Novel Molecular and Mechanical Properties of Egg Case Silk from Wasp Spider, *Argiope bruennichi*†

Ai-Chun Zhao,‡,§ Tian-Fu Zhao,‡ Koichi Nakagaki,‡ Yuan-Song Zhang,‡ Yang-Hu SiMa,‡,|| Yun-Gen Miao,‡,⊥ Kunihiro Shiomi,‡ Zenta Kajiura,‡ Yoko Nagata,¶ Masayuki Takadera,▽ and Masao Nakagaki*,‡

Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University, Ueda 386-8567, Japan, The Key Sericultural Laboratory of Agricultural Ministry, College of Sericulture and Biotechnology, Southwest University, Chong Qing 400716, China, Faculty of Agricultural Science and Technology, Suzhou University, Su Zhou 215006, China, College of Animal Sciences, Zhejiang University, Hang Zhou 310029, China, College of Science and Technology, Nihon University, Tokyo 101-8308, Japan, and Department of Kansei Engineering, Faculty of Textile Science and Technology, Shinshu University, Ueda 386-8567, Japan

Received November 25, 2005; Revised Manuscript Received January 19, 2006

ABSTRACT: Araneoid spiders use specialized abdominal glands to produce up to seven different protein-based silks/glues that have various mechanical properties. To date, the fibroin sequences encoding egg case fibers have not been fully determined. To gain further understanding of a recently reported spider silk protein gene family, several novel strategies were utilized in this study to isolate two full-length cDNAs of egg case silk proteins, cylindrical silk protein 1 (*CySp1*, 9.1 kb) and cylindrical silk protein 2 (*CySp2*, 9.8 kb), from the wasp spider, *Argiope bruennichi*. Northern blotting analysis demonstrated that *CySp1* and *CySp2* are selectively expressed in the cylindrical glands. The amino acid composition of raw egg case silk was closely consistent with the deduced amino acid composition based on the sequences of *CySp1* and *CySp2*, which supports the assertion that *CySp1* and *CySp2* represent two major components of egg case silk. CySp1 and CySp2 are primarily composed of remarkable homogeneous assemble repeats that are 180 residues in length and consist of several complex subrepeats, and they contain highly homologous C-termini and markedly different N-termini. Our results suggest a possible link between *CySp1* and *CySp2*. In addition, comparisons of stress/strain curves for dragline and egg case silk from *Argiope bruennichi* showed obvious differences in ultimate strength and extensibility, and similarities in toughness.

Silk is a proteinaceous polymer secreted by specialized exocrine glands in several groups of arthropods. Spiders are unique in their production and use of different silks throughout their lifetime, whereas insects use silk only during specific periods. Orb-web spiders are known to be able to use specialized sets of abdominal silk glands to manufacture up to seven types of silk fibers and glues (*1*). The different and unique mechanical properties of these silks enable them to be used for a variety of purposes, ranging from arresting prey to protecting offspring enclosed in the egg case. However, all of the silks characterized to date are composed almost solely of proteins having a predominance of Ala, Gly, and Ser, and undergo essentially irreversible transformation from soluble proteins into insoluble fibers. The exceptional mechanical properties (*2*) and the number of potential applications of spider silks have made them the focus of numerous molecular and structural studies (*3*−*6*).

Spider egg case silk is synthesized by the cylindrical glands (*1*, *7*, *8*), which differ from all other silk glands of spiders in that they are not used daily. Egg cocoons are produced only once or twice in the lifetime of a typical spider (*9*). The egg case of a spider serves to protect the enclosed offspring during the early stages of their lives. It must be sufficiently robust to resist such threats as predator/parasitoid invasion and temperature fluctuations (*10*−*12*). In addition, cylindrical glands are found only in female spiders and are very prominent in mature female adults but are hardly detectable in juveniles. Cylindrical silk protein synthesis is initiated in these glands at sexual maturation (*6*). The *ADF-2* clone obtained by Guerette et al. (*13*) from a cylindrical gland complementary DNA (cDNA) library of *Araneus diadematus* and *ECP-1* reported by Hu et al. (*14*) are likely to be two constituents of the egg case, but their predicted amino acid compositions differ markedly from that reported for this silk. A major component of the egg case fibers of several spiders was recently reported (*15*, *16*); however, the partial sequences only included the repetitive region and nonrepetitive carboxy terminus, lacking the amino-terminal region.

Recently, a simple method (vector-capping) for preparation of high-quality full-length cDNA libraries was developed by Kato et al. (*17*). This method comprises only three steps: (1) a process for synthesizing first-strand cDNA with reverse transcriptase using a vector primer containing a dT tail at

† The present work was supported in part by a Grant-in-aid from the Ministry of Education, Science, Sports and Culture of Japan.

* To whom correspondence should be addressed. Phone: +81-268-21-5335. Fax: +81-268-21-5331. E-mail: nakagak@giptc.shinshu-u.ac.jp.

‡ Department of Applied Biology, Shinshu University.

§ Southwest University.

|| Suzhou University.

⊥ Zhejiang University.

¶ Nihon University.

▽ Department of Kansei Engineering, Shinshu University.

10.1021/bi052414g CCC: $33.50
Published on Web 02/17/2006

one end of a linear vector; (2) a process for joining first-strand cDNA to the other end of the vector primer with T4 RNA ligase; and (3) a process for replacing mRNA with DNA. In addition, the screening of cDNA libraries based on a statistical approach was found to be feasible for cloning new genes with high expression in specific tissues during a specific period (*18*). To further elucidate the structure and function of spider egg case silk, we investigated the genes of these silk proteins using a novel procedure that combines the above methods. This is the first report of two full-length cDNAs and protein sequences of the spider egg case silk gene family. Our results provide extensive insight into the relationship between protein structure and function.

MATERIALS AND METHODS

Spider Care and Silk Gland Preparation. Mature female *Argiope bruennichi* spiders, in which the cylindrical glands are very prominent, were collected in Ueda city of Nagano prefecture, Japan, during September and October. Some of the collected spiders were dissected under a dissection microscope to harvest the silk glands. The harvested silk glands were used for RNA isolation. The remaining spiders were housed separately in plastic tanks with a cardboard frame inserted to aid web construction under ambient conditions, and they were fed 2−3 grasshoppers weekly until egg cases were constructed. Egg cases were collected on the day after construction, and egg masses were removed. Egg cases were then stored individually in closed beakers at room temperature until studied.

Scanning Electron Microscopy and Measurement of Material Properties. Egg case silk was coated to a thickness of ~3.5−10 nm with platinum−palladium alloy in an E-1010 ion sputter autocoater (Hitachi, Tokyo). Samples were observed with an S-2380N scanning electron microscope (SEM; Hitachi, Tokyo) operated with an acceleration voltage of 10 kV. The diameter of egg case silk was measured in four different SEM samples. Tests were performed at ambient temperature and humidity, which ranged from 23 to 25 °C and 30−38%, respectively.

