Molecular Characterization and Evolutionary Study of Spider Tubuliform (Eggcase) Silk Protein†

Maozhen Tian* and Randolph V. Lewis

Department of Molecular Biology, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071

Received February 27, 2005; Revised Manuscript Received April 5, 2005

ABSTRACT: As a result of hundreds of millions of years of evolution, orb-web-weaving spiders have developed the use of seven different silks produced by different abdominal glands for various functions. Tubuliform silk (eggcase silk) is unique among these spider silks due to its high serine and very low glycine content. In addition, tubuliform silk is the only silk produced just during a short period of time, the reproductive season, in the spider's life. To understand the molecular characteristics of the proteins composing this silk, we constructed tubuliform-gland-specific cDNA libraries from three different spider families, *Nephila clavipes*, *Argiope aurantia*, and *Araneus gemmoides*. Sequencing of tubuliform silk cDNAs reveals the repetitive architecture of its coding sequence and novel amino acid motifs. The inferred protein, tubuliform spidroin 1 (TuSp1), contains highly homogenized repeats in all three spiders. Amino acid composition comparison of the predicted tubuliform silk protein sequence to tubuliform silk indicates that TuSp1 is the major component of tubuliform silk. Repeat unit alignment of TuSp1 among three spider species shows high sequence conservation among tubuliform silk protein orthologue groups. Sequence comparison among TuSp1 repetitive units within species suggests intragenic concerted evolution, presumably through gene conversion and unequal crossover events. Comparative analysis demonstrates that TuSp1 represents a new orthologue in the spider silk gene family.

Silk has been used by arachnids as construction material for over 400 million years (*1*). Spiders are unique in their production and use of different silks throughout their lifetime, whereas insects use silk only during specific periods. Orb-web-weaving spiders are known to be able to produce up to seven different protein-based silk fibers. Of the seven silks, major ampullate silk (dragline and web frame silk), minor ampullate silk (auxiliary spiral), and flagelliform silk (core capture fiber) are known for their unique mechanical properties (*2*). Studies on these silk proteins have shown that the conserved sequence motifs correspond to their specific mechanical properties of strength and elasticity (*3*). In contrast, the great sequence diversifications of the known silk proteins from non-orb-web-weaving spiders suggest that the silk protein design in spiders is far more complicated than suggested by the orb-weaver silk proteins (*2*).

Studies from the sequenced spider silk cDNAs, genomic DNAs, or polymerase chain reaction products indicate that silk genes have a uniform molecular organization (*4*−*9*). They are large in size and contain nonrepetitive amino and carboxy termini with the repetitive sequences in the middle (*4*−*9*). Extensive differences have been observed in the repetitive sequence among paralogous genes in silk gene family members, while the carboxy terminus shows high sequence similarity among both paralogous and orthologous genes (*8*, *9*).

Four spider silks, major ampullate, minor ampullate, flagelliform, and aciniform silk, have been characterized at the DNA level (*2*, *4*−*8*). Four amino acid motifs, polyAla[1] (A_n), $(GA)_n$, Gly-Pro-Gly-X_n, and $(Gly\text{-}Gly\text{-}X)_n$, make up most of the silk proteins predicted from their DNA sequences in major ampullate, minor ampullate, and flagelliform silk (*2*, *9*, *10*). However, these motifs are poorly represented in aciniform silk proteins (*8*). In all cases, the amino acid compositions of the silks match those of the predicted proteins (*4*−*9*). The DNA or protein sequences of tubuliform (eggcase or cocoon silk), aggregate (viscous glue), and pyriform (attachment disk) silk have not been characterized as yet.

Tubuliform silk, also known as eggcase silk, is unique among all the known silks in that it is the only silk produced by the female spider during a single period in a spider's lifetime, the reproductive season. Studies show that tubuliform glands undergo structural and morphological changes just prior to egg laying and eggcase forming (*11*). Tubuliform silk is also atypical in that it has a very low glycine but high serine content. Tubuliform silk has a tensile strength (1.3×10^9 N m^{-2}) and elasticity (5%) similar to those of minor ampullate silk (1.6×10^9 N m^{-2}, 5%) (*12*). However, what makes it different from other silks is its low ability to

† This work is supported by a grant from the NSF (LOC4408) to R.V.L.

* To whom correspondence should be addressed. Phone: (307) 766-6380. Fax: (307) 766-5098. E-mail: tianmz@uwyo.edu.

[1] Abbreviations: MaSp1, major ampullate silk protein 1; MaSp2, major ampullate silk protein 2; MiSp1, minor ampullate silk protein 1; MiSp2, minor ampullate silk protein 2; Flag, flagelliform silk protein; AcSp1, aciniform silk protein 1; TuSp1, tubuliform silk protein 1; G/Gly, glycine; A/Ala, alanine; I, isoleucine; L, leucine; V, valine; Y, tyrosine; F, phenyalanine; T, threonine; S/Ser, serine; D, aspartate; E, glutamate; N, asparagine; Q, glutamine.

