Egg Case Protein-1

A NEW CLASS OF SILK PROTEINS WITH FIBROIN-LIKE PROPERTIES FROM THE SPIDER
LATRODECTUS HESPERUS*

Received for publication, November 1, 2004, and in revised form, March 23, 2005
Published, JBC Papers in Press, March 29, 2005, DOI 10.1074/jbc.M412316200

Xiaoyi Hu‡§¶, Kristin Kohler¶§¶, Arnold M. Falick¶, Anne M. F. Moore¶, Patrick R. Jones‡,
O. David Sparkman¶, and Craig Vierra**

From the ‡Department of Chemistry and the ¶Department of Biological Sciences, University of the Pacific, Stockton, California 95211 and Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, Berkeley, California 94720

Spiders produce multiple types of silk that exhibit diverse mechanical properties and biological functions. Most molecular studies of spider silk have focused on fibroins from dragline silk and capture silk, two important silk types involved in the survival of the spider. In our studies we have focused on the characterization of egg case silk, a third silk fiber produced by the black widow spider, Latrodectus hesperus. Analysis of the physical structure of egg case silk using scanning electron microscopy demonstrates the presence of small and large diameter fibers. By using the strong protein denaturant 8 M guanidine hydrochloride to solubilize the fibers, we demonstrated by SDS-PAGE and protein silver staining that an abundant component of egg case silk is a 100-kDa protein doublet. Combining matrix-assisted laser desorption ionization tandem time-of-flight mass spectrometry and reverse genetics, we have isolated a novel gene called ecp-1, which encodes for one of the protein components of the 100-kDa species. BLAST searches of the NCBI nr protein data base using the primary sequence of ECP-1 revealed similarity to fibroins from spiders and silkworms, which mapped to two distinct regions within the ECP-1. These regions contained the conserved repetitive fibroin motifs poly(Ala) and poly(Gly-Ala), but surprisingly, no larger ensemble repeats could be identified within the primary sequence of ECP-1. Consistent with silk gland-restricted patterns of expression for fibroins, ECP-1 was demonstrated to be predominantly produced in the tubuliform gland, with lower levels detected in the major and minor ampullate glands. ECP-1 monomeric units were also shown to assemble into higher aggregate structures through the formation of disulfide bonds via a unique cysteine-rich N-terminal region. Collectively, our findings provide new insight into the components of egg case silk and identify a new class of silk proteins with distinctive molecular features relative to traditional members of the spider silk gene family.

* This work was supported by Collaborative Research at Undergraduate Institutions Grant DBI-0112165 from the National Science Foundation and Scholarly Activities Achievement Grant from the University of the Pacific. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** § Both authors should be considered as first authors.

This paper is available on line at http://www.jbc.org
with tubuliform silks containing substantially higher levels of serine relative to the silk from the minor ampullate, major ampullate; and flagelliform glands (17). Because tubuliform silks have different chemical compositions and mechanical properties relative to other araneoid silks, we hypothesized that tubuliform silks contain unique proteins assembled into the egg case silk fiber. In our search for genes encoding egg case silk proteins, we have isolated a novel cDNA that codes for a tubuliform-restricted protein. Based upon the pattern of expression and the abundance of the protein in egg case fibers, we have named this gene product ECP-1 (egg case protein-1). The primary sequence of ECP-1 represents the only completed sequence for egg case silk proteins. Analysis of the predicted primary sequence of ECP-1 reveals similarity to published fibroin proteins from spiders and silkmoths. By SDS-PAGE analysis of egg case silk protein samples, we also demonstrate ECP-1 monomers are assembled into higher ordered structures and comprise abundant components of egg case silk. Our findings suggest that ECP-1 represents a new class of silk proteins belonging to the spider gene silk family.

MATERIALS AND METHODS

Scanning Electron Microscopy—Egg case silk was coated to a thickness of ~14–20 nm with gold alloy in a Pelco SC-7 auto-sputter coater with an FTM-2 film thickness monitor. Samples were examined on a Hitachi S-2600 S.E. operated with an accelerator voltage of 3 kV. The diameter of both strands was measured to the nearest 0.01 μm at three distant places along the S.E. sample. Tests were done at ambient temperature and humidity, which ranged from 22 to 25°C and 30–36%, respectively.

Characterization of Egg Case Proteins by Differential Solubilization—Individual egg cases from female black widow spiders were cut open using sterilized scissors to remove the eggs. The egg case was discarded if any eggs were broken during the isolation procedure. Silk from each egg case was extracted with 8M GdnHCl (3 ml of solution per mg of silk) for 10 min with agitation. The viscous supernatant solution was removed and conserved for dialysis. The solid residue was then exchanged with 1 ml of water, which left the solid white core fiber. The water extract solution was also conserved for dialysis. A second, fresh 8 M GdnHCl solution (same volume to weight ratio as the initial solution) was then added to the silk residue, and the mixture was agitated continuously with a vertical shaker. Samples of this supernatant were collected after 10 min and after 5, 15, 25, and 40 h. Each of the silk solutions obtained from initial GdnHCl and water extracts, and multiple samples from the second GdnHCl extraction was dialyzed against three changes of 100 ml of 50 mM Tris-HCl (pH 7.8) for 4 h each, using 1000 molecular weight cut-off dialysis tubing (Sigma). The proteins from the dialyzed samples were separated on SDS-PAGE with a 4–20% polyacrylamide gradient gel (Bio-Rad). Separated proteins were visualized by silver staining (ProteoSilver™ Plus silver stain kit, Sigma). The results are shown in Fig. 2, lanes 1–8. Broad range molecular weight markers were used to determine protein sizes (Bio-Rad).