Tensile properties of different spider silks were investigated at room temperature using a tensile tester (Tensilon UTM-111-100, Toyo Baldwin Co., Ltd.) with a gauge length of 20 mm and a strain rate of 10 mm/min. Average values from four measurements were taken to calculate the tensile modulus and tensile strength of the strands of each silk.

Amino Acid Composition Analyses. Percent amino acid composition analysis of the individual egg case outer covers was accomplished in collaboration with the College of Science and Technology of Nihon University. Protein samples were subjected to acid hydrolysis in preparation for amino acid analysis with an L-8800 amino acid analyzer (Hitachi, Tokyo).

RNA Isolation and cDNA Library Construction. Silk glands (50−100 mg) were dissected from euthanized *A. bruennichi* and then rapidly homogenized in 1 mL of ISOGEN. Total RNA and mRNA were extracted from the glands using the protocols described in the manuals for ISOGEN and Micro-FastTrack 2.0 (Nippon Gene, and Invitrogen, respectively). cDNA synthesis and construction of the cDNA library were accomplished in collaboration with Hitachi Instrument Service Co., Ltd. Strategies for cDNA library construction are described in detail by Kato et al. (*17*).

cDNA Library Screening and DNA Sequencing. ElectroMAX DH12S cells (Invitrogen Life Technologies) were transformed by electroporation, and more than 5000 recombinant plasmids were purified by alkaline lysis with SDS (*19*). The novel method used to screen the cDNA library was based on the hypothesis that the proportion of egg case fibroin gene mRNA is highest among the mRNAs from the silk glands of mature spiders. Partial sequences of about 1600 recombinant plasmids with >2 kb insert fragments were sequenced and analyzed using an ABI Prism genetic analyzer 3100. Full sequencing of the insert fragments in the clones of interest was then performed. The sequence of the cylindrical silk gene was determined by producing a series of nested deletions using a deletion kit containing exonuclease III.

Northern Blotting. Aliquots of about 5 μg of total RNA were prepared from the cylindrical glands, major ampullate glands, minor ampullate glands, and flagelliform glands of five spiders. Total RNAs together with RNA size markers were electrophoresed on a 1% agarose gel (1 × MOPS, 0.66 M formaldehyde) (*19*) and then blotted. Using the hybridization conditions given in the manual for the DIG nucleic acid detection kit (Boehringer Mannheim), we probed the Northern blots with a Digoxygenin-labeled probe. Relative positions were evaluated using RNA size markers. One probe was used in Northern blotting experiment. The probe was obtained by digesting clone *CySp1*[1] with *Sac*I. The resulting 540-bp band was purified from the gel. The obtained fragment was labeled with DIG-HIGH-Primer reagent.

Comparative Analyses. Computer analyses of DNA and amino acid sequences were conducted using the Genetyx package (Windows version, Genetyx, Inc.) and a Sequencer 4.1.4 (Demo version), respectively. Comparisons were performed through the Internet using new cylindrical gland gene sequences and previously published spider silk genes.

RESULTS

Physical Structure and Mechanical Properties of Wasp Spider Egg Case Silk. In an effort to better understand the physical structure and mechanical properties of egg case silk, we combined a mechanical testing system with an SEM to examine the physical structure and mechanical properties of egg case silk collected from wasp spider (Figure 1). The scanning electron micrograph shows that egg case silk is composed of fibers of two different diameters (Figure 1C). The larger diameter fibers, which are produced by the cylindrical silk gland, represent the major components of egg case silk. The diameter of the large fibers was 7−8 μm, whereas the diameters of the smaller fibers were approximately 600 nm. The smaller diameter fibers, which are likely secreted by the aciniform glands, are a very minor component. The mechanical properties of dragline and egg case silk from *A. bruennichi* were measured with a mechanical testing system (Figure 2). Average data (mean ± standard deviation) for the two types of silk from *A. bruennichi* are shown in Table 1. Comparisons of mean values for the silks revealed that ultimate strength and

[1] Abbreviations: Cy, cylindrical; Sp, silk protein or spidroin.

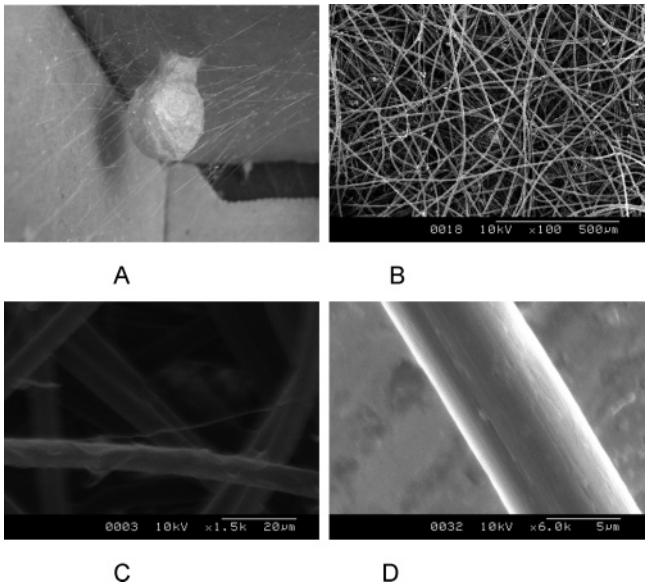


FIGURE 1: Physical structure of an egg case collected from a wasp spider. (A) photo of a fresh egg case; (B) SEM photo of egg case silk at 100× magnification; (C) SEM photo of egg case silk at 1500× magnification showing the presence of fibers of two different diameters; (D) SEM photo of egg case silk at 6000× magnification showing the diameter of the large fiber is 7−8 μm. This structure is similar to the egg case of *Argiope aurantia* (*27*).

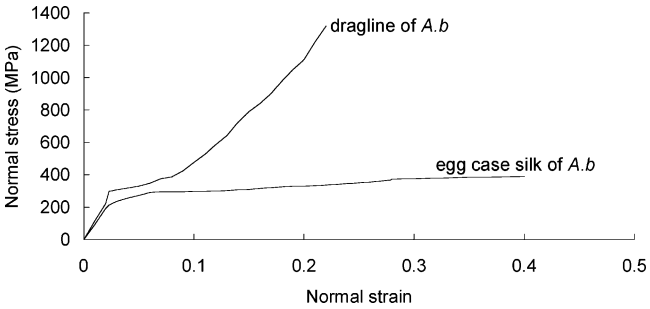


FIGURE 2: Normal stress−strain curves for dragline and egg case silk fibers from *Argiope bruennichi (A.b)*. 1 GPa = 1000 MPa.