10.1021/bi050366u CCC: $30.25
Published on Web 05/06/2005

withstand bending before breaking, a glasslike or crystalline behavior (*12*). On the basis of its unique features, it seems likely that new amino acid motifs are present in tubuliform silk protein. To expand our knowledge on spider silk gene family members and better understand the molecular characteristics underlining the unique mechanical properties of tubuliform silk fibroins, three cDNA libraries were constructed and positive clones were isolated from the tubuliform gland of *Argiope aurantia* (*Ag. aurantia*), *Araneus gemmoides* (*Ar. gemmoides*), and *Nephila clavipes* (*N. clavipes*). The predicted proteins were compared to other silk proteins and to the known composition of the fibers.

MATERIALS AND METHODS

Spider Collection and Dissection. Female adults of *N. clavipes* and *Ag. aurantia* were purchased from Hatari Invertebrates (Portal, AZ). *Ar. gemmoides* were collected locally. The spiders were maintained at 24 °C for 12 h in light/12 h in dark in individual chambers and fed with frozen crickets weekly. Tubuliform glands were isolated, flash frozen, and stored in −70 °C freezers.

mRNA Extraction. Tubuliform glands were homogenized in TRI Reagent (Molecular Research Center, Inc.) using a Polytron homogenizer (IKA Labortechnik, Ultra Turrax). Total RNA was extracted following the manufacturer's instructions. mRNA was extracted from total RNA using Dynabeads oligo$(dT)_{25}$ (Dynal Inc., Lake Success, NY). The concentration of the isolated mRNA was determined using a spectrophotometer (Beckman Instruments, Inc., Fullerton, CA).

cDNA Library Construction. cDNA was synthesized from isolated mRNA using the SuperScript Choice System kit (Invitrogen Corp., Carlsbad, CA). To reduce mRNA secondary structure, 0.5% Tween-20 and 8% DMSO had to be used in the synthesis of the cDNA second strand. The synthesized blunt-ended cDNAs were ligated to an *Eco*RI adaptor and size-fractionated by a Chroma Spin+TE-1000 gel filtration column (Clon Tech Laboratories Inc., Palo Alto, CA). Large fragments (>500 bp) were ligated into the Lambda ZAP II vector (Stratagene, La Jolla, CA). The ligation was then used to package and transfect *Escherichia coli* XL1-blue MRF′ cells. After tittering of the package reaction, the library was amplified and stored in −70 °C. Mass excision was used to excise the phagemid with helper phage. The excised phagemids were plated to media containing IPTG and X-gal for blue-white selection. Blue colonies were selected randomly for plasmid preparation and glycerol stocks.

Several combinations of restriction enzymes were used to check for the presence of inserts and their sizes. Clones with inserts were sent for sequencing in our USDA sequencing center (Laramie, WY). By blast search for homologous sequences, and determination of repetitive sequence/carboxy-terminal sequence, potential positive clones were identified. Amino acid composition was also calculated from sequence information and compared with published data for tubuliform silk. The positive clones were then used as a probe by random labeling with [α-^{32}P]dCTP to hybridize to the phage library for more positive inserts. The EZ::TN ⟨TET-1⟩ transposon insertion kit was used to sequence larger inserts (Epicenter, Madison, WI). Two primers (5′-TAACGCATTAGCTAAGGC-3′, 5′-GGCAGCAGAATAAGACAACGC-3′) were also designed on the basis of the C-terminus sequence to sequence those that could not be finished from the vector primer, but were too short to use a transposon.

Northern Blot. mRNA was purified as described above. A 200 ng portion for *Ag. aurantia*, 380 ng for *Ar. Gemmoides*, and 300 ng for *N. clavipes* were electrophoresed on a denaturing formaldehyde agarose gel and blotted to Hybond-NX nylon membranes (Amersham Biosciences) for hybridization. For each species, corresponding major ampullate gland mRNA (200 ng) was also loaded as a negative control. The repetitive fragments obtained from the respective cDNA clones were randomly labeled with [α-^{32}P]dCTP as probes for each species. Hybridization was conducted at 45 °C overnight. Following a series of washes, the membranes were exposed to X-ray film (Fuji, Japan) and developed.

Protein Isolation for Amino Acid Analysis. Tubuliform glands were homogenized in SDS buffer (2% SDS, 200 mM Tris−Cl, 1 mM EDTA). Followed by extraction in SDS buffer one more time, the pellet was rinsed in H_2O and then acetone. The pellet was dried, rinsed with formic acid, and dried again.

Sequence Analysis. ClustalW was used to construct alignments using default settings (Mac Vector 7.2, Accelrys Inc., San Diego, CA). Parsimony analysis was done using PAUP* (Swofford, 2002) with gaps treated as both missing data and characters. Bootstrap with 1000 replicates was used with branch-and-bound search for the analysis.

