted primary sequence of ECP-2 (Figures 3A and 5B). However, peptide ions 1384.7 and 1848.0 contained one missed cleavage event, whereas ions 1316.7 and 1648.8 were found to contain an N to D transition that likely resulted from hydrolysis. One could speculate that Asn residues promote in-chain peptide cleavage by losing a primary amine. This reaction could occur when the sample is treated with guanidine and boiled. Peptides from TuSp1

A)



B)

$[M+H]^+$	origin	Sequence confirmed by MS/MS
619.4	ECP-1	FSLPR
1129.5	ECP-2	LPYSTEYTR
1302.7	TuSp-1	APAFADAVSQAVR[±]
1316.7	ECP-2	LLESDGFGPIIR [$]
1384.7	ECP-2	VRLPYSTEYTR[&]
1502.7	ECP-1	TAGVSQSFGTAFSSR
1613.8	ECP-1	AAGFSGAQGSSGLQYGR
1648.8	ECP-2	FLDGEDNLWSLPSR [$]
1697.9	unknown	Sequence could not be determined
1750.8	ECP-1	KLCYQGECLTEVPNR [*&]
1848.0	ECP-2	GLYPLRDGQLDFQGLR[&]
1877.0	TuSp-1	AIEQNFVSSLASSPSLAR[±]
2069.0	unknown	Sequence could not be determined
2158.1	Tusp-1	AAQTLGLSNAPAFADAVSQAVR[±]
2450.3	unknown	Sequence could not be determined
2467.3	TuSp-1	GVGASPFQYANAVSNAFGQLLR
3064.5	TuSp-1	LPANSLSGVSASVNLFNSPNAATSFLNCLR
4023.0	TuSp-1	(TFY)SVSNADQAANLAYNLGLNLAPSLGLNPSEALASSLR

[&]1384.7, 1848.0 and 1750.8 each contain one missed cleavage
[$] 1316.7 and 1648.8 contain transformation from N to D by hydrolysis
[*] 1750.7 contains one intramolecular-disulfide bond
[±] 1302.7, 1877.0 and 2158.1 display breaks between NA (Fig. 5C) and reference [14]
Parentheses indicate confirmed sequence after examination of the genomic DNA sequence

C)

```
  1 MVWLTSTVLL ASLLGTLGLP ANSLSGVSAS VNIFNSPNAA TSFLNCLRSN IESSPAFPXQ
 61 EQADLDSIAE VILSDVSSVN TASSATSLAL STALASSLAE LLVTESAEED IDNQVVALST
121 ILSQCFVETT GSPNPAFVAS VKSLLGVLSQ SASNYEFVET ADESIAANVP GITTTGNFTP
181 LNAIEQNFVS SLASSPSLAR TFYSVSNADQ AANIAYNLGL NIAPSLGIPN SEALASSLRQ
241 AVASAGEGVN SSTYANLVSN TLGQFLSSQG ILTQSNAASL ASTLASAISQ SAAQSSAFAQ
301 SQAAAQAFSQ AASRSASQSA AQAGSSSTST TTTTSQAASQ AASQSASSSY SAASQSAFSQ
361 ASSSALASSS SFSSAFSSAS SASAVGQVGY QIGLNG
```

D)

```
TuSp1   93   ALASSLAELLVTESAEED-------IDNQVVAL-STILSQCFVETTGSPNPAFVASVKSL  144
             ALASS+AEL++ ES+  D       I N   AL + ++S     TTGSPN  FV  V+ L
FLAG    18   ALASSIAELVIAESSGGDVQRKTNVISN---ALRNALMS-----TTGSPNEEFVHEVQDL  69

TuSp1  145   LGVLSQ  150
             + +LSQ
FLAG    70   IQMLSQ
```