Table 1: Comparison of Physical Properties of Egg Case and Dragline Silks from *A. bruennichi*

original	diameter (μm)	tensile breaking strain	breaking stress (GPa)	breaking energy (MJ/m^3)
egg case silk	8.04 ± 0.35	0.40 ± 0.07	0.39 ± 0.03	128.6 ± 26.6
dragline silk	3.47 ± 0.14	0.22 ± 0.04	1.32 ± 0.17	134.5 ± 50.0

extensibility of egg case silk apparently differed from the values for dragline silk, but toughness values were similar. The mechanical properties of egg case silk are inferior to those of dragline silk, which is one of the strongest biomaterials reported to date, but as with previously reported silk proteins, the mechanical properties of egg case silk may be ideal for certain biomaterial applications.

Isolating and Sequencing Cylindrical Gland cDNA Clones. More than 5000 recombinant plasmid DNAs were purified from cultures. The 5′-termini of insert fragments (of >2 kb) of about 1600 recombinant plasmids were sequenced using a specific-vector primer. Homology analysis of these sequences was conducted using Sequencing and Genetyx. Of the 1448 sequences, 453 were assigned to 31 groups and 995 could not be grouped. The number of sequences in the different groups varied from 6 to 94. The two largest groups contained 94 and 72 sequences from different recombinant clones, respectively. These two groups were analyzed further. Two recombinant clones from the two groups were completely sequenced and designated *CySp1* and *CySp2* (contraction of cylindrical silk protein or spidroin). *CySp1* and *CySp2* are 9.1 and 9.8 kb in length, respectively. Like other silk proteins, translation of the two full-length cylindrical silk protein cDNAs featured iterated repeat units flanked by nonrepetitive C- and N-termini (Figure 4). Translated Blast searches identified both repetitive and C-terminal regions among the sequences most similar to other spider silk proteins. Sequence data have been deposited with the DDBJ (accession numbers are AB242144 for *CySp1* and AB242145 for *CySp2*).

Egg Case Silk Amino Acid Composition. The predicted amino acid compositions of *CySp1* and *CySp2* were compared with the compositions obtained by direct analysis of egg case silk (Table 2). The egg case silk of *A. bruennichi* was largely composed of Ala (25.45%) and Ser (22.26%). These compositions closely matched those reported for cylindrical silk glands and egg case silk from other araneoid spiders, as indicated in Table 2. Moreover, the predicted amino acid compositions of *CySp1* and *CySp2* closely matched the compositions from *A. bruennichi* egg case silk, which indicate that *CySp1* and *CySp2* are two major cylindrical fibroin components.

CySp1 and CySp2 Are Selectively Expressed in the Cylindrical Silk Glands. The cylindrical glands are present only in females, and the development of these glands parallels the maturation of the ovaries (*6*, *15*). Thus, morphological and developmental studies suggest the cylindrical silk gland is primarily responsible for egg case silk production. We performed Northern blot analysis to examine the mRNA expression profiles of *CySp1* and *CySp2*. Northern blot analysis indicated that *CySp1* and *CySp2* are selectively expressed in the cylindrical silk glands, with no expression in the major ampullate, minor ampullate, and flagelliform glands (Figure 3). In addition, Northern blot analysis apparently demonstrated that cylindrical silk protein genes have two different transcript sizes (∼9 and 10 kb, respectively), which is consistent with the sequencing data for *CySp1* and *CySp2*.

CySp1 and CySp2 Primary Structures. 1. N-Terminal Region. The identified *CySp1* and *CySp2* sequences encode putative peptides of 2983 and 3218 amino acids, respectively (Figure 4). The first 179 residues of *CySp1* and 83 residues of *CySp2* exhibit no obvious sequence iterations. These two expanses represent the N-terminal sequences of spider egg case silk, which has only been reported for two types of spider silk to date (*18*, *20*). These regions have different amino acid compositions and sequences when compared with the repetitive portions of both *CySp1* and *CySp2*. Following these N-terminal regions, the remaining sequences of *CySp1* and *CySp2* proceed without interruption to the core repeat regions. Secretory signals are key elements of silk proteins, such as spider silks and silkworm silk, which must be transported across the endoplasmic reticulum and secreted; therefore, we analyzed the two N-terminal sequences with signalP3.0 software (eukaryotic option) (*21*) to determine whether a likely signal peptide was present. The analysis suggested a signal peptide sequence was retained in the

Table 2: Amino Acid Compositions (mol %) of Luminal Contents or Fiber of Cylindrical Glands, and Predicted Cylindrical Silk Protein Genes[a]

species source	*A. aurantia* raw egg case outer cover[b]	*A. aurantia* extracted egg case outer cover[b]	*N. clavata* luminal contents[c]	*N. edulis* luminal contents[d]	*A. diadematus* luminal contents[e]	*A. bruennichi* raw egg case outer cover	predicted *CySp1*	predicted *CySp2*	predicted silk-like repeats from *ADF-2*[f]	predicted *L. hesperus ECP-1*[g]
Ala	25.79	27.13	28.28	27.31	24.44	25.45	25.77	25.88	27.04	23.28
Arg	3.14	1.95	3.01	2.18	1.49	2.82	1.51	1.43	0.00	6.12
Asp/Asn	6.93	6.96	3.25	3.87	6.26	6.85	6.51	5.15	0.00	6.33
Cys	NA	NA	0.00	0.09	NA	0.11	0.07	0.16	0.00	1.72
Gln/Glu	7.34	7.67	8.27	8.59	8.22	7.41	7.40	7.48	12.24	7.30
Gly	10.34	8.76	8.74	8.61	8.63	11.72	9.15	8.05	47.45	16.31
His	0.25	0.19	0.00	0.32	trace	0.45	0.00	0.12	0.00	0.32
Ile	1.92	1.7	2.14	2.79	1.69	1.63	1.31	1.03	0.00	1.93
Leu	6.11	6.76	6.65	7.41	5.73	6.53	6.87	6.46	4.08	2.47
Lys	0.64	0.27	0.61	1.64	1.76	0.50	0.00	0.12	0.00	0.54
Met	0.05	0	0.00	1.44	NA	0.08	0.03	0.03	0.00	0.21
Phe	3.72	4.23	3.73	3.85	3.22	3.43	4.12	3.95	0.00	3.86
Pro	1.34	0.6	0.00	NA	0.59	0.89	0.97	0.84	0.51	0.86
Ser	22.37	24.17	23.81	20.18	27.61	22.26	25.67	28.27	4.08	19.31
Thr	4.05	3.64	6.52	5.89	3.44	3.88	4.02	3.95	0.00	3.97
Tyr	1.82	1.25	1.02	2.05	0.95	1.69	1.14	1.18	4.08	2.15
Val	4.20	4.72	3.97	2.75	5.97	4.31	5.40	5.81	0.51	3.22

[a] NA, not available. *A. diadematus* = *Araneus diadematus; A. aurantia* = *Argiope aurantia; A. bruennichi* = *Argiope bruennichi; L. hesperus* = *Latrodectus hesperus; N. clavata* = *Nephila clavata; N. edulis* = *Nephila edulis*. [b] Data from Foradori et al. (*27*). [c] Data from Zhao et al. (*18*). [d] Data from Dicko et al. (*29*). [e] Data from Andersen (*12*). [f] Accession number is AAC47009. [g] Accession number is AAX92677.