RESULTS

Cloning and Sequencing of TuSp1. Numerous attempts to obtain high-quality cDNA libraries from tubuliform glands failed because of poor yields in the second strand cDNA synthesis. A wide variety of conditions were tested to develop the method described in the Materials and Methods. Due to the high divergence of previously published silk fibroin sequences in silk gene family members, we were unsuccessful in using probes based on those sequences to identify tubuliform silk cDNA clones (*2*, *8*). Instead, the libraries were screened by restriction enzyme digestion and direct sequencing of large-sized clones. Previous data from the carboxy terminus of numerous spider silk proteins indicated that the sequence of this region was conserved in all instances (*2*, *4*−*10*, *13*, *14*). Thus, we were able to identify likely tubuliform silk cDNA clones on the basis of homology to these sequences. Once likely clones were identified, the library was probed again with the sequenced tubuliform silk cDNA fragments. The sequenced cDNA clones in each library are only partial, and differ in length from each other.

All the sequenced cDNA clones are in two groups; one contains both the repetitive sequence and the nonrepetitive carboxy terminus, while the other has only the repetitive region. However, neither contains the amino terminus. The longest clone obtained from screening each of the three libraries is 1.7 kilobases (kb) for *Ag. aurantia*, 2.0 kb for *Ar. Gemmoides*, and 1.8 kb for *N. clavipes*, which were used in subsequent analysis (GenBank accession numbers AY855098, AY855099, AY855100, AY855101, and AY855102). The encoded tubuliform silk proteins are named TuSp1 (tubuliform spidroin 1). Blast searches of the translated amino acid sequence indicate that the repetitive sequences of tubuliform silk protein share no sequence

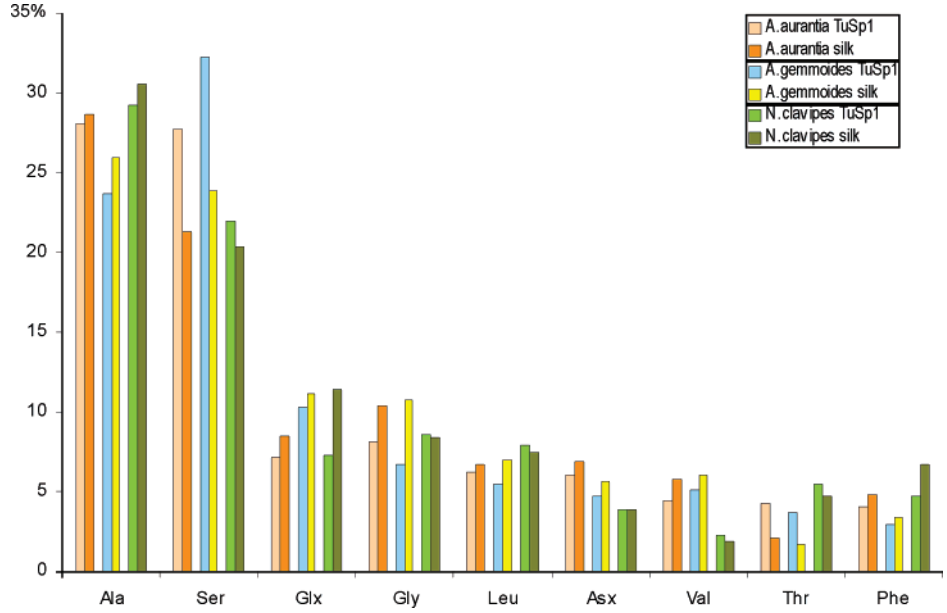

FIGURE 1: The amino acid compositions predicted from TuSp1amino acid sequences are compared with the data derived from *Ag. aurantia*, *Ar. gemmoides*, and *N. clavipes* tubuliform silk protein. Only the most abundant amino acids are shown (see Supporting Information Table 1 for complete composition comparison). Due to acid hydrolysis in the amino acid composition analysis, Gln and Glu (Glx) and Asn and Asp (Asx) cannot be distinguished. The predicted amino acid composition was calculated on the basis of the Tusp1 repetitive sequence in each species, and inclusion of the carboxy-terminal sequence in the calculation did not change the data significantly.

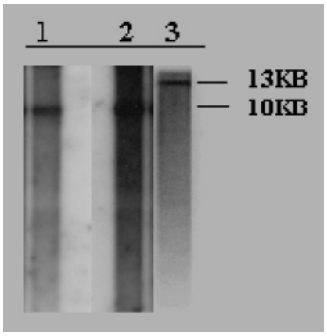

FIGURE 2: Northern blot analysis of the tubuliform silk gene transcript for *Ag. aurantia*, *Ar. gemmoides*, and *N. clavipes*. mRNA (200 ng for *Ag. aurantia* (lane 1), 380 ng for *Ar. gemmoides* (lane 2), and 300 ng for *N. clavipes* (lane 3)) was electrophoresed and blotted to membranes, and then hybridized with homologous probes.

similarity with known spider silk fibroins, although the carboxy-terminal region shows high sequence conservation with known silk gene family members. The predicted amino acid compositions from *Ag. aurantia*, *Ar. Gemmoides*, and *N. clavipes* TuSp1s were compared with compositions obtained by direct analysis of the respective silks (Figure 1).