Tryptic Digestions of Egg Case Protein and Mass Spectrometric Analysis—Sequencing grade trypsin (trypsin gold, Promega) was dissolved in 50 μl acetate acid at a concentration of 1 mg/ml. This solution was then diluted with 50 mM NH4HCO3, pH 7.5 to give a 2 μg/ml stock solution of trypsin. In-solution digest of the 40-h extract of egg case silk were prepared by mixing 100 μl of the ECP protein solution with 10 μl of trypsin solution and incubating at 37°C overnight. The digest was then completely dried using a vacuum centrifuge and redisolved in 10 μl of 0.1% aqueous trifluoroacetic acid. Peptides were extracted and desalted with a C18 Zip-Tip (Millipore) according to the manufacturer’s instructions. In order to test whether any ECP species were also included with the peripheral proteins from the first GdnHCl extraction, we prepared an in-gel tryptic digest of the 100-kDa bands from this extract following a published protocol (18). Briefly, the 100-kDa bands in Fig. 2, lane 2, were excised from the gel. After destaining, the pieces were minced into fine particles with sealed pipette tips, washed with 25 mM NH4HCO3, 50% acetonitrile, and then dried in a vacuum centrifuge. The dried pieces were rehydrated and reduced with 10 mM dithiothreitol, alkylated with 55 mM iodoacetamide, dried again, and then rehydrated and digested overnight with 250 ng of trypsin. Peptides were extracted with two changes of 30 μl of 50% acetonitrile, 5% formic acid solution. The extracts were combined and then reduced in volume to about 10 μl to concentrate the sample, after which the solutions were desalted with C18 Zip-Tips.

In-gel digestion was also carried out for the ECPs in the 40-h extracts (100-kDa bands in Fig. 2, lane 8). In order to enhance peptide sequence coverage, 300 μl of the 40-h extracts were concentrated to 40 μl using PAGE Prep™ Protein Clean-up and Enrichment kit (Pierce), 20 μl of which were loaded on a separate 4–20% gradient gel. The 100-kDa bands were excised and subjected to in-gel trypsin digestion.

The three digests (the in-gel positive tryptic digests of ECPs and the in-gel digest of the 100-kDa bands from the first GdnHCl wash) were analyzed with a MALDI tandem TOF mass spectrometer (4700 Proteomics Analyzer, Applied Biosystems, Foster City, CA). One-half microliter of each desalted digest solution was mixed with an equal volume of a 10 mg/ml solution of a-cyano-4-hydroxycinnamic acid in 50/50 acetonitrile, 0.1% aqueous trifluoroacetic acid and spotted on the MALDI target. For the in-gel digestion, two negative controls were used as follows: a mixture of 100 μl of 50 mM Tris-HCl (pH 7.8) and 10 μl of trypsin solution, and a mixture of 100 μl of ECP sample and 10 μl of 50 mM NH4HCO3 (no trypsin) were treated in the same way as the actual digests. For the in-gel digests, a piece of blank gel was digested as described above.

Monoisotopic masses of all significant peptides generated from the trypsin digests were measured in a mass spectrometer and the monoisotopic masses were used to synthesize degenerate oligonucleotides corresponding to the underlined regions. PCRs containing the forward primer 5’-CAAGGWACAAACGGWTYT-3’ (encodes for QQQQFP; Y = T or C; W = A or T) and the reverse primer 5’-NGNCCRAANCCRTC-3’ (specifies DGFP; R = A or G; N = A, G, T, C) were designed and amplified a 525-bp fragment of ECP-1 from a cdna library produced from black widow silk glands (corresponded to amino acids 353–527 Fig. 4). The 525-bp fragment was sequenced as described previously (19). To amplify the 5’ end of the ECP-1 cDNA, the forward primer (anchor) from the pGALA-AD library vector (5’-CGATGATGAA-GATACCCAC-3’) was used with the gene-specific reverse primer 5’-CTTATTGTAATTCCAGGAGGACC-3’ (nucleotides 1506–1527, corresponding to amino acids 482–492). The product from the pGALA-AD forward and reverse primer resulted in an ~1.4-kb product. A similar approach was used to obtain the 3’ end of the cDNA. The reverse primer from the pGALA-AD library vector 5’-CGACAGTTGAAGTGAATCTC-3’ and the forward primer 5’-GAGGAGAGGAGCGCTGTT-5’ (nucleotides 1405–1427, corresponding to residues 464–470) amplified a product that was ~1.7 kb. Both 5’- and 3’-RACE products were purified using the QiAquick gel extraction kit according to the manufacturer’s instruction (Qiagen) and sequenced as described above.