FIGURE 5: TuSp1, ECP-1, and ECP-2 are assembled into the large-diameter egg case core fiber. (A) MALDI−TOF analysis of the tryptic fragments generated from black widow egg case core fiber after removal of the coating proteins and small-diameter fiber material. Peptide ion 1877.0 is not labeled on the spectrum. (B) Peptide sequences obtained from MALDI MS/MS analysis of the fragments generated from in-solution tryptic digestion of the core egg case silk components, shown in A. All peptides generated were either components of TuSp1, ECP-1, or ECP-2. (C) N-Terminal sequence of TuSp1 was retrieved using a genomic DNA promoter walking strategy (*28*). Bold regions denote peptides identified in the N-terminal region of TuSp1 after tryptic digestion and MS/MS analysis, shown in B. Two Cys residues that could potentially be involved in disulfide bridge formation are indicated as bold and underlined species. The boxed region denotes the nonrepetitive N-terminal region of TuSp1. The arrow denotes the predicted cleavage location for the N-terminal signal sequence using a signal secretion prediction program (*19*). (D) Alignment of the nonrepetitive region of TuSp1 with the N-terminal region of *N. clavipes* FLAG silk (*29*) using the protein−protein BLAST algorithm. Numbers indicate the position in the primary sequence, and dashes are inserted gaps. Pluses indicate amino acids with similar side-chain properties.

were also observed, corresponding to peptides of *m*/*z* 1302.7, 1877.0, 2158.1, 2467.3, 3064.5, and 4023.0 (Figure 5B), the majority of which were generated by a combination of tryptic digestion and acid hydrolysis of Asn residues. Asn residues have been reported to represent hot spots for nonenzymatic degradation of proteins (*25*). Sequences for peptides *m*/*z* 1697.9, 2069.0, and 2450.3 could not be successfully deduced by MS/MS. Collectively, our data support the assertion that the core egg case fibroin is a trimeric complex consisting of TuSp1, ECP-1, and ECP-2. This is supported by the fact that all of the egg case core tryptic peptides sequenced by MS/MS analysis could be accounted for within the primary sequences of TuSp1 and the ECPs.

The N Terminus of TuSp1 Contains a Nonrepetitive Architecture. Partial cDNAs corresponding to the 3′ end of the gene have been published (*12*−*14*). Because some of the peptide ions sequenced from the core fiber (Figure 5B) could not initially be identified from the reported primary sequences of TuSp1 (partial sequence), we reasoned that these peptides might reside in the N-terminal region of TuSp1. To retrieve the N terminus of TuSp1, we used a genomic DNA walking approach. This methodology successfully led to the isolation of a genomic DNA fragment containing the N-terminal sequence of TuSp1 (Figure 5C). A nonrepetitive region consisting of 190 amino acids was found, containing two Cys residues. Three of the peptides that initially were not able to be accounted for within TuSp1, ECP-1, or ECP-2 (Figure 5B), which included peptide ions with masses 1877.0, 3064.5, and 4023.0, were located in the N terminus of TuSp1. Protein−protein BLAST searches using the N-terminal 190 nonrepetitive region of the NCBInr database identified weak similarity to flagelliform silk (Figure 5D; *E* value of 0.002).

DISCUSSION

Recent studies have indicated that one of the major constituents of egg case silk in spiders is TuSp1. Several labs have published partial cDNAs encoding TuSp1 from a variety of different species that correspond to the C-terminal region of TuSp1 (*12*−*14*); however, no N-terminal information on TuSp1 has been reported. Although predicted amino acid compositions from these translated TuSp1 cDNA sequences closely align with the amino acid compositions of raw egg case silk, differences in the percentages of certain residues have been noted (*12*, *14*). These differences are likely reflective of the presence of other proteins that are assembled in the egg case fibers. To elucidate the molecular identities of proteins assembled in tubuliform silks, we have used a proteomic approach coupled with reverse genetics to reveal the constituents of black widow spider egg case silk. Our studies have led to the identification of a third protein that is found in black widow egg cases, named ECP-2. Common repetitive sequences reported in traditional fibroins were present within the primary sequence of ECP-2, including GA repeats and polyalanine blocks (Figure 3A). Protein−protein BLAST searches (blastp) of the NCBInr database show that ECP-2 is 52% identical at the amino acid level to ECP-1, another protein previously identified in spider egg case silk (*15*). Using this algorithm, no other proteins in the database showed significant similarity to ECP-2. Analysis of the primary sequence of ECP-2 demonstrates an N-terminal Cys-rich area, containing 16 Cys within the first 153 amino acids. Strikingly, alignments with ECP-1 reveal conservation with respect to the number and spacing of these Cys residues (Figure 3B). Because ECP-1 and ECP-2 both lack recognizable ensemble repeats (large blocks that contain subrepeat elements) that are commonly found in traditional fibroins, the question arises regarding their function in tubuliform silks. One possibility is that the ECPs function as intermolecular cross-linkers that play structural roles in the fiber, perhaps, by forming disulfide bridges via their N termini. Previous experiments conducted by our laboratory have shown that ECPs accumulate after prolonged treatment with reducing agent, which is supportive of higher ordered protein complexes containing ECP molecules (*15*). However, whether ECPs are directly involved in disulfide bond linkage, in particular with the N-terminal region of TuSp1, will require further biochemical analyses. We have demonstrated that the C-terminal region of ECP-2 contains a cluster of $(GA)_n$ repeats. These repeats are similar to motifs identified in silkmoth silks, which contain multiple repeats of a hexapeptide GAGAGA that form β sheets that comprise the crystalline regions of the fiber. Spider minor ampullate silks, which are strong nonelastic silks used in web formation, are composed of two proteins (MiSp1 and MiSp2) that are predominantly β-sheet structure and contain large numbers of GA repeats (*21*). The presence of GA repeats in ECP-2 supports the assertion that this molecule may have a structural role in the fiber.

