Protein Science (1998), 7:667–672. Cambridge University Press. Printed in the USA.

Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like "spacer regions"

MARK A. COLGIN[1] AND RANDOLPH V. LEWIS

Department of Molecular Biology, University of Wyoming, P.O. Box 3944, Laramie, Wyoming 82071-3944

(RECEIVED June 4, 1997; ACCEPTED November 12, 1997)

Abstract

Spider minor ampullate silk is a strong non-elastic deformably stretchable silk used in web formation. This silk from *Nephila clavipes* is composed of two proteins, MiSp 1 and 2, whose transcripts are 9.5 and 7.5 kb, respectively, as determined by northern blots. Both MiSp proteins are organized into a predominantly repetitive region and a small nonrepetitive carboxy terminal region. These highly repetitive regions are composed mainly of glycine and alanine, but also contain tyrosine, glutamine, and arginine. The sequences are mainly GGX and GA repeats. The repetitive regions are interrupted by nonrepetitive serine-rich spacer regions. Although the sequences of the spacer regions differ from the repetitive regions, sequences of the spacers from different regions of the proteins are nearly identical. The sequence differences between major and minor ampullate silks may explain the differing mechanical properties of the fibers.

Keywords: minor ampullate; protein; sequence; spider silk

An orb web, the typical spider web, is constructed of several different silk types, each composed primarily of protein and synthesized by a different gland (Lucas, 1986; Vollrath, 1992). These silks vary in their mechanical properties over a very wide range of tensile strength and elasticity (Vollrath, 1992). The spider silks provide the opportunity for the development of a whole new class of biomaterials. The best studied spider silk is dragline (major ampullate) silk from *Nephila clavipes* (Lewis, 1992). Minor ampullate silk is a related, but less well characterized fiber. Both major and minor ampullate silks have high tensile strengths with minor ampullate silk irreversibly deforming when stretched and major ampullate silk being elastic (Stauffer et al., 1994; Gosline et al., 1986). The partial primary structures of the two major ampullate silk proteins, MaSp 1 and MaSp 2, are known (Xu & Lewis, 1990; Hinman & Lewis, 1992). Structural analyses indicate regions of crystallinity associated with the β-sheet structures responsible for the extraordinary tensile strength of the fiber (Warwicker, 1960; Dong et al., 1991; Gillespie et al., 1994; Simmons et al., 1996; Parkhe et al., 1997). No biophysical studies have yet determined the structure of MaSp 2 but its predicted structure is that of an elastin-like protein, which could exist in a stacked β-turn or coil structure, which may be essential for fiber elasticity.

To develop a model explaining the strength and inelasticity of minor ampullate silk, we investigated the proteins in these fibers. The results of those studies are described in this paper. We present here the first description of substantial minor ampullate silk protein and DNA sequences. Minor ampullate silk fibers contain two proteins, which are predominantly composed of glycine and alanine. However, unlike the known spider silk proteins, minor ampullate silk proteins are not simply homogeneously repetitive. Instead, highly conserved nonrepetitive, serine-rich spacer regions alternate with the highly repetitive regions.

Materials and methods

RNA isolation

Adult female *Nephila clavipes* spiders were silked to stimulate RNA synthesis and glands were harvested within 24 h. Total RNA was purified from the minor ampullate glands of *Nephila clavipes* using a mild SDS RNA extraction procedure (Sambrook et al., 1989). Minor ampullate glands were identified on the basis of their morphology (j-shaped glands) and size (about half the size of the major ampullate glands). Poly(A) + RNA was purified using oligo(dT) Sepharose (Sambrook et al., 1989).

cDNA library construction

cDNA was made using a variation of the Riboclone (Promega; Madison, WI) cDNA synthesis system. The primers used for first

Reprint requests to: Randolph V. Lewis, Department of Molecular Biology, University of Wyoming, P.O. Box 3944, Laramie, Wyoming 82071-3944; e-mail: silk@uwyo.edu.

[1]Current address: Department of Biochemistry and Molecular Biology, Colorado State University, Fort Collins, Colorado 80523.

strand synthesis were *Not* I primer/adapter [AATTCGCGGC CGC(T)$_{15}$] (1.2 μg, Promega) and MIKE1 [A/T,G/C,G/C, A/T, G/C,G/C] (40 ng, University of Kentucky). The primers were incubated with 20 μg total RNA for 5 min at 70 °C and gradually cooled to room temperature. cDNA was synthesized using the cDNA Synthesis System from Promega (Madison, WI). A final concentration of 0.1 mM nicotinamide adenine dinucleotide and 2 units *Escherichia coli* DNA ligase were added to the second strand reaction mixture. cDNA was fractionated by size using a Sephacryl S-1000 gel filtration column (Sambrook et al., 1989). cDNA was ligated into the phosphatased *Sma* I site of pGEM 3Zf(-) (Promega; Madison, WI). Plasmids were transformed into *E. coli* SURE competent host cells (Stratagene; La Jolla, CA).

cDNA library screening

The colonies were lifted, cells lysed, and the DNA was fixed and denatured on nylon membranes. Cell debris was removed by washing the membranes and scraping with a spatula. Hybridization was performed according to manufacturer's protocol (Quikhyb; Stratagene, La Jolla, CA). The cDNA library was screened by colony hybridization with the ^{32}P end-labeled oligonucleotide probe (GC NCCNGCNCCNGCNCC, where N = A, G, C, or T). This probe was designed based on minor ampullate silk peptide sequence data which showed substantial stretches of GA repeats.