Previously reported spider silk gene transcripts are very large and range in size from 4.4 to 15.5 kb (*4*−*10*, *14*). The lengthy transcripts of silk fibroin have been reported to severely limit the synthesis of full-length cDNAs (*4*, *5*, *7*, *14*). Northern blotting was used to estimate the transcript size of tubuliform silk gene for the three species (Figure 2). *Ag. aurantia* and *Ar. gemmoides* produce transcripts of about 10 kb, while *N. clavipes* produces a larger transcript of about 13 kb.

Sequence comparison within each species reveals the repetitive units of TuSp1 for each species. The length of the repeat units differs slightly among the three species (Figure 3). The *Ag. aurantia* TuSp1 repeat unit is 180 amino acids (aa), *Ar. gemmoides* is 182 aa, and *N. clavipes* is 171 aa. The amino acid motifs A_n, S_n, $(SA)_n$, $(SQ)_n$, and GX (X represents Q, N, I, L, A, V, Y, F, and D) are common to TuSp1 repetitive sequences in the three spider species. Unlike polyAla in previously described silk fibroin sequences, the TuSp1 repeat has very short alanine stretches, with A_3 the longest, scattered throughout the repetitive sequences. Interestingly, *Ag. aurantia* and *N. clavipes* TuSp1's share a stretch of threonine (T_5) which is not seen in any of the known araneoid silk fibroins, while *Ar. gemmoides* and *N. clavipes* TuSp1's contain a QQ motif in their repeat units. Despite the differences in the repeat units, all three proteins contain a 99 aa nonrepetitive carboxy terminus.

High Sequence Similarity of TuSp1. The individual repeat units of TuSp1 within each species were compared to each other (see Supporting Information Figures 4−6). Three repeat units of 180 aa each corresponding to 540 base pairs (bp) of DNA sequence were aligned for *Ag. aurantia* TuSp1 at both the DNA and amino acid levels. There are only 14 variation sites of the 540 bp unit, 5 of which cause amino acid differences and 9 of which are synonymous substitutions. This results in a 97% sequence identity (see Supporting Information Figure 4). A similar result is found when *N. clavipes* TuSp1 repeats are aligned. Eight nonsynonymous variations and six synonymous substitutions of the 513 bp (171 aa) repeat sequence lead to a 96% sequence identity in amino acid level and 97% similarity in DNA level (see Supporting Information Figure 5). Alignment of *Ar. gemmoides* TuSp1 repeats of 182 aa each (546 bp of DNA sequence) exhibits an even higher similarity. Of the seven nucleotide variations in the 546 bp repeat unit, only one causes an amino acid difference which results in a greater than 99% amino acid sequence identity and greater than 98% similarity in DNA level (see Supporting Information Figure 6).

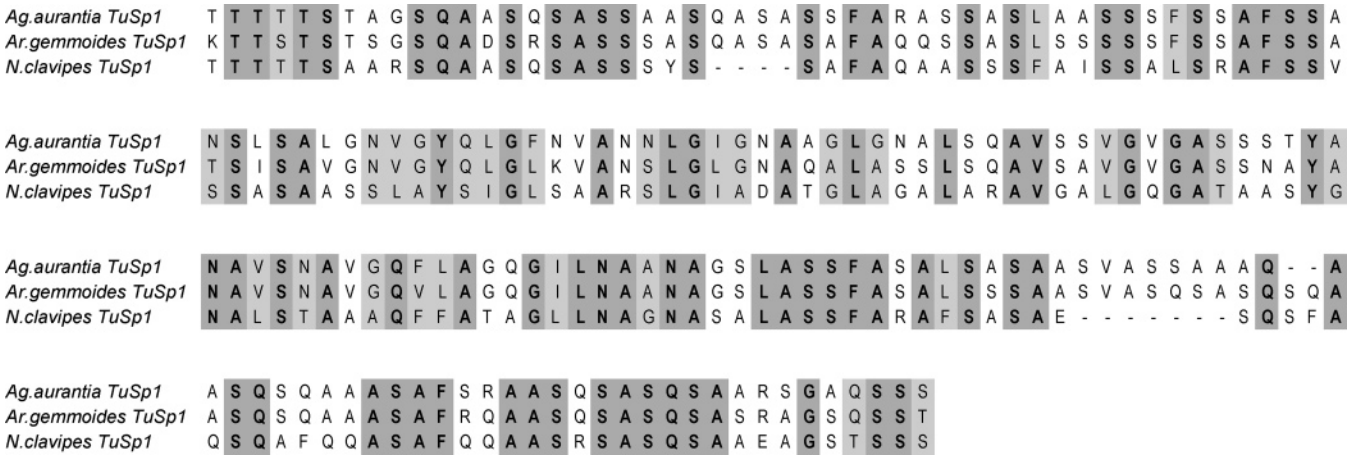

FIGURE 3: Amino acid sequence alignment of the TuSp1 repeat units among *Ag. aurantia*, *Ar. gemmoides*, and *N. clavipes*. Amino acids are indicated as single letters. The alignment was created using ClustalW default parameters. Shaded regions with bold letters indicate identical amino acids. Lightly shaded regions indicate similar amino acids. Genbank accession numbers: AY855098, AY855099, AY855100, AY855101, and AY855102.