Amino Acid Composition of the Core Silk Filament—The core silk fibers were subjected to amino acid analysis at the Protein Chemistry Laboratory of Texas A & M University as described previously (20). Briefly, vapor phase hydrolysis of the core proteins by 6 M HCl was employed to generate the constituent amino acids. Amino acids obtained were derivatized with o-phthalaldehyde and 9-fluoromethylchlo-roformate. The derivatized amino acids were separated by reversed phase high performance liquid chromatography with UV detection.

Real Time PCR Analysis with SYBR Green Detection—Reverse transcription reactions were carried out as described previously (21). Typically, 2 μl of a reverse transcription reaction were used for real time PCR analysis using the DyNamo SYBR Green qPCR kit (MJ Research).
Real time PCR fluorescence detection was performed in 96-well plates using an Opticon II instrument (MJ Research). Amplification products were monitored by SYBR Green detection and routinely checked using dissociation curve software and 1% agarose gel electrophoresis. Reactions were performed in triplicate. Oligonucleotides used for the analysis of ECP-1 were the forward and reverse primers 5’-GAATCCAG-TAGTGCCTCCCAATT-3’ (nucleotides 1110–1132) and 5’-TTGTGACTCTCTCCTCCTTGACT-3’ (nucleotides 1293–1314), respectively. Primers were selected using the Beacon Designer 2.0 software (PREMIER Biosoft International).

RESULTS

Physical Structure of Black Widow Egg Case Fibers—To understand more regarding the physical structure of egg case silk (Fig. 1A), we used a scanning electron microscope to examine the physical structure of egg cases collected from black widow spiders (Fig. 1B). The scanning electron micrograph shows that egg case silk is composed of fibers of two different diameters (Fig. 1C). The larger diameter fibers, which represent the major component of egg case silk, are produced by the tubuliform gland (22). The diameter of the large fibers was 4–5 μm, whereas the diameters the smaller fibers were on the order of 500 nm. The smaller diameter fibers, which constitute a minor component, are likely synthesized by the aciniform gland (23).

Identification of Egg Case Proteins—To examine the biochemical features and the number of proteins assembled into egg case silk, we dissolved egg cases collected from black widow spiders in 8 M GdnHCl. Analysis of the egg case extracts by SDS-PAGE followed by silver staining revealed a broad distribution of molecular masses, ranging from ~10 to ~300 kDa (Fig. 2, lane 2). Analysis of egg cases taken from six different black widow spiders generated similar protein patterns (data not shown). The broad distribution of molecular weight proteins is likely a result of translational pausing of fibroin mRNAs (14), protein degradation, and/or the presence of glue proteins.

To identify major structural proteins in the silk fiber, we performed a solubilization time course assay with GdnHCl. Egg case silk was repeatedly extracted with 8 M GdnHCl over a 40-h time period. Many egg case proteins dissolved immediately upon the first exposure to the GdnHCl solution (Fig. 2, lane 2). A distinct color change was observed after the first GdnHCl extraction and water wash, with the egg case changing from a yellowish hue to white. The treatment of silk with GdnHCl or other types of chaotropic agents has been reported to solubilize efficiently the hydrophobic proteins. SDS-PAGE analysis of the water wash following the first GdnHCl extraction of the silk produced a protein doublet with an apparent molecular mass of ~100 kDa (Fig. 2, lane 3). This protein doublet had the same apparent size as the strong 100-kDa species observed in the...
initial GdnHCl extract (compare Fig. 2, lanes 2 and 3). No other proteins were detected in the water wash (Fig. 2, lane 3). Because silk fibroins are not water-soluble, it is very probable that these species represent the diluted sample from the initial treatment because the 100-kDa doublet was the strongest band in the initial GdnHCl extraction. When the residue after the water wash (about 70% of the total mass) was extracted with 8 M GdnHCl for a second time, the remaining core fibroins and associated proteins displayed reduced solubility. Samples extracted for less than 15 h failed to generate sufficient proteins for detection by silver staining following SDS-PAGE (Fig. 2, lanes 4–6). However, after 25 h, a faint protein doublet was detected (Fig. 2, lane 7). This protein doublet became relatively stronger at 40 h (Fig. 2, lane 8), and these species were named egg case proteins or ECPs. No higher molecular weight proteins were detected by the SDS-PAGE at this point.

Trypsin Digestion of ECPs and Mass Spectrometric Sequence Analysis—We then performed in-solution tryptic digestions of the protein doublet described in the secondary 40-h GdnHCl extract fraction (Fig. 2, lane 8). This digest was found to contain numerous peptides whose molecular masses were determined by MALDI-MS analysis (Fig. 3A). Seven peptides were sequenced using high energy collision-activated dissociation (CAD), those with precursor ion masses (MH+ monoisotopic) of 855.4, 1316.7, 1502.7, 1613.8, 1623.8, 1754.8, and 3424.5. The product ion spectra from the peptides of MH+ 1316.7, 1613.8, and 3424.5 are shown in Fig. 3, D–F, respectively. The N termini of the 1613.8 and 3424.5 peptides could not be determined, but a significant amount of sequence information was derived. The primary sequences of the remaining peptides were also determined (Table I). The sequences of the peptides at m/z 1613.8 and 1623.8 were nearly identical, having only a single amino acid difference (Table I). The selection of ions for sequence analysis was deliberately biased toward peptide masses that were repeatedly observed from different black widow spiders (data not shown). Analysis of the derived peptide sequences using the algorithm BLAST revealed no significant similarity to any polypeptides in the NCBI nr protein database.