Similar to the reported expression pattern for TuSp1 and ECP-1, ECP-2 mRNA production was shown to be predominantly restricted to the tubuliform gland (Figure 4A). During our dissections, we routinely observed two different colors of the tubuliform gland. Some were completely white, whereas others displayed a white-orange appearance. We classified those glands with no visible color as white and those with a combination of colors (white and orange) as orange. Our studies demonstrated that ECP-2 mRNA levels were higher in the orange than the white tubuliform gland. Similar results were obtained for ECP-1 (data not shown). Orange tubuliform glands have only been identified in females, and their appearance correlates well with the laying of egg cases. In the orange tubuliform gland, we observed ~2.6-fold higher levels of ECP-2 mRNA relative to ECP-1, but the explanation for the elevated ECP-2 mRNA levels is currently unclear. When ECP transcript levels were compared to TuSp1, our RT-qPCR analyses demonstrate that TuSp1 mRNA levels are considerably higher in tubuliform tissue. TuSp1 transcript levels were determined to be ~48- and ~19-fold higher relative to ECP-1 and ECP-2, respectively. Higher levels of TuSp1 mRNA are consistent with the fact that TuSp1 proteins comprise the major constituent of the large-diameter core fiber found in egg case silk. This is further supported by the amino acid composition profile of egg cases, which closely reflects the predicted amino acid percentages predicted from translation of the retrieved TuSp1 cDNA sequences.

Previously, we have shown that egg cases from black widow spiders consist of a network of filaments that are distinguished by two different diameter sizes. The larger diameter fibers, which are 4−5 μm, constitute the bulk of the egg case material (*15*). In our studies, the treatment of egg cases with GdnHCl results in the solubilization of the smaller diameter fibers, as well as partial disruption of the

large-diameter fiber. This is supported by the examination of the insoluble fibrous material after GdnHCl treatment (about 70% of the total egg case mass) using the scanning electron microscope, which results in the detection of predominantly large-diameter filaments. In our studies, tryptic digestion of the large-diameter core fiber (HFIP-treated egg case material) generated 18 different peptide ions. A total of 15 peptides were successfully sequenced by MS/MS analysis and found to be constituents of ECP-1, ECP-2, or TuSp1. On the basis of these findings, our data support that the large-diameter fiber of the egg case is trimeric in nature.