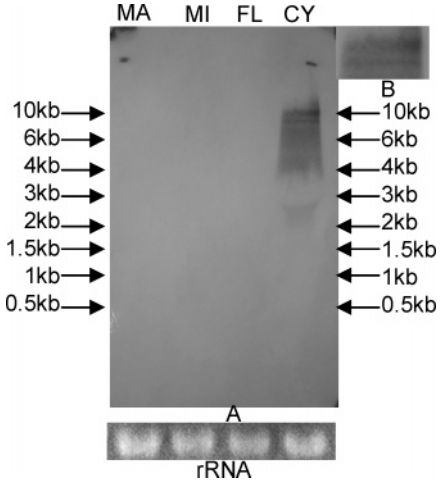
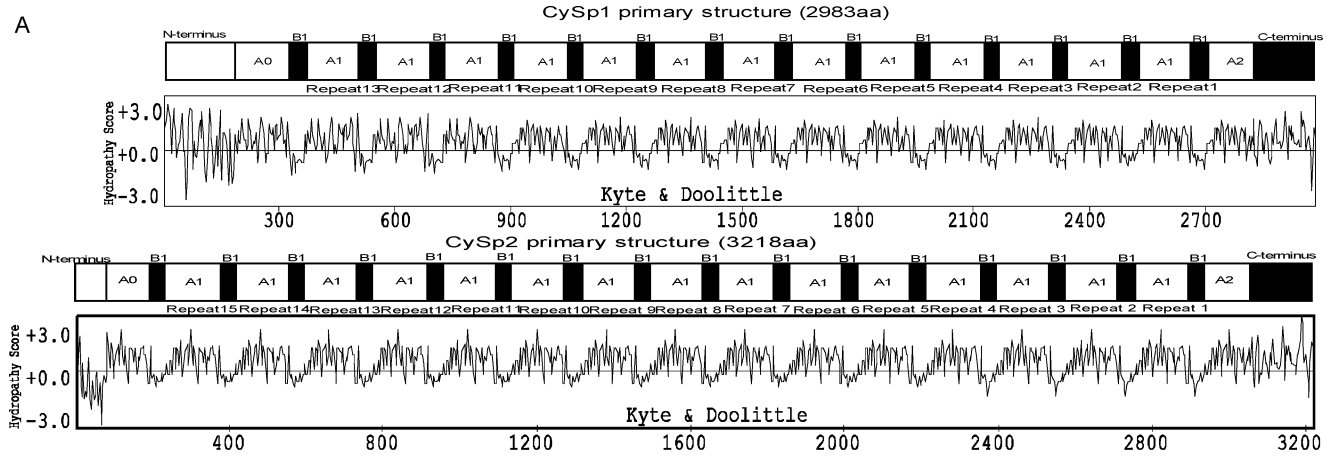


FIGURE 3: Northern blot analysis of silk gland specific distribution of *CySp1* and *CySp2*. Total RNA (5μg) from major ampullate glands (MA), minor ampullate glands (MI), flagelliform glands (FL), and cylindrical glands (CY) was probed with a 540-bp repetitive sequence. Panel B shows an amplified photo of the bands in panel A. Ribosomal RNA stained with ethidium bromide was used to confirm RNA integrity and relative quantity. The sizes of RNA markers are shown by arrows.

CySp1 sequence, while there were no apparent signaling sequences in *CySp2*. The same analysis for the two flagelliform sequences (AAC38846 and AAF36091) was performed, and like *CySp2*, the results suggest that the two flagelliform sequences do not possess any obvious signal peptides.

Finally, we aligned the N-terminal sequences from *CySp1* and *CySp2* found in this study with those of only three other spider sequences, which reflects the challenges posed by the cloning and sequencing of silk sequences, as truncation of longer cDNA partial clones from the spider *Nephila clavipes* is routinely observed in *Escherichia coli* (*22*). Analysis showed the N-terminal sequences of *CySp1* from two different spiders are highly homologous (65% identity), and they are also considerably homologous (71% homology in a continuous 83-aa region) to those (115 aa) of the two flagelliform silks, which are 98% identical each other. Surprisingly, the N-terminal sequence of *CySp2* shared little homology with the other proteins (Figure 5).

2. Core Repetitive Region. The deduced molecular masses were 277 kDa for *CySp1* and 298 kDa for *CySp2*, and about 90% of the two molecules consist of repeats. The two molecules are comprised of predominantly hydrophobic and neutral residues, and only about 3% of the residues are hydrophilic. Comparison of the repetitive sequences of the two molecules revealed that they were nearly indistinguishable. The repetitive regions of *CySp* (a generic term for *CySp1* and *CySp2* in this paper) have two characteristics, which differ from several previously reported spider silk proteins (*20*, *23*−*25*): (1) four short motifs, A_n, $(GA)_n$, GGX, and GPGGX, are scarcely present in the core repetitive regions of *CySp*, and several complex repetitive units are seen (Figure 4B) and (2) unequal allelic recombination, which has been reported in several spider and silkworm silks, are not apparently generated in *CySp* because the assemble repeat units in *CySp* lack any variability in length and marked conservation of the sequences is seen. All of the sequences of the various assemble repeat units of *CySp* are as long as 180 aa, excluding Repeat 11 (181 aa) in *CySp1*, and differ by only 24 residues from one another (data are shown in Supporting Information, Figure 6). In addition, the repetitive nature of *CySp* sequences with alternating hydrophilic and hydrophobic blocks can be visualized using hydrophathy plots (Figure 4A). On the basis of the Kyte−Doolittle plots, the core repetitive regions are composed of four types of complex repeat (Figure 4A); Repeats A1 and B1 consist of 139 and 41 residues, respectively. Furthermore, repeats A0 and A2 are also major portions of repeat A1, but only represent the two ends of the repetitive regions. The proportion of hydrophobic residues in repeat A1 is up to 60%, while that of hydrophilic and neutral residues in repeat



B

CySp1:

1 ***MVWLTSIAFLVGFLGAVSS***QSVAVTAVPSVFSSPNLASGFLQCLTFGIGNSPAFPTQEQQDLDAIAQVILNAVSSNTGATAS **N-terminus**

83 ARAQALSTALASSLTDLLIAESAESNYSNQLSELTGILSDCFIQTTGSDNPAFVSRIQSLISVLSQNADTNIISSAGIPSVSGR

166 RGAGGLGFDNTAR----------

905 AAAQASASSFAQASSASLAASSAFSSAFSSANSLSALGNVGYQLGFNVANTLGIGNAPGLGNALSQAASSVGVGASSSA **Repeat 10**

983 YANAVSNAVGQFLAGQGVLNAGNAGSLASSFANALSASAASVASSAAAQSAFQSQAAASAFSRAASQSASRGAARSGA

1060 QSSSTTTTTSTSGSQAASQSASS----------

2821 NSALSVGSRVSSPSYGALSPIAAGPNFISTGLNVGGASVGGPFDSLSQSLPTSLQTALAPIVSSSGLGSSAATARVSSLAN **C-terminus**

2901 SFASAISSSGGSLSVPTFLNLLSSVGAQVSSSSSLSSLEVTNEVLLEAIAALLQVINGGSITSVDLRYVPNAQQDLVNALSG*

CySp2:

1 MPFRSSVVTISCSYRRSLSLQQSKFGQRFPSVSHIWNRKFSRISYSRTTRLGCHCPGDTQCRFKQHWRHSIGQSSGFKY **N-terminus**

80 SACI----------

1136 AASQASASSFAQASSASLAASSSFSSAFSSANSLSALGNVAYQLGFNVANTLGIGNAPGLGAALSQAVSSVGVGASSSTY **Repeat 10**

1215 ANVVSNAVGQFLAGQGVLNAANAGSLASSFASALSASAASVASSAAAQSASQSQAAASAFSRAASQSASQSAARSGAQ

1292 SSSTTTTTSASGSQAASQSASS----------

3052 NSALSIGSRVSSPSYGVFSPIAAGPNSISTGLNVGGASIGGPFATLSQSLPTSLQTALAPIVSSSGLGSSAATARVSSLANSI **C-terminus**

3134 ASAISSSGGSLSVPTFLNLLSSIGAQVSSSSSLSSSSEVTTQVLLEAIAALLQVINGAQITSVNFSNVSNVNRALVDSLVGSF

A*

FIGURE 4: Molecular structures of *CySp1* and *CySp2*. **(**A) Schematic diagrams and the corresponding Kyte−Doolittle hydropathy plots of *CySp* primary structures. The numbers on the plots indicate the positions of amino acid residues deduced from *CySp* primary sequences. Repeat A0, A2, A1, and the N-termini are represented with white tetragonal blocks, while Repeat B1 and the C-termini are represented with black ones. Full individual assemble repeat units starting from the C-termini are numbered between schematic diagrams and the corresponding Kyte−Doolittle hydropathy plots. (B) The primary sequences of the N-termini, a full assemble repeat unit, and the C-termini of *CySp*. The sequences of the N-termini, a full assemble repeat unit, and the C-termini of *CySp* were divided by continual dashes. The presumed signal peptide is shown as bold italics. Cys residues represented in the primary sequences are highlighted in black block. In full assemble repeat units, the sequences of Repeat B1 are underlined; the remainder is the sequences of Repeat A1. The motifs, which could contribute to β-sheet secondary structures of egg case silk, are indicated in tetragon. The * is a stop codon.

B1 is more than 70%. In the two Kyte−Doolittle plots, the spikes of repeat B1 are largely below the abscissa axis, while those of repeat A1 are largely above (Figure 4A).

3. C-Terminal Region. The carboxy-terminal sequences from several orb-web spider silk proteins and *CySp* were used for multialignment and systematic analysis (see Supporting Information, Figure 7). On alignment, these sequences fall into separate groups, depending on the type of silk. The termini of cylindrical silk of different spiders are well conserved. Systematic analysis shows that TuSp1 or *CySp* groups together, while MiSp1 clusters with MiSp2, MaSp1 forms a clade with MaSp2, and Flag also forms a clade with AcSp1. This suggests that *CySp* or TuSp1 represents a new member of the spider silk gene family or an orthologous gene group. In addition, systematic analysis of C-terminal sequences indicates that *A. bruennichi* and *A. aurantia* group tightly and then form a clade with *Nephila clavata*, which is consistent with previous results that *A. bruennichi* and *Argiope aurantia* are more closely related than either is to *Nephila* (*26*).

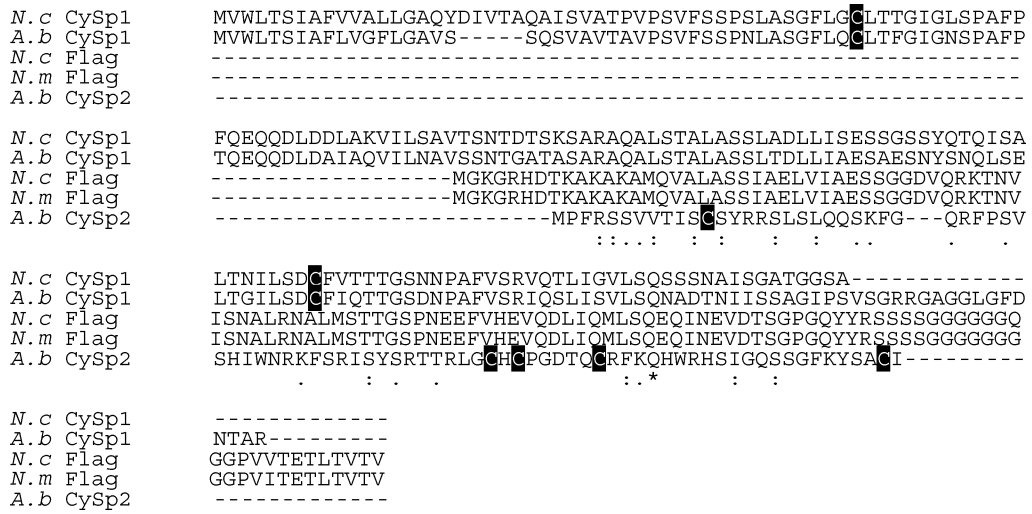


FIGURE 5: CLUSTAL W alignment of the N-terminal sequences of several spider silk proteins. At the bottom of the alignment, asterisks represent identical residues, one dot (.) represents conserved substitutions, and two dots (:) represent semiconserved substitutions. *N.* and *A.* in this figure are the abbreviations of *Nephila* and *Argiope*, respectively. The accession numbers are AB218974 for *Nephila clavata (N.c) CySp1*, AAC38846 for *Nephila clavipes (N.c) Flag*, and AAF36091 for *Nephila madagascariensis (N.m) Flag*. *A.b* is an abbreviation of *Argiope bruennichi.*