Most of the peptides from the in-solution digest of the 40-h extract (Fig. 3A) were also found in the in-gel digest of the 100-kDa proteins in the 40-h extract (Fig. 3B), indicating that the 100-kDa proteins are the major components of the 40-h extract. This is reasonable, because only the 100-kDa bands were observed upon SDS-PAGE analysis after 40 h of extraction.

MALDI-TOF mass spectrometric analysis of the in-gel tryptic fragments from the 100-kDa bands observed in the initial GdnHCl extract (Fig. 2, lane 2, and Fig. 3C) also had many of the same peptides as those observed in the in-solution and in-gel digestion of the ECPs (from the second GdnHCl extraction) (Fig. 2, lane 8, and Fig. 3, A and B), indicating that some ECPs are also dissolved during the initial GdnHCl extraction.

Isolation of the ecp-1 Gene—We then searched for cDNAs that encoded the proteins found in the ECPs. A cDNA library was prepared from mRNA collected from silk-producing glands, and the library was screened by PCR using different combinations of degenerate oligonucleotides that corresponded to the peptide sequences obtained by MS/MS (Table I). The forward and reverse primers designed from the peptides QQGGGF-SETLSQSDSR and LLESDFGGPPIR (the underlined region denotes the primer construction region), respectively, successfully amplified a 525-bp cDNA fragment (amino acid residues corresponded to 355–527 (Fig. 4A)). The amplified gene segment was named egg case protein-1 (ECP-1). No other combinations of primers generated gene-specific amplified products. By using the nucleotide sequence data obtained from the 525-bp gene piece, we amplified the remaining ends of the cDNA by using a modified form of rapid amplification of cDNA ends (RACE). Two overlapping cDNA fragments obtained from RACE were used to reconstruct a 2.9-kb cDNA fragment that coded for ECP-1 (Fig. 4A; GenBankTM accession number A9Y94149). In protein–protein BLAST searches against the entire sequence of ECP-1, the top matches corresponded to published fibroins (Nephila clavipes MiSp1, GenBankTM accession number AAC14589.1; Bombyx mori silk fibroin heavy chain, GenBankTM accession number AAB31861.1; Pleurotus tristis fibroin 1, GenBankTM accession number AAK30610.1; Araneus ventricosus major ampullate dragline silk protein 2, GenBankTM accession number AAN58281.1) and one cell wall protein from a bacterium (Streptococcus pneumoniae cell wall surface anchor family protein, GenBankTM accession number NP_346206.1). The regions of ECP-1 showing similarity to fibroins were clustered into two main areas, which spanned amino acid residues 180–342 and 600–890. The identities ranged in these areas from 35 to 45%, with the statistical E values ranging from 2 × 10–36 to 2 × 10–41. No similarity was found to the conserved, nonrepetitive C-terminal region identified in other fibroins (7). Analysis of the N terminus of ECP-1 revealed the presence of 16 cysteines within the first 145 residues; these potentially could be used to form intermolecular disulfide bridges with other ECP-1 molecules or conventional fibroin molecules. This region showed no similarity to any other protein in the data base. As expected for secreted proteins, the primary sequence of ECP-1 contains an N-terminal signal sequence, with the predicted cleavage site between residues 20 and 21 (24). The putative signal sequence likely functions as a signal for the secretion of ECP-1 into the glandular lumen. Because signal sequences on secreted proteins are located on the N terminus, this observation supports our hypothesis that the primary sequence of ECP-1 is complete (Fig. 4A; see underlined region).

Two peptides sequenced by MS/MS showed 100% identity to translated regions of the ECP-1 cDNA (m/z 855.4 and 1502.7 (Table I and Fig. 4A)). Peptides with m/z 1613.8 and 3424.5, which could not be fully sequenced at their N termini by MS/MS, were both found to match peptide masses in the ECP-1 protein after theoretical digestion with trypsin (Fig. 4A). The two peptides used to design the degenerate oligonucleotides were not found as 100% matches within our translated cDNA sequence. Both peptides showed ~40–50% amino acid identity with respect to the predicted ECP-1 sequence. The peptide QQGGGF-SETLSQSDSR matched the residues QQGGFS in the translated ECP-1, whereas the peptide LLESDFGGPPIR contained identical matches to the residues ESDGF (Fig. 4A). Thus, exact matches for the sequences corresponding to the peptides at m/z 1316.7, 1623.8, and 1754.8 were not found within the predicted ECP-1 sequence. Nevertheless, ~80% of the theoretical masses of ECP-1 tryptic fragments within the mass ranging from 600 to 4000 Da were found in the peptide map produced by tryptic digestion of the 100-kDa protein species (Fig. 3, A–C, and Table II), and some peptides were further shown to be from ECP-1 by MS/MS (boldface peptide masses in Table II; spectra not shown). The peptides observed covered more than 50% of the ECP-1 protein sequence (Table II). This confirms that ECP-1 was a component of the 100-kDa protein doublet but also reinforces the idea that at least one other ECP was present, because many common peptides in the three peptide maps could not be found within the ECP-1 sequence. The predicted molecular mass of ECP-1 is 88.0 kDa, whereas the observed molecular mass is ~100-kDa. One explanation for this difference could be post-translational carbohydrate addition; other silk proteins have been shown to be subject to glycosylation (25).
An observed common feature with silk fibroins from other organisms is in codon usage bias where, in most cases, the spider selects the base A or T at the wobble position (5, 6, 26).