DNA sequencing

Deletion fragments and *Pst* I and *Pvu* II fragments were generated for DNA sequencing. Deletion subclones were generated using the Erase-a-Base Method according to the manufacturers procedures (Promega, Madison, WI). One to 2 micrograms of plasmid DNA were denatured at 85°C for 5 min in 0.2 N NaOH-20 mM EDTA and the DNA sequenced according to the Sequenase protocol for DNA sequencing with 7-deaza-dGTP (United States Biochemical; Cleveland, OH).

Polymerase chain reactions

Probes for Southern and Northern blotting were generated using a PCR-based labeling procedure (Mertz & Rashtchian, 1994). PCR reactions were 25 cycles and catalyzed by Deep Vent DNA polymerase (New England Biolabs, Beverly, MA). Reactions were carried out in a volume of 20 μL and conditions were as follows: 10 mM KCl, 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 0.825 μM ^{32}P-a-dCTP, 1.65 μM dATP-dTTP-dGTP, 10% dimethylsulfoxide, 200 pmol primers, template DNA (amount varies), and 2 units of Deep Vent DNA polymerase. Denaturation and extension temperatures were 98 °C and 72 °C, respectively. Annealing temperatures were set at 10 °C below the primer T_ms.

Spacer region probes were generated using the following primers: SPAC5′ (GGATCCTCTGCAGGAAATGCT) and SPAC3′ (GGATCCAGATGATGCAGAAGG). MiSp 1 cDNA (100 ng) was used as a template. The resulting PCR product is the 400 bp spacer region composed of non-repetitive sequence. The MiSp 1-specific probe was generated using the following primers: MIP1+ (ACAA CTTTGQCGTCTTATGGACC) and MIP1− (GGTCCATAAGA CGTCAAAGTTGT). MiSp 1 cDNA (100 ng) was used as a template. The resulting PCR product is a 1400 bp region composed entirely of MiSp 1 repetitive sequence. The MiSp 2 specific probe was generated using the following primers: MIP1+ (ACAACTTT GQCGTCTTATGGACC) and MIP3− (TTTAGCCCATTACAA CTTGCATG). MiSp 2 cDNA (100ng) was used as template DNA. The product is a 600 bp fragment which encodes the MiSp 2 C-terminal nonrepetitive region as well as MiSp 2 repetitive region.

Northern blotting

Minor ampullate gland total RNA (0.5–3 μg) was dissolved in Premix Solution (1 X MOPS, 6.5% formaldehyde, 50% formamide). RNA was incubated at 65 °C for 15 min and placed immediately at 0 °C. Formaldehyde Loading Buffer (1 mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol, 50% glycerol) was added before loading onto an agarase gel. Agarose gels (percentages varied) contained 0.66 M formaldehyde and 1 X MOPS. The running buffer was 1 X MOPS. After electrophoresis, the RNA was transferred overnight to Zeta Probe nylon membrane (Bio-Rad Laboratories, Richmond, CA) by capillary transfer using 10 X SSC. The membrane was rinsed with 6 X SCC and baked at 80 °C for 30 min. The RNA was stained by methylene blue (0.02% methylene blue, 0.3 M NaOAc [pH 5.5]) to visualize and mark RNA size markers. The membrane was destained in water. Prehybridization and hybridization were carried out at 65 °C, as described in the manufacturer's protocol for Quikhyb (Stratagene; La Jolla, CA). The membranes were washed twice for 15 min at room temperature with 2XSSC-0.1% (w/v) SDS, once for 30 min at 60°C with a 0.1XSSC-0.1% (w/v) SDS, then blotted dry, and subjected to autoradiography. MiSp 1 and spacer-specific PCR probes (described in the previous section) were used in two independent northern blotting experiments.

Southern blotting

Ten adult female *Nephila clavipes* spiders were sacrificed for genomic DNA preparation. Exoskeletons were removed from the abdomens. Whole abdomens (minus the silk glands) were ground into a powder and genomic DNA was extracted as previously described (Herrmann & Frischauf, 1987). Southern blotting was performed as previously described (Sambrook et al., 1989). Genomic DNA (50 μg/reaction) was digested with various restriction enzymes and precipitated with ethanol. Samples were loaded onto a 0.7% agarose gel and electrophoresed overnight. The gel was stained with ethidum bromide and photographed. The DNA was denatured by soaking the gel in 1.5 M NaCl–0.5 N NaOH for 45 min. The gel was rinsed with dH_2O and neutralized by soaking in 1 M Tris-HCl (pH 7.4)–1.5 M NaCl for 30 min. Denatured DNA was transferred to a nylon filter (Hybond-N; Amersham, UK) using capillary transfer with 10 X SSC as the transfer buffer. DNA was transferred for 18 h and fixed to the membrane by baking at 80 °C for 2 h. Membranes were rinsed with dH_2O to remove salts. Hybridization was performed at 65 °C overnight using the QuikHyb method. Membranes were washed blotted dry, and subjected to autoradiography. Each of the three probes generated by PCR were used in separate Southern blotting experiments. This allowed for identifying MiSp 1 and MiSp 2-specific genomic DNA fragments.

Results

Eight partial cDNAs from minor ampullate silk gland mRNA were completely or partially sequenced representing over 8 kilobases.