In addition to high similarity of the TuSp1 repeat sequences within species, sequence comparison of the TuSp1 repetitive units among *Ag. aurantia*, *Ar. gemmoides*, and *N. clavipes* shows that the repeat units are also conserved between TuSp1 orthologues with a 46% sequence identity and an additional 12% sequence similarity (Figure 3). However, no significant sequence similarity is observed when comparisons are made with their paralogous genes from other glands. It has been reported in the silk fibroin gene family that sequences among particular orthologue groups can be highly conserved, while paralogous genes or cDNAs exhibit an extensive divergence (*2*, *8*, *9*). Pairwise alignments of the TuSp1 repetitive units between each pair of the TuSp1 orthologues reveal different levels of conservation. The alignment of the *Ag. aurantia* and *Ar. gemmoides* TuSp1 repeat units has 79% sequence similarity, while the *Ag. aurantia*/*N. clavipes* alignment has 53% sequence identity, and the *Ar. gemmoides*/*N. clavipes* pair has 52% similarity (see Supporting Information Figure 7).

Consistent with the conservation of TuSp1 repeat units among species, the 99 aa nonrepetitive carboxy-terminal region also shows nearly 50% identity with an additional 15% similarity in amino acid level among the three species (see Supporting Information Figure 8). This conservation has been observed in all of the silk fibroin proteins characterized thus far (*4*−*10*, *13*−*15*). The carboxy-terminal sequences representing the known silks and species were used for a parsimony analysis since the repetitive sequences differ to such an extent that they cannot be used in this analysis (see Supporting Information Figure 9). The result shows that TuSp1's group together, while MiSp1 clusters with MiSp2, and MaSp1 forms a clade with MaSp2.

Codon Usage Analysis. It is well-known that the DNA sequences encoding silk fibroin proteins have strongly biased codon usage in which adenine (A) and thymine (T) are highly preferred as the third base of three bases encoding one amino acid (*4*, *5*, *9*). The codon usages for the three most abundant amino acids, alanine, serine, and glycine, and the total base composition are analyzed for TuSp1 coding sequences and compared with those representing all the known silk fibroins in orb-weaving spiders (see Supporting Information Figure 10). Compared with the extremely biased codon usage (up to 94%) for alanine (GCN) in the sequences encoding major ampullate, minor ampullate, and flagelliform silk proteins, the alanine codons of TuSp1 coding sequences are only moderately biased toward A and T, with 66% the highest in the *Ar. gemmoides* TuSp1 coding sequence. Similar results are observed in the glycine and serine codons in which A and T are highly preferred in the wobble position in MaSp1, MaSp2, MiSp1, MiSp2, and Flag coding sequences, whereas they are selected only slightly higher than guanine (G) and cytosine (C) in tubuliform silk protein coding sequences. Even though A and T are strongly selected in the third position in DNA sequences encoding MaSp1, MaSp2, MiSp1, MiSp2, and Flag, they are only moderately preferred in tubuliform silk proteins, and even less used in the alanine codon of the AcSp1 encoding sequence. However, A/T and C/G contents are similar in all the silk protein coding sequences despite the divergences in their usage in the wobble position for encoding specific amino acids.

DISCUSSION

The tubuliform silk cDNA clones reported in this study were obtained from sequencing randomly selected clones after confirmation of the inserts with restriction enzyme digestion. Similar procedures have been reported in identifying silk gene transcripts from cDNA libraries due to the great sequence divergence among known araneoid silk fibroins (*2*, *8*). The sequenced tubuliform silk cDNAs for all three spiders are partial sequences including the repetitive region and nonrepetitive carboxy terminus, but lacking the amino-terminal region. It has been reported that the lengthy silk gene transcripts severely limit the full-length cDNA products (*7*). Northern blot analysis shows that the tubuliform silk gene transcript is also very large, ranging from 10 to 13 kb in the three species (Figure 2). Additionally, the use of harsh denaturing conditions for a successful second strand cDNA synthesis strongly suggests that secondary structure formation is also contributing to the short cDNA lengths found.

Variations in transcript size among TuSp1 orthologues (Figure 2) may result partially from replication slippage due to the internally repetitive structure of TuSp1 encoding sequences. Length polymorphism due to replication slippage and homologous recombination events has also been reported

in alleles of spider silk and silkworm silk (*Bombyx mori*, *B. mori*) fibroin genes (*10*, *15*−*17*).