Analysis of the codon usage for the three most abundant amino acid residues in ECP-1, glycine, alanine, and serine, shows that all conform to this same principle (Table III). The observed bias...
at the wobble position supports a fibroin-like nature for ECP-1.

Amino Acid Composition of ECP-1 and Raw Egg Case Silk—
The predicted amino acid composition of ECP-1 differed from silks collected from *Latrodectus hesperus* (Table IV). There was slightly less serine, glutamic acid, and threonine in comparison to raw egg case silk. In addition, glycine levels were substantially higher in ECP-1 than either raw egg case silk or GdnHCl-treated egg case silk. These differences in amino acid compositions are consistent with ECP-1 representing only one of the constituents of tubuliform silks.

Expression of ECP-1 Is Predominantly Restricted to the Tubuliform Gland—To examine the mRNA levels of ECP-1, we collected total RNA from a variety of different tissues from the black widow spider for RT-qPCR analysis. Equivalent amounts of total RNA were reversed-transcribed from each tissue and were used for RT-qPCR. The tubuliform gland showed the highest level of expression, with lower levels being detected in the major and minor ampullate glands (Fig. 5). The tubuliform
gland was found to synthesize ∼5-fold more ECP-1 mRNA than the major and minor ampullate glands. The detection of the ECP-1 mRNA in the tubuliform gland supported the current hypothesis that this tissue synthesizes egg case silk components. Although the major and minor ampullate glands contained lower levels of ECP-1 mRNA, these tissues synthesized ∼3-fold more mRNA than the flagelliform, aggregate, fat and ovary tissues (Fig. 5). The low level of ECP-1 in the fat, ovaries, aggregate and flagelliform tissues likely reflected base-line levels of gene transcription. Northern blot analysis of total RNA isolated from the tubuliform gland confirmed the detection of an mRNA transcript with a size of ∼3,000 nucleotides,
which is supportive of the cDNA representing a full-length sequence (data not shown).

**ECP-1 Participates in Higher Molecular Weight Aggregates—**

The observation of 16 cysteines within the first 145 amino acids (H11011 11%) implied that the N terminus of ECP-1 may be involved in the formation of higher order structures, perhaps with other ECP-1 molecules or conventional fibroins. To examine whether ECP-1 forms disulfide bonds, we performed a time course experiment.

**TABLE I**

| M + H⁺ | De novo sequence | Location of peptide sequence |
|--------|------------------|-----------------------------|
| 855.4  | EFGSYPR          | 614–620                     |
| 1316.7 | LLESDGQPITIR     | NA                          |
| 1502.7 | TQAGCSSPFAISR    | 338–352                     |
| 1613.8 | (AAGF)SGAQSSQLQGR| 488–505                     |
| 1623.8 | SQAGSSPLQYGR     | NA                          |
| 1754.8 | QQQGFSLELSQSDSR  | NA                          |
| 3424.5 | (AVSSGIT)GNSANDAESFAASESESFAQQAAGSR | 382–417 |

**Fig. 4.** The primary sequence of ECP-1 shows similarity to spider and silk moth fibroins. A, translation of the nucleotide sequence from the ECP-1 cDNA contains an open reading frame. The longest open reading frame encodes a protein 932 amino acids in length. The conceptual translation product predicts an 88.0-kDa protein with an estimated pI of 9. One potential start codon with a good Kozak sequence was found within the ECP-1 open reading frame (31). Peptides sequences determined by MS/MS that are found within the open reading frame are indicated in boldface. In addition to the serine-rich C terminus (large box), there is a stretch of fibroin-like GA motifs and Ala tracts (both regions colored purple). Arrows indicate the initial protein segment obtained after translation of the retrieved cDNA by PCR using the degenerate oligonucleotides. Cysteine residues potentially involved in disulfide bond linkages are shown in red boldface. B, schematic representation of the primary sequence of ECP-1 demonstrates the presence of fibroin-like domains. The grid represents the amino acid sequence beginning from the N terminus. The boxes above the grid represent the areas that show similarity to fibroins identified by protein-protein BLAST searches. These regions contain poly(Ala) and poly(Gly-Ala) tracts; (GA)ₙ/Aₙ repeats have been confirmed in other silk types to form β-sheet structures (32). Several regions rich in cysteine, alanine, serine or glycine are indicated: Cys_rich, Ala_rich, Ser_rich, and Gly_rich.
**TABLE II**

Peptide fingerprints of ECP-1 found by MALDI-TOF analysis of three tryptic digests