To determine whether TuSp1 proteins were partially stripped off the filaments upon initial GdnHCl treatment, we performed in-solution tryptic digestion and MALDI MS/MS analysis on proteins extracted from egg cases after GdnHCl incubation. Surprisingly, no traces of TuSp1 peptides were detected after tryptic digestion and MS/MS analysis (data not shown). Unlike ECP-1 and ECP-2, which can be partially peeled off with initial GdnHCl treatment (Figure 1), TuSp1 peptides were found only after the GdnHCl-treated insoluble fibers were processed with HFIP and trypsin (Figure 5). The detection of TuSp1 peptides only after examination of insoluble proteins through the use of HFIP and trypsin is likely explained by the following events: (1) the low polarity of the solvent HFIP facilitated the breakdown of the insoluble fibers by disrupting local hydrophobic interactions that maintain the crystalline regions of the fiber, which resulted in the freeing of more monomeric fibroin units for proteolytic cleavage events, and (2) repeated boiling of the GdnHCl-treated silk (prior to the resuspension in HFIP) resulted in acid hydrolysis of the TuSp1 polypeptide backbone. Because TuSp1 contains relatively low amounts of Arg and Lys (combined ~1% of protein), the ability to generate peptides from tryptic cleavage alone is somewhat limited. Hydrolysis in combination with tryptic digestion increases the likelihood of TuSp1 peptide formation. Consistent with this proposal is the identification of three different TuSp1 peptides that were produced by a combination of hydrolysis and tryptic cleavage (*m*/*z* 1302.7, 1877.0, and 2158.1; Figure 5B). A fourth TuSp1 peptide, which corresponded to *m*/*z* 3064.5, was generated by the removal of the signal sequence at the N terminus and one tryptic cleavage (parts B and C in Figure 5). Only one TuSp1 peptide resulted from two tryptic cleavage events (*m*/*z* 4023.0; parts B and C in Figure 5). Large molecular-weight proteins (>200 kDa) were detected after egg cases were treated initially with GdnHCl (without boiling) when assayed by SDS−PAGE analysis and silver staining (Figure 1). However, in-gel tryptic digestion of these bands, followed by MALDI tandem TOF analysis, produced peptide sequences that were not represented within TuSp1 or the ECPs (data not shown). These large molecular-weight proteins likely represent constituents of the smaller diameter egg case fiber. Currently, there are two explanations for the detection of the ECPs in the initial GdnHCl treatment: (1) they are also constituents of the smaller diameter fibers, or (2) GdnHCl initiates their removal from the larger diameter fibers. Further experiments will be directed at addressing this issue, as well as whether TuSp1 molecules are extracted upon initial GdnHCl treatment.

ECP-1 and ECP-2 display similar mRNAs sizes, as well as protein masses, which may suggest that they are produced by alternative RNA splicing. Somewhat consistent with this view is the fact that nucleotide−nucleotide BLAST searches of the NCBInr database show that the DNA sequence of ECP-2 has three distinct regions of similarity to the nucleic acid sequence of ECP-1, which include nucleotides 25−458 (identities 357/437 or 81%), 1298−1416 (identities 110/120 or 91%), and 1915−1954 (identities 37/40 or 92%). Alternatively, ECP-1 and ECP-2 could be encoded by distinct genes. Retrieval of the genomic DNA fragments containing the ECPs should help resolve this issue. The detection of low levels of ECP-2 mRNA in the minor ampullate gland (Figure 4A) could also suggest that ECP-2 has a broader function in minor ampullate silks, such as dragline silk. Further experiments will need to be conducted to address this issue.

In summary, we have combined MS with reverse genetics to elucidate the molecular constituents of tubuliform silks in spiders. We have cloned the cDNA encoding a third protein component of the black widow egg case silk fibers, which we have dubbed ECP-2. Moreover, using primary sequence information obtained from TuSp1 and the ECP cDNAs, as well as experimentally obtained peptide fragments generated from a combination of acid hydrolysis and enzymatic digestion, we demonstrate that the large-diameter core fibers of black widow egg case silk consist of trimeric complexes, containing TuSp1, ECP-1, and ECP-2. The structure is somewhat different from dragline silk, which has been shown to contain two proteins, which include MaSp1 and MaSp2 (*21*, *22*, *26*). Although we have also detected MaSp1 and MaSp2 mRNAs in the tubuliform gland (data not shown), similar to other reports (*13*), we have been unable to detect MaSp1 and MaSp2 peptides in the core fibers of egg case silk. This demonstrates that MaSp1 and MaSp2 proteins are not assembled into the core egg case filaments, despite the presence of the MaSp1 and MaSp2 transcripts in the tubuliform gland. The exact reason for the presence of these mRNAs within this gland is unclear. In the future, similar experimental approaches combining MS and reverse genetics should allow for the rapid identification of other molecules integrated into different core fiber silk types, as well as the elucidation of minor constituents assembled into silks. This information will provide 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 Jennifer-Grace Cabug, Stephanie Davis, Caroline Lai, Mercedes Lenoncio, Olivia Martinez, Grace Mercado, Amy Nguyen, Robert Passamano, Stephanie Phan, Jenie Rocetes, and Joseph Schooler for their contributions on the RT-qPCR analysis.