DISCUSSION

Our full-length cDNA clones of the two cylindrical silk proteins were obtained by sequencing randomly selected clones after confirming the inserts using a statistical approach. The cylindrical glands are much more prominent than other glands in mature female *A. bruennichi* spiders, which suggests that the cylindrical silk protein genes are very strongly expressed in the cylindrical glands during this period. Based on this characteristic of cylindrical glands, we supposed that cylindrical silk protein mRNA would account for the largest portion of transcripts in the silk glands of mature female spiders, which was confirmed by the research of Garb and Hayashi (*16*). Our research has shown that the present procedures are feasible in identifying silk gene transcripts from cDNA libraries (*18*). Because the harsh denaturing conditions required for successful second-strand cDNA synthesis of lengthy silk gene transcripts severely limit full-length cDNA products (*20*), most silk protein genes reported to date are partial sequences that only include part of the repetitive region and nonrepetitive C-terminus, and lack the N-terminal region. In this study, we obtained two full-length cylindrical silk protein cDNA clones using a novel procedure. In addition, sequencing results showed that the two cylindrical silk genes are 9.1 and 9.8 kb in length, which is consistent with the results of Northern blot analysis (Figure 3).

The amino acid compositions deduced from the *CySp1* and *CySp2* sequences were compared with the data obtained from direct measurements of egg case silk protein from *A. bruennichi* (Table 2). Generally, the data are consistent, which indicates that *CySp1* and *CySp2* are two major components of cylindrical silk protein. However, slightly lower Gly levels and higher Ser levels are observed in the deduced amino acid compositions when compared with the silk from *A. bruennichi* (Table 2). These slight differences may have been due to the presence of other proteins in the glands. ADF-2 (*13*) and ECP-1 (*14*) were reported to be two likely constituents of the cylindrical silk protein. Their predicted amino acid compositions differ markedly from those reported for this silk, but their high Gly and low Ser contents are clear (Table 2). It has also been reported that MaSp1 and MaSp2 are expressed in cylindrical glands (*16*). All of these proteins are characterized by higher Gly and lower Ser levels (*18*). It is also possible that the outer cover of the egg case contains other types of silk from different silk glands. We found that slightly lower Gly levels and higher Ser levels are observed in luminal amino acids or among amino acids from the extracted outer egg case when compared with the corresponding raw egg case outer cover (Table 2). Small diameter fibers, which are likely from aciniform glands, were found on structural observation of the egg case. It has also been reported that fibers from piriform glands certainly occur in the egg case (*27*). Finally, complete acid hydrolysis of egg case silk protein may partially destroy Ser, which would lead to the decreased concentration of this amino acid.

Spider silks reported previously are renowned for their ability to absorb energy without failing (toughness). Our characterization of the modulus, strength, extensibility, and toughness of dragline silk from *A. bruennichi* was similar to the average values obtained in previous studies of major ampullate silks from closely related orb weavers, such as *Argiope trifasciata* and *Araneus diadematus* (Table 3). The ultimate strength of *A. bruennichi* cylindrical silk is about 1/3 that of its dragline silk, but the extensibility of cylindrical silk is about twice that of dragline silk. The toughness of cylindrical silk was also remarkable and was similar to its dragline silk.

Few structural analyses have been performed on cylindrical silk primarily due to the lack of full sequence information. The secondary structures of egg case silk, which were detected by fiber X-ray diffraction, contain β-sheets with a larger b dimension of X-ray diffraction unit cell and lack of orientation parallel to the fiber axis (*28*). Furthermore, side group spacing was defined for major ampullate (polyAla, 10.77 Å), minor ampullate ($(GA)_n$, 10.45 Å), and egg cocoon silks (16.2 Å) (*28*). Recently, the results from Fourier self-deconvolution (FSD) of Fourier transform infrared (FTIR)

Table 3: Comparisons of Mechanical Properties of Spider Silk[a]

material	Young's modulus (GPa)	strength (MPa)	extensibility	toughness (MJ/m^3)
A. bruennichi dragline silk	11.8	1320	0.22	134.5
A. bruennichi cylindrical silk	9.1	390	0.40	128.6
A. trifasciata major ampullate silk[b]	9.3	1290	0.22	145
A. trifasciata minor ampullate silk[b]	8.5	342	0.54	148
A. trifasciata aciniform silk[b]	9.6	687	0.86	367
A. diadematus major ampullate silk[c]	10	1100	0.27	160
A. diadematus flagelliform silk[c]	0.003	500	2.7	150

[a] *A. bruennichi* = *Argiope bruennichi*; *A. trifasciata* = *Argiope trifasciata*; *A. diadematus* = *Araneus diadematus*. [b] Data from Hayashi et al. (*37*). [c] Data from Gosline et al. (*34*).

spectra suggested that β-sheets exist in the cylindrical gland and that β-sheets account for 5% of the secondary structure (*29*). More recently, NMR spectroscopy showed that approximately 70% of Ala present in the *Latrodectus hesperus* TuSp1 sequences are present in β-sheet secondary structures (*15*). The complex repetitive architecture of *CySp1* and *CySp2*, which contain more amino acids with large side chains than other silk proteins, could better explain the larger b dimensional value and lack of orientation parallel to the fiber axis in β-sheets when compared with major and minor ampullate silks. On the basis of these results, short polyAla blocks and neighboring scattered Ala residues, such as AAAQASA and AAASA motifs (Figure 4B), within the primary sequences of *CySp1* and *CySp2* could contribute to β-sheet secondary structures in egg case silk. In comparison to GA repeats and polyAla regions in minor ampullate, major ampullate, and silkworm silks, the AAAQASA and AAASA motifs contain larger side chains and more hydrophilic Gln and Ser residues, which is consistent with much larger side chain spacing and a lack of orientation in the β-sheets. Gln and Ser have larger side chains than Ala, and thus, the side chain spacing of the β-sheet formed by AAAQASA and AAASA motifs must be larger than that of polyAla; on the other hand, Gln and Ser have poor hydrophobicity, while Ala has good hydrophobicity, and this may result in nonorientation of the β-sheet in egg cocoon silks due to a lack of hydrophobic force to hold the sheet parallel to the fiber axis (*30*, *31*).