A indicates peptide map obtained by in-solution tryptic digest of the 100-kDa protein doublet extracted for 40 h (Fig. 3A). B indicates peptide map from in-gel digest of 100-kDa proteins in 40-h extract (Fig. 3B). C indicates peptide map from in-gel digest of 100-kDa proteins in initial GdnHCl extract (Fig. 3C). Sequence coverage for A-C were 59, 52, and 51%, respectively; peptide masses shown in boldface were also sequenced by MS/MS and found to match sequences within ECP-1.

| M + H⁺ | Location in ECP-1 | Found in | M + H⁺ | Location in ECP-1 | Found in |
|--------|-------------------|---------|--------|-------------------|---------|
| 619.4 | (619.3)ᵇ | 169–173 | A, B, C | 651.3 | 247–252 | A, B, C |
| 758.4 | | 621–628 | C | 808.4 | 457–465 | C | A, B, C |
| 812.5 | | 162–168 | A, B, C | 852.4 | 329–337 | A, B, C |
| 855.4 | 614–620 | A, B, C | 901.4 | 797–806 | A, B, C |
| 1008.5 | 588–597 | A, B, C | 1086.6 | 675–688 | A |
| 1212.6 | 216–230 | A, C | 1343.7 | 253–267 | A, B, C |
| 1386.7 | 695–712 | A, B, C | 1497.7 | 781–796 | A, B, C |
| 1459.7 | 851–868 | B | 1502.7 | 338–352 | A, B, C |
| 1518.7 | 807–824 | A, B, C | 1527.7 | 598–613 | A, B, C |
| 1551.7 | 231–246 | A, C | 1613.7 | 489–505 | A, B, C |
| 1784.8 | 570–587 | A, C | 1817.9 | 268–287 | A, B, C |
| 1831.9 | 829–850 | A, B | 1852.9 | 761–780 | A, B, C |
| 1905.9 | 653–674 | A, B, C | 2366.0 | 466–488 | A, B, C |
| 2567.1 | 353–377 | A, B, C | 2870.2 | 540–568 | A, B, C |
| 2885.4 | 713–750 | A, B, C | 3243.4 | 506–539 | A, B, C |
| 3299.4 | 174–211 | A | 3424.5 | 362–417 | A, B, C |

ᵇ Location is based upon primary sequence of ECP-1 (Fig. 4).

**TABLE III**

Codon usage (%) of the three most abundant amino acids in the translated ECP-1 show codon bias at the wobble position for Gly, Ala, and Ser

| Gly | Ala | Ser |
|-----|-----|-----|
| GGA | GGT | GCT | GCA | TCT | TCA | AGT |
| 56% | 30% | 45% | 43% | 23% | 27% | 27% |

experiment with egg case proteins removed from the GdnHCl wash in the presence of dithiothreitol (Fig. 6). Longer dithiothreitol treatment with equivalent amounts of GdnHCl-washed egg case silk resulted in the accumulation of ECP-1, which is consistent with formation of disulfide bonds in the silk. Other protein bands observed in Fig. 2 are not readily observed on this gel, as the silver staining reaction was terminated prior to their detection. The lower species of the protein doublet was confirmed to contain ECP-1 by mass spectrometry after gel excision and in-gel tryptic digestion (data not shown).

**DISCUSSION**

The silk filaments produced by orb and cob weaving spiders, as well as silk moths, rank among nature’s most highly engineered structural biomaterials, displaying, in many situations, combinations of strength and toughness that exceed man-made materials. Spider silk properties can be modulated, in part, by araneoid spiders using seven distinct glands that manufacture silk fibers with diverse mechanical properties. At the molecular level, the wide functionality of silk (web construction, prey capture, reproduction, and movement) is determined to a large degree by subtle changes in amino acid sequence, composition, and mechanical processing. Much emphasis has been placed on retrieving cDNAs encoding silk proteins in an attempt to understand the relationship between primary sequence and material properties. Primary sequences of fibroins have been reported from a variety of different species, and sequence alignments have demonstrated common molecular features, which include the conservation of a nonrepetitive C-terminal region as well internal, iterated motifs containing either polyalanine (Aₓ) stretches, alternating glycine and alanine couplets (GAₓ), (GGXₓ) repeats, or glycine-proline-glycine (GPGXₓ) modules that form ensemble repeats. Analysis of the primary sequence of ECP-1 by the computer program Prosite identified two alanine-rich stretches (residues 180–326 and 661–886), two serine-rich regions (residues 342–660 and 886–928), and one glycine-rich block (amino acids 448–756) (Fig. 4B). Regions showing similarity to fibroins, as identified by protein BLAST searches, corresponded to the two alanine-rich regions and contain either Aₓ stretches or GA couplets (Fig. 4, A and B). The identities ranged in these areas from 35 to 45%, with the statistical E values ranging from 2 x 10⁻¹⁰ to 2 x 10⁻⁴¹ (Ala-rich region I spans ~160 residues; Ala-rich region II stretches ~300 amino acids). In addition to GA repeats and Aₓ stretches, analysis of the ECP-1 codon preference demonstrated bias for alanine, glycine, and serine codons preferring A or T at the wobble position, a property reported in traditional fibroins (Table III). Furthermore, removal of GdnHCl from an enriched fraction containing ECP-1 by dialysis (Fig. 6, lane 6) followed by cooling, lead to the precipitation of ECP-1, which is a phenomenon often observed with traditional silks.