REFERENCES

1. Gosline, J., DeMont, M. E., Denny, M. W. (1986) The structure and properties of spider silk, *Endeavour 10*, 31−43.
2. Vollrath, F. (1992) Spider webs and silk, *Sci. Am. 266*, 70−76.
3. 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.
4. 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.

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. 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.
7. 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.
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. Kovoor, J. (1987) Comparative structure and histochemistry of silk-producing organs in arachnids, in *Ecobiology of Spiders* (Nentwig, W., Ed.) Vol. B, pp 159−186, Springer-Verlag, Berlin, Germany.
10. Foradori, M. J., Kovoor, J., Moon, M. J., and Tillinghast, E. K. (2002) Relation between the outer cover of the egg case of *Argiope aurantia* (Araneae: Araneidae) and the emergence of its spiderlings, *J. Morphol. 252*, 218−226.
11. Candeles, G., Candelas, T., Ortiz, A., and Rodriguez, O. (1983) Translational pauses during a spider fibroin synthesis, *Biochem. Biophys. Res. Commun. 116*, 1033−1038.
12. Tian, M., and Lewis, R. V. (2005) Molecular characterization and evolutionary study of spider tubuliform (eggcase) silk protein, *Biochemistry 44*, 8006−8012.
13. Garb, J. E., and Hayashi, C. Y. (2005) Modular evolution of egg case silk genes across orb-weaving spider superfamilies, *Proc. Natl. Acad. Sci. U.S.A. 102*, 11379−11384.
14. Hu, X., Lawrence, B., Kohler, K., Falick, A. M., Moore, A. M., McMullen, E., Jones, P. R., and Vierra, C. (2005) Araneoid egg case silk: A fibroin with novel ensemble repeat units from the black widow spider, *Latrodectus hesperus*, *Biochemistry 44*, 10020−10027.
15. Hu, X., Kohler, K., Falick, A. M., Moore, A. M., Jones, P. R., Sparkman, O. D., and Vierra, C. (2005) Egg case protein-1. A new class of silk proteins with fibroin-like properties from the spider *Latrodectus hesperus*, *J. Biol. Chem. 280*, 21220−21230.
16. Jiménez, C. R., Huang, L., Qiu, Y., and Burlingame, A. L. (1998) *Current Protocols in Protein Science: In-Gel Digestion of Proteins for MALDI−MS Fingerprint Mapping*, p 16.4.1−16.4.5., John Wiley and Sons, Inc., New York.
17. Kohler, K., Thayer, W., Le, T., Sembhi, A., Vasanthavada, K., Moore, A. M., and Vierra, C. A. (2005) Characterization of a novel class II bHLH transcription factor from the black widow spider, *Latrodectus hesperus*, with silk-gland restricted patterns of expression, *DNA Cell. Biol. 24*, 371−380.
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. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0, *J. Mol. Biol. 340*, 783−795.
20. Zhou, C. Z., Confalonieri, F., Medina, N., Zivanovic, Y., Esnault, C., Yang, T., Jacquet, M., Janin, J., Duguet, M., Perasso, R., and Li, Z. G. (2000) Fine organization of *Bombyx mori* fibroin heavy chain gene, *Nucleic Acids Res. 28*, 2413−2419.
21. 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.
22. 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.
23. Tomizawa, H., Yamada, H., and Imoto, T. (1994) The mechanism of irreversible inactivation of lysozyme at pH 4 and 100 °C, *Biochemistry 33*, 13032−13037.
24. Scaloni, A., Dalla Serra, M., Amodeo, P., Mannina, L., Vitale, R. M., Segre, A. L., Cruciani, O., Lodovichetti, F., Greco, M. L., Fiore, A., Gallo, M., D'Ambrosio, C., Coraiola, M., Menestrina, G., Graniti, A., and Fogliano, V. (2004) Structure, conformation, and biological activity of a novel lipodepsipeptide from *Pseudomonas corrugata*: Cormycin A, *Biochem. J. 384*, 25−36.
25. 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.
26. 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.
27. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
28. Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A., and Lukyanov, S. A. (1995) An improved PCR method for walking in uncloned genomic DNA, *Nucleic Acids Res. 23*, 1087−1088.
29. 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.

BI052105Q