Four amino acid motifs, A_n, $(GA)_n$, $(Gly\text{-}Gly\text{-}X)_n$, and $Gly\text{-}Pro\text{-}Gly\text{-}X_n$, in different combinations and arrangements form the majority of previously reported spider silk proteins (*2*, *4*−*7*, *14*). It has been proposed that these conserved motifs directly correspond to the extraordinary mechanical properties in silk fibers (*3*, *9*). For example, polyAla forms a β-sheet crystalline structure that is thought to be responsible for the high tensile strength of dragline silk, while the $Gly\text{-}Pro\text{-}Gly\text{-}X_n$ motif may form a β-spiral structure providing the silk with elasticity (*3*). However, these amino acid motifs are rarely represented in tubuliform silk proteins. Instead, TuSp1 contains several new motifs such as S_n, $(SA)_n$, $(SQ)_n$, and GX (X represents Q, N, I, L, A, V, Y, F, and D). Some of these motifs are also found in silk proteins of basal taxa spiders such as the mygalomorph *Euagrus chisoseus* (*E. chisoseus*) and *Plectreurys tristis* (*P. tristis*) (*2*). In addition, A_n, $(GA)_n$, $(Gly\text{-}Gly\text{-}X)_n$, and $Gly\text{-}Pro\text{-}Gly\text{-}X_n$ are poorly present in aciniform silk protein AcSp1 (*8*). Since tubuliform silk (eggcase silk) and aciniform silk (wrapping silk) do not serve as the structural silk fibers as do major or minor ampullate or flagelliform silk, and primitive spiders do not even construct webs, the presence of these motifs may not be crucial for silks not involved in prey capture. This result supports the argument of the correlation among protein primary sequence, secondary structure, and mechanical properties of orb-web silks (*3*).

The amino acid composition derived from the TuSp1 sequence is compared with the data obtained from analyzing the corresponding tubuliform silk protein for all three spiders (Figure 1). Generally, they match well with each other, which suggests that TuSp1 is the major protein component in the tubuliform gland. However, a slightly higher serine level is observed in the predicted amino acid composition than from the silk of *Ag. aurantia* and *Ar. gemmoides* (Figure 1). This difference may be due to the partial nature of the TuSp1 sequence since it is known that the nonrepetitive amino-terminal sequence is different in amino acid composition from the repetitive sequence (*7*). Additionally, other proteins may exist in the gland. Last, the acid digestion conditions needed for complete hydrolysis of tubuliform silk protein may partially destroy serine, which could lead to the decreased concentration of this amino acid.

In addition to the high serine but low glycine content, tubuliform silk protein also contains more amino acids with large side chains than other silk fibroins, such as valine, leucine, isoleucine, and phenyalanine (Figure 1). The relatively high levels of these amino acids are consistent with the results of fiber X-ray diffraction (*18*). The data show that eggcase silk has a larger *b* dimensional value in the β-sheet than major and minor ampullate silk, indicating the presence of large-side-chain amino acids (*18*). Secondary structure predictions of the TuSp1 sequence also indicate large stretches of β-sheet are likely to occur with this protein as the fiber X-ray diffraction shows. β-Sheet has been proposed as the key structure for the formation of fibers and for the strength characteristics of spider silk (*18*). Furthermore, the presence of large-side-chain amino acids in tubuliform silk proteins also agrees with the results of tubuliform silk transmission electron microscopy (TEM) (*19*). The TEM diffraction pattern shows the presence of streaks in tubuliform silk that is an indication of variations in β-sheet spacing (*19*). The complex molecular architecture of tubuliform silk protein with the presence of large-side-chain amino acids could better explain the highly frustrated crystalline secondary structure shown in TEM that might be responsible for the lower stiffness of tubuliform silk.

Before the identification of TuSp1, ADF-2 was reported as a protein synthesized in the *Araneus diadematus* (*Ar. diadematus*) tubuliform gland (*13*). However, due to the mismatch in amino acid composition between the data derived from the ADF-2 sequence and those from the silk, the authors suggested that there might be more than one protein in the tubuliform gland (*13*). It has been shown that two proteins can be present in one silk such as MaSp1 and MaSp2 in dragline silk (*4*, *5*), and MiSp1 and MiSp2 in minor ampullate silk (*6*). Common features of the sequences of MaSp1, MaSp2 and MiSp1, MiSp2 suggest that the combined proteins forming major and minor ampullate silks must have functional applications in terms of strength and elasticity. It is not clear how the combination of ADF-2 and TuSp1 would produce any functional advantages for tubuliform silk. In this study, TuSp1 is the only protein sequence found after three tubuliform gland specific cDNA libraries for *Ag. aurantia*, *Ar. gemmoides*, and *N. clavipes* have been screened, and hundreds of clones in all three libraries have been sequenced. DNA probes used to screen the library included ones that would have identified sequences from ADF-2. In addition, the amino acid composition derived from TuSp1 closely matches the data obtained from the silk protein. Therefore, we propose that TuSp1 is the major component in the tubuliform gland, although it is possible that ADF-2 mRNA can be expressed at a low level.