Furthermore, the function of the small peptide motifs (GGX, A_n/$(GA)_n$, or GPGXX) have been widely studied, as they are present in the primary sequences of most of silk proteins characterized to date. On the basis of previous physical studies, the peptide motifs have been assigned different structural roles (*20*, *28*, *32*). The GGX motif probably forms a helix with three amino acids per turn, called a 3_{10} helix (*33*). The $(GA)_n$/A_n motifs have been found to form a β-sheet (*30*, *33*). The GPGXX motif repeat unit has been suggested to be involved in β-turn spirals (*30*). Recent results have shown that $(GA)_n$/A_n motifs may form a very short crystal-forming domain, and the remainder of repeat motifs (GGX, GPGXX, and spacer) may direct the formation of amorphous domains within the core repetitive regions of different silk proteins (*34*). The crystal-forming domain would mainly contribute to the strength of silk protein, while the amorphous domains would mainly contribute to the extensibility of silk proteins (*34*). As a result of the inability to detect any protein secondary structure other than β-sheet, several groups have suggested that other motifs excluding $(GA)_n$/A_n form random coils (*35*). When combined with β-sheet regions as anchors, the random-coil segments would provide elasticity.

The repetitive sequences of cylindrical silk proteins are different from most sequences previously characterized in araneoid silk proteins (Table 4). However, on the basis of the analysis above, we speculate that the considerable strength of egg case silks could result from the presence of β-sheets, and that the excellent extensibility of egg case silks could partially be due to large-scale complex repetitive blocks with anomalous amino acid sequences, such as the amorphous regions and spacers with more complex amino acid compositions seen in silkworm and spider silks, respectively (*20*, *36*). When compared with dragline silk proteins expressed by the major ampullate silk glands, cylindrical silk proteins contain lower proportions of crystal-forming sequence elements, which is consistent with the cylindrical silk having lower strength when compared with the corresponding

Table 4: Comparison of the Core Repeat Sequences of Spider Silks and Silkworm Silk

original (access no.)	primary repetitive motifs	organization of core repeat sequences	length of an assemble repeat unit (aa)
A. bruennichi CySp1	devoid	$A0B1(A1B1)_{13}A2$	180
A. bruennichi CySp2	devoid	$A0B1(A1B1)_{15}A2$	180
A. trifasciata AcSp1 (AAR83925)	devoid	tandem of high conservation 200aa-long repeat units	200
N. clavipes MaSp1 (AAA29380)	GGX, $(GA)_n$, A_n	$GAG(A)_{6-7}GGA$, GQ GX $(GGX)_n$	19−46
N. clavipes MaSp2 (AAA29381)	GPGXX, $(GA)_n$, A_n	$(GPGXX)_{2-7}$, GGX,$(GPGXX)_n$, A_{6-10}	31−49
N. clavipes MiSp1 (AAC14589)	GGX, $(GA)_n$, A_n, minor spacer	minor spacer, $(GGX)_{2-3}$, $(GA)_{3-4}$, A_5,	not available
N. clavipes MiSp2 (AAC14591)	GGX, $(GA)_n$, A_n, minor spacer	minor spacer, $(GA)_n$, $(GGX)_n$, A_n	not available
N. madagascariensis Flag (AAF36091 and AAF36092)	GPGXX, GGX, Flag spacer	$(GP(GGX)_{1-4}Y)_n$, $GRGGA_n$, GGX_n, Flag, spacer	368−411
Bombyx mori H-fibroin (AAF76983)	$(GAGAGS)_n$, $(GX)_n$, amorphous regions	$(GAGAGS)_{5-15}$, $(GX)_{5-15}$, GAAS, amorphous regions	194−641

dragline silk. On the other hand, cylindrical silk proteins containing higher proportions of amorphous sequence elements possess greater elasticity arising from the large-scale extension of these coiled amorphous chains.

Recent investigations have discovered a spider silk protein composed of highly homogeneous ensemble repeats largely lacking the amino acid motifs seen in other silk proteins (*37*). *CySp1* and *CySp2* appear to be similar to this recently discovered spider silk protein, being composed of complex, nearly identical ensembles that are devoid of GGX, A_n, $(GA)_n$, and GPGXX motifs (Table 4). Unequal allelic recombination in spider silk genes (*38*) and the length variation of *Bombyx mori* H-fibroin (*39*) demonstrate the importance of this mechanism in silk genes with short repetitive motifs, but larger DNA blocks without intrinsic repeats are less likely to be internally misaligned during replicating and crossing over (*18*). This conclusion is further supported by the primary structures of *CySp* and *AcSp1*, which have nearly identical ensemble repeats.

In addition, secretory signals are key elements of silk proteins, such as spider and silkworm silks, which must be transported across the endoplasmic reticulum and secreted. Surprisingly, our investigation showed that the *CySp2* identified in this study did not possess a signal peptide sequence. Similar results have been reported for flagelliform silk sequences (*18*, *40*). Furthermore, recent analysis of the N-terminal sequences of 270 mature secreted and cell-surface proteins by Zhang and Henzel (*41*) showed that computational predictions are only 78.1% accurate when compared with experimental data; therefore, it is possible that these proteins, including *CySp2*, fall into the 'nonpredictable' category. In addition, the C-termini of silks from orb-weaver spiders contain one or two Cys residues whose positions are well-conserved within flagelliform silks and dragline silks (see Supporting Information, Figure 7). It has been suggested that the Cys in the C-terminal end of these spider silks may function in dimer formation or linkage between spidroin I and spidroin II (*40*), which is similar to the linkage between the H-Chain and L-Chain via one of the Cys residues present in the C-termini of lepidopteran silk (*42*). Interestingly, the C-termini of aciniform and cylindrical silks sequences from two recently characterized spider gene families contain no Cys (see Supporting Information, Figure 7), but the N-termini of *CySp1* and *CySp2* reported in this study contain several Cys residues, five in *CySp2* and two in *CySp1* (Figure 4B). The function of N-terminal Cys is uncertain, but it is possible that N-terminal Cys residues have a role on linkage between *CySp1* and *CySp2*, similar to the function of C-terminal Cys residues seen in other silks. The *CySp1*−*CySp2* complex, which would possess a signal peptide, would then be transported across the endoplasmic reticulum and secreted to the Golgi.

In summary, characterization of two full-length cylindrical silk proteins, *CySp1* and *CySp2*, has resulted in several new findings. First, sequencing and amino acid composition analysis suggest that *CySp1* and *CySp2* are two major components of egg case silk. Furthermore, other minor proteins are present in egg case silk and cylindrical glands (Table 2). Second, the molecular structures of cylindrical silk proteins could contribute to further understanding of its unique molecular and mechanical properties. The molecular characteristics of the N-termini of the new spider gene family are particularly interesting. Third, a novel procedure was used to successfully obtain full-length cDNA products of highly repetitive and lengthy spider silk genes. This technique facilitates the full characterization of other spider silks of interest. Finally, egg case silk protein may provide a new biomaterial source for medical applications and scaffolds for tissue engineering, as it is characterized by its hydrophobic nature (*CySp* contains about 3% hydrophilic residues), good air permeability, heat preservation, and moisture retention.