Despite ECP-1 having fibroin features, ECP-1 exhibits several distinctive characteristics relative to published fibroins. For example, compared with traditional fibroins, which have large protein molecular masses (>275 kDa), ECP-1 has a relatively smaller molecular mass (~100 kDa). In addition, ECP-1 lacks the highly conserved, nonrepetitive C-terminal region found in typical spider fibroins, as well as any recognizable internal, iterated motifs containing GGX and GPGXₓ, which assemble to form ensemble repeats. Although polyalanine stretches can be found, there are fewer of these regions than in classical fibroins. Furthermore, the GA repeats found in ECP-1 are periodically interrupted with the amino acids Arg, Glu, Thr, or Ser, which is typically not observed in traditional fibroins. The significance of the interruption of the couplets with amino acids with polar side chains is currently unknown but may have a similar function as the amino acid sequence of the β-sheet-forming crystalline region of fibroins (27). Another distinctive feature of ECP-1 is the presence of 16 cysteine residues clustered near the N terminus. There is little known regarding N-terminal sequences of traditional spider fibroins. One of the only N-terminal sequences reported is from flagelliform silk (11, 28). However, the N-terminal region of ECP-1 shows no similarity to flagelliform silk; in particular, ECP-1 has a cysteine-rich N terminus, whereas flagelliform silk lacks cysteine residues in this region. Collectively, because of the distinct physical and chemical properties of ECP-1, our data indicate ECP-1 represents a new class of silk fibroins.

Although different sized egg case proteins were observed...
by the SDS-PAGE analysis, we believe ECP-1 represents an important component of the fiber for the following reasons: 1) levels of ECP-1 selectively increase following prolonged treatment with reducing agent (Fig. 6); 2) analysis of several proteins with molecular masses lower than 100 kDa show similar peptide maps relative to ECP-1, which suggests translational pausing or protein degradation in silk manufacturing (data not shown); and 3) prolonged treatment of the silk filament with GdnHCl leads to the accumulation of ECP-1 (Fig. 2). Although ECP-1 is an important constituent of egg case silk, our data also indicate that other silk proteins are present in the fiber, as the predicted amino acid composition of ECP-1 deviates from the composition of raw egg case silk (Table IV).

Elucidation of the precise structural role of ECP-1 in tubuliform silk will involve further investigation. We have shown that ECP-1 can form disulfide bonds, presumably through the N-terminal cluster of cysteine residues, which result in the formation of higher aggregate complexes (Fig. 6). The ability of the N-terminal region of ECP-1 to form disulfide bonds, coupled with two distinct regions with fibroin-like properties, implies that ECP-1 may interact with fibroins and play a structural role in egg case silk. Another possibility, given the high serine

![Fig. 6. ECP-1 forms higher ordered structures through the formation of disulfide bonds.](image)

The results are expressed as an average of multiple samples (n). Published values for dragline and egg case silk of *L. hesperus* are presented for comparison. The values are expressed per 100 residues (mol percent).

|                  | Gly | Ala | Glu | Tyr | Pro | Arg | Asp | Ser | Leu | Ile | Thr |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Predicted ECP-1  | 16.3| 23.3| 4.1 | 2.1 | 0.9 | 6.1 | 1.4 | 19.3| 2.5 | 1.9 | 3.9 |
| GdnHCl washed egg case silk of *L. hesperus* (n = 1) | 7.9 | 25.8| 10.6| 1.4 | 1.6 | 1.1 | 2.3 | 25.9| 4.9 | 2.5 | 4.9 |
| Raw egg case silk of *L. hesperus* (n = 3)*a* | 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 |
| Dragline of *L. hesperus* (n = 5)*a* | 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 |

*a* Data are from Casem et al. (20).
content, is that ECP-1 functions as a glue protein or, perhaps, as a water shell protein. However, when comparing the primary sequence of ECP-1 to sericin family members, which are well known glue proteins identified from the silk moth B. mori, ECP-1 displayed no significant similarity at the protein level. In addition, another distinctive feature of sericins is their ability to be removed by washing the cocoon silk with water (29). Several attempts to selectively remove ECP-1 from egg case silk with water treatment alone failed to remove ECP-1 (data not shown). Although spider glue proteins have not been studied extensively at the molecular level, the silk gland responsible for glue protein production has been reported to be the aggregate gland (30). Because ECP-1 is predominantly expressed in the tubuliform gland, with essentially no expression detected in the aggregate gland, ECP-1 seems unlikely to be a glue protein. Although the mRNA levels of ECP-1 were the highest in the tubuliform gland, lower levels were detected in the major and minor amputate glands. This could imply that ECP-1 has a broader function in other silk types, i.e., it may help to control the strength and elasticity of silk, depending upon its abundance.