Sequence comparison shows that TuSp1 repetitive sequences are more homogeneous within species than among species (Figure 3; see Supporting Information Figures 4−6). The only divergence among repeat units within species is due to single-base substitutions (see Supporting Information Figures 4−6). The high homogeneity among TuSp1 repeat units within species in all three spiders is an indication of within-gene concerted evolution, probably through gene conversion and unequal crossover events. Similar results have been reported in other spider silk gene family members and silkworm silk (*8*, *10*, *15*, *17*).

Although comparative analysis demonstrates the sequence similarity of the TuSp1 repeat unit among the three spiders (Figure 3), pairwise alignments between two spiders suggest different levels of conservation (see Supporting Information Figure 7). The result shows that *Argiope* and *Araneus* are more similar to each other than either is to *Nephila* in terms of repeat unit length and sequence similarity. This result is not surprising given the differences in these three species. Morphological evidence shows that *Nephila* belongs to the derived araneoids, while *Argiope* and *Araneus* belong to Araneidea that diverged from derived araneoids about 125 million years ago (*20*).

In comparison with the high sequence homogeneity within species, the repetitive sequences show greater divergence between species (Figure 3). In addition to point mutations, there are insertions/deletions of multiple codons in the repetitive sequences among the three species. The divergence, in part, may be due to the different selection pressures on tubuliform silk among the three spider families given that

distinct motifs are observed in the TuSp1 repeat unit in some species, but not in others.

The carboxy-terminal sequences representing the known silks and species were used for a parsimony analysis (see Supporting Information Figure 9). The result shows that TuSp1 groups together, while MiSp1 clusters with MiSp2, and MaSp1 forms a clade with MaSp2, which suggests that TuSp1 represents a new orthologue group of the spider silk gene family. Consistent with previous results that *Ar. gemmoides* and *Ag. aurantia* are more closely related than either is to *N. clavipes*, parsimony analysis of the carboxy-terminal sequences indicates that *Ar. gemmoides* and *Ag. aurantia* group tightly and then form a clade with *N. clavipes*. The parsimony analysis also suggests that tubuliform silk may have evolved at least at the same time as major and minor ampullate silk, which is inconsistent with systematic studies that indicate that major and minor silk evolved before tubuliform silk (*21*−*23*).

Although not as strong as seen in major and minor ampullate and flagelliform silk proteins, biased codon usage is observed in the most abundant amino acids in tubuliform silk proteins (see Supporting Information Figure 10). The biased codon usage may be due to mRNA secondary structure, the repetitive nature of silk gene transcripts, and the gland-specific tRNA pool. It has been reported in silkworm *B. mori* that the biased codon usage was determined by mRNA or chromatin structure, but not by tRNA population (*24*). They found that the codon usage for the three most abundant amino acids alanine, glycine, and serine was distinctly different in the repetitive and joint regions which alternate through the silk fibroin sequence (*24*). However, other studies have shown that, prior to protein synthesis, the corresponding silk gland accumulates a large amount of tRNAs specific for the abundant amino acids. Interestingly, like the silkworm *B. mori* silk gland, the major ampullate gland of *N. clavipes* develops two isoaccepting tRNA forms, one of which is specific for the gland (*25*). Unlike *B. mori*, which produces relative proportions of tRNAs in the silk gland according to the abundance of corresponding amino acids (*26*), the *N. clavipes* major ampullate gland accumulates more alanine tRNA although glycine is the most abundant amino acid in the gland (*25*). The difference may be due to the different distributions of alanine in these silks. Alanine is dispersed alternately with glycine or serine in silkworm silk, while it is in stretches of six to eight consecutive residues in MaSp1 and MaSp2, which may need a large pool of alanine tRNA to optimize the continuous demand (*5*).

In summary, characterization of tubuliform silk protein TuSp1 has resulted in several novel findings. First, sequencing and amino acid composition analysis suggest that TuSp1 is the major protein in the tubuliform gland (Figure 1). Additionally, the molecular characteristics of tubuliform silk protein could contribute to the understanding of its unique mechanical properties. Furthermore, sequence analyses reveal the presence of novel amino acid motifs as well as sequence conservation within and among species (Figure 3; see Supporting Information Figures 4−6). Finally, comparative analysis of the carboxy termini indicates that TuSp1 represents a new orthologous group of the silk fibroin gene family (see Supporting Information Figure 9). Further characterization of the tubuliform silk gene sequence will help to clarify the exact processes that control the evolution of this gene. Future structural study of tubuliform silk protein will provide evidence correlating protein sequence, structure, and mechanical properties.

ACKNOWLEDGMENT

We thank Dr. P. Langer, Dr. D. Jarvis, and Dr. M. Hinman for improving the paper, Dr. M. Hinman for the amino acid analysis, Dr. G. K. Brown and Dr. C. Y. Hayashi for help in comparative analysis, and two anonymous reviewers for valuable comments on the paper.