ACKNOWLEDGMENT

The present work was supported in part by a Grant-in-aid from the Ministry of Education, Science, Sports and Culture of Japan. We thank two anonymous reviewers for valuable comments on the paper.

SUPPORTING INFORMATION AVAILABLE

The comparative and systematic analyses of repeat unit and C termini sequences. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Lucas, F. (1964) Spiders and their silk, *Discovery 25*, 20−26.
2. Gosline, J. M., DeMont, M. E., and Denny, M. W. (1986) The structure and properties of spider silk, *Endeavour 10*, 37−43.
3. Candelas, G. C., Candelas, T., Ortiz, A., and Rodriguez, O. (1983) Translational pauses during a spider fibroin synthesis, *Biochem. Biophys. Res. Commun. 116*, 1033−1038.
4. Candelas, G. C., and Cintron, J. (1981) A spider fibroin and its synthesis, *J. Exp. Zool. 216*, 1−6.
5. Candelas, G. C., and Lopez, F. (1983) Synthesis of fibroin in the cultured glands of *Nephila clavipes*, *Comp. Biochem. Physiol. 74B*, 637−641.
6. Candelas, G. C., Ortiz, A., and Molina, C. (1986) The cylindrical or tubiliform glands of *Nephila clavipes*, *J. Exp. Zool. 237*, 281−285.
7. Peters, V. H. M. (1955) Uber der spinneraparat von *Nephila madagascariensis*, *Z. Naturforsch. 10B*, 395−404.
8. Comstock, J. H. (1948) *The Spider Book*, Comstock Publication Co., Ithaca, New York.
9. Austin, A. D. (1985) The function of spider egg sacs in relation to parasitoids and predators, with special reference to the Australian fauna, *J. Nat. Hist. 19*, 359−376.
10. Hieber, C. (1985) The "insulation" layer in the cocoons of *Argiope aurantia* (Araneae: Araneidae), *J. Therm. Biol. 10*, 171−175.
11. Hieber, C. (1992) Spider cocoons and their suspension systems as barriers to generalist and specialist predators, *Oecologia 91*, 530−535.
12. Andersen, S. (1970) Amino acid composition of spider silks, *Comp. Biochem. Physiol. 35*, 705−711.
13. Guerette, P., Ginzinger, D., Weber, B., and Gosline, J. (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family, *Science 272*, 112−115.
14. 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.
15. 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.
16. 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.
17. Kato, S., Ohtoko, K., Ohtake, H., and Kimura, T. (2005) Vector-capping: a simple method for preparing a high-quality full-length *cDNA* library, *DNA Res. 12*, 53−62.
18. Zhao, A., Zhao, T, Sima, Y., Zhang, Y., Nakagaki, K., Miao, Y., Shiomi, K., Kajiura, Z., Nagata, Y, and Nakagaki, M. (2005)

Unique molecular architecture of egg case silk protein in a spider, *Nephila clavata*, *J. Biochem. 138*, 593−604.
19. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, New York.
20. 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.
21. Bendtsen, J. D., Nielsen, H., Heijne, G. V., and Brunak, S. (2004) Improved prediction of signal peptides: SignalP 3.0, *J. Mol. Biol. 340*, 783−795; (http://www.cbs.dtu.dk/services/SignalP/).
22. Arcidiacono, S., Mello, C., Kaplan, D., Cheley, S., and Bayley, H. (1998) Purification and characterization of recombinant spider silk expressed in *Escherichia coli*, *Appl. Microbiol. Biotechnol. 49*, 31−38.
23. Xu, M., and Lewis, R. (1990) Structure of a protein superfiber: spider dragline silk, *Proc. Natl. Acad. Sci. U.S.A., 87*, 7120−7124.
24. Hinman, M., and Lewis, R. (1992) Isolation of a clone encoding a second dragline silk fibroin, *J. Biol. Chem. 267*, 19320−19324.
25. Colgin, M., and Lewis, R. (1998) Spider minor ampullte silk proteins contain new repetitive sequences and highly conserved non-silk-like "spacer regions", *Protein Sci. 7*, 667−672.
26. Coddington, J., and Levi, H. (1991) Systematics and evolution of spiders (Araneae), *Annu. Rev. Ecol. Syst. 22*, 565−592.
27. 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.
28. Parkhe, A. D., Seeley, S. K., Gardner, K., Thompson, L., and Lewis, R. V. (1997) Structural studies of spider silk proteins in the fiber, *J. Mol. Recognit. 10*, 1−6.
29. Dicko, C., Knight, D., Kenney, J. M., and Vollrath, F. (2004) Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. Temperature and concentration effects, *Biomacromolecules 5*, 2105−2115.
30. 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.
31. Vollrath, F., and Knight, D. P. (2001) Liquid crystalline spinning of spider silk, *Nature 410*, 541−548.
32. Simmons, A. H., Michal, C. A., and Jelinski, L. W. (1996) Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk, *Science 271*, 84−87.
33. Dong, Z., Lewis, R. V., and Middaugh, C. R. (1991) Molecular mechanism of spider silk elasticity, *Arch. Biochem. Biophys. 284*, 53−57.
34. Gosline, J. M., Guerette, P. A., Ortlepp, C. S., and Savage, K. N. (1999) The mechanical design of spider silks: from fibroin sequence to mechanical function, *J. Exp. Biol. 202*, 3295−3303.
35. Temonia, Y. (1994) Molecular modeling of spider silk elasticity, *Macromolecules 27*, 7378−7381.
36. 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.
37. 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.
38. Beckwitt, R., Arcidiacono, S., and Stote, R. (1998) Evolution of repetitive proteins: spider silks from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae), *Insect Biochem. Mol. Biol. 28*, 121−130.
39. Manning, R. F., and Gage, L. P. (1980) Internal structure of the silk fibroin gene of *Bombyx mori. II*. Remarkable polymorphism of the organization of crystalline and amorphous coding sequences, *J. Biol. Chem. 255*, 9451−9457.
40. Bini, E., Knight, D. P., and Kaplan, D. L. (2004) Mapping domain structures in silks from insects and spiders related to protein assembly, *J. Mol. Biol. 335*, 27−40.
41. Zhang, Z., and Henzel, W. J. (2004) Signal peptide prediction based on analysis of experimentally verified cleavage sites, *Protein Sci. 13*, 2819−2824.
42. Inoue, S., Tanaka, K., Arisaka, F., Kimura, S., Ohtomo, K., and Mizuno, S. (2000) Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio, *J. Biol. Chem. 275*, 40517−40528.

BI052414G