To date no laboratory has been able to genetically engineer silks that exhibit properties identical to those found in nature. The complete elucidation of the molecular players assembled into silk, combined with refinement of silk processing procedures in vitro, should help facilitate the engineering process. One potential missing ingredient could be intermolecular linkers, which help direct the proper assembly of fibroin molecules into their natural structures. Future experiments will be directed at determining the interactions of ECP-1 with fibroins, potential conservation of ECP-1 in other species, and the influence of ECP-1 on the biomaterial properties of egg case silk.

Acknowledgments—We thank Catherine Wells and Ronda Rufsvold for their contributions with respect to the S.E.M. data.

REFERENCES
1. Farli, R. F. (1996) Biology of Spiders, 2nd Ed., pp. 1–336, Oxford University Press, New York.
2. Guerette, P. A., Ginzinger, D. G., Weber, B. H., and Gosline, J. M. (1996) Science 272, 112–115.
3. Gatey, J., Hayashi, C., Motriuk, D., Woods, J., and Lewis, R. (2001) Science 291, 2603–2605.
4. Hayashi, C. Y. (2002) Exs (Basel) 92, 209–223.
5. Xu, M., and Lewis, R. V. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 7120–7124.
6. Hinman, M. B., and Lewis, R. V. (1992) J. Biol. Chem. 267, 19320–19324.
7. Beckwith, R., and Arribas, S. (1994) J. Biol. Chem. 269, 6661–6663.
8. Arribas, S., Mello, C., Kaplan, D., Cheley, S., and Bayley, H. (1998) Appl. Microbiol. Biotechnol. 49, 31–38.
9. Lawrence, B. A., Vierra, C. A., and Moore, A. M. (2004) Biomacromolecules 5, 689–695.
10. Colgin, M. A., and Lewis, R. V. (1998) Protein Sci. 7, 667–672.
11. Hayashi, C. Y., and Lewis, R. V. (1998) J. Mol. Biol. 275, 773–784.
12. Hayashi, C. Y., Blackledge, T. A., and Lewis, R. V. (2004) Mol. Biol. Evol. 21, 1950–1959.
13. Hayashi, C. Y., Shipley, N. H., and Lewis, R. V. (1999) Int. J. Biol. Macromol. 24, 271–275.
14. Cerdales, G., Cerdales, T., Ortiz, A., and Rodriguez, O. (1983) Biochem. Biophys. Res. Commun. 116, 1033–1038.
15. Mello, C. M., Senecal, K., Yeung, B., Vouros, P., and Kaplan, D. (1994) in Silk Polymers: Materials Science and Biotechnology (Kaplan, D., Wade, W. W., Farmer, B., and Viney, eds) Vol. 544, pp. 67–79, American Chemical Society, Washington, D. C.
16. Gosline, J. M., Guerette, P. A., Ortlepp, C. S., and Savage, K. N. (1999) J. Exp. Biol. 202, 3295–3303.
17. Vollrath, F., and Knight, D. P. (2001) Nature 410, 541–548.
18. Jimenez, 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, pp. 16.4.1–16.4.5, John Wiley & Sons, Inc., New York.
19. Mitsuhashi, Y., Mugyia, M., Youngblom, J., Funes-Duran, M., Miller, R., Expeleta, J., Rigby, N., and Vierra, C. (2000) Biochim. Biophys. Acta 1492, 320–329.
20. Casem, M. L., Turner, D., and Houchin, K. (1999) Int. J. Biol. Macromol. 24, 103–108.
21. Nguyen, L., Round, J., O’Connell, R., Geurts, P., Funes-Duran, M., Wong, J., Jongeward, G., and Vierra, C. A. (2001) Nucleic Acids Res. 29, 4423–4432.
22. Sekiguchi, K. (1950) Sci. Rep. Toho Kysaku Daigaku Sect. B 8, 23–32.
23. Foradori, M. J., Kovoor, J., Moon, M. J., and Tillinghast, E. K. (2002) J. Morphol. 252, 218–226.
24. Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. (2004) J. Mol. Biol. 340, 783–796.
25. Hoffman, R., Schmidt, E. R., and Case, S. T. (1996) J. Biol. Chem. 271, 8609–8615.
26. Tian, M., Liu, C., and Lewis, R. (2004) Biomacromolecules 5, 657–660.
27. Jin, H. J., and Kaplan, D. L. (2003) Nature 424, 1057–1061.
28. Hayashi, C. Y., and Lewis, R. V. (2000) Science 287, 1477–1479.
29. Altman, G. H., Horan, R. L., Lu, H. H., Moreau, J., Martin, I., Richmond, J. C., and Kaplan, D. L. (2002) Biomaterials 23, 4131–4141.
30. Tillinghast, E. K. (1981) Naturwissenschaften 68, 526–527.
31. 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.
32. Parkhe, A. D., Seeley, S. K., Gardiner, K., Thompson, L., and Lewis, R. V. (1997) J. Mol. Recognit. 10, 1–6.
33. Draper, N., Walker, E. A., Bujalska, J. I., Tomlinson, J. W., Chalder, S. M., Arlt, W., Lavery, G. G., Bedendo, O., Ray, D. W., Laing, I., Malunowicz, E., White, P. C., Hewison, M., Mason, P. J., Connell, J. M., Shackleton, C. H., and Stewart, P. M. (2003) Nat. Genet. 34, 434–439.