SUPPORTING INFORMATION AVAILABLE

Most of the comparative analyses and codon usage comparisons. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Shear, W. A., Palmer, J. M., Coddington, J. A., and Bonamo, P. M. (1989) A Devonian spinneret: early evidence of spiders and silk use, *Science 246*, 479−481.
2. Gatesy, J., Hayashi, C. Y., Motriuk, D., Woods, J., and Lewis, R. (2001) Extreme diversity, conservation, and convergence of spider silk fibroin sequences, *Science 291*, 2603−2626.
3. Hayashi, C. Y., Shipley, N. H., and Lewis, R. V. (1999) Hypothesis that correlate the sequence, structure and mechanical properties of spider silk proteins, *Int. J. Biol. Macromol. 24*, 271−275.
4. Xu, M. and Lewis, R. (1990) Structure for a protein superfiber: spider dragline silk, *Proc. Natl. Acad. Sci. U.S.A. 87*, 7120−7124.
5. Hinman, M., and Lewis, R. (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.
6. Colgin, M. A., and Lewis, R. (1998) Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk like "spacer regions", *Protein Sci. 7*, 667−672.
7. Hayashi, C., and Lewis, R. (1998) Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks, *J. Mol. Biol. 275*, 773−784.
8. Hayashi, C., Blackledge, T. A., and Lewis, R. (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.
9. Hayashi, C. Y. (2002) Evolution of spider silk proteins: insight from phylogenetic analyses, in *Molecular Systematics and Evolution: Theory and Practice* (DeSalle, R., Giribet, G., Wheeler, W., Eds.) pp 209−223, Brikhauser, Cambridge, MA.
10. Hayashi, C., and Lewis, R. (2000) Molecular architecture and the evolution of a modular spider silk protein gene, *Science 287*, 1477−1479.
11. Moon, M. J. (2003) Fine structural analysis of the cocoon silk production in the garden spider, *Argiope aurantia*, *Korean J. Biol. Sci. 7*, 35−41.
12. Stauffer, S. L., Coguill, S. L., and Lewis, R. V. (1994) Comparison of physical properties of three silks from *Nephila clavipes* and *Araneus gemmoides*, *J. Arach. 22*, 5−11.
13. Guerette, P. A., Ginzinger, D. G., Weber, B. H. F., and Gosline, J. M. (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family, *Science 272*, 112−115.
14. Tian, M.-Z., Liu, C.-Z., and Lewis, R. (2004) Analysis of major ampullate silk cDNAs from two non-orb-weaving spiders, *Biomacromolecules 5*, 657−660.
15. Beckwitt, R., Arcidiacono, S., and Stote, R. (1998) Evolution of repetitive proteins: spider silk from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae), *Insect Biochem. Mol. Biol. 28*, 121−130.
16. Manning, R. F., and Gage, L. P. (1980) Internal structure of the silk fibroin gene of *Bombyx mori*, *J. Biol. Chem. 255*, 9451−9457.
17. Mita, K., Ichimura, S., and James, T. C. (1994) Highly repetitive structure and its organization of the silk fibroin gene, *J. Mol. Evol. 38*, 583−592.
18. Parkhe, A. D., Seeley, S. K., Gardner, K., Thompson, L., and Lewis, R. (1997) Structural studies of spider silk proteins in the fiber, *J. Mol. Recognit. 10*, 1−6.

19. Barghout, J. Y. J., Thiel, B. L., and Viney, C. (1999) Spider (*Araneus diadematus*) cocoon silk: a case of non-periodic lattice crystals with a twist? *Int. J. Biol. Macromol. 24*, 211−217.
20. Selden, P., and Gall, J. (1990) Lower cretaceous spiders from sierra de montsech, northeast Spain, *Palaeontology 33*, 257−285.
21. Glatz, L. (1972) Der spinnapparat haplogyner spinnen (Arachida, Araneae), *Z. Morphol. Tiere 72*, 1−25.
22. Coddington, J., and Levi, H. (1991) Systematics and evolution of spiders (Araneae), *Annu. Rev. Ecol. Syst. 22*, 565−592.
23. Platnick, N., Coddington, J., Forster, R., and Griswold, C. (1991) Spinneret morphology and the phylogeny of haplogyne spiders, *Am. Mus. Novit. 3016*, 1−73.
24. Mita, K., Ichimura, S., and Zama, M. (1988) Specific codon usage pattern and its implication on the secondary structure of silk fibroin mRNA, *J. Mol. Biol. 203*, 917−925.
25. Candelas, G. C., Arroyo, G., Carrasco, C., and Dompenciel, R. (1990) Spider silk glands contain a tissue-specific alanine tRNA that Accumulates *in vitro* in response to the stimulus for silk protein synthesis, *Dev. Biol. 140*, 215−220.
26. Sprague, K. U. (1975) *Bombyx mori* silk proteins. Characterization of large polypeptides, *Biochemistry 14*, 925−931.

BI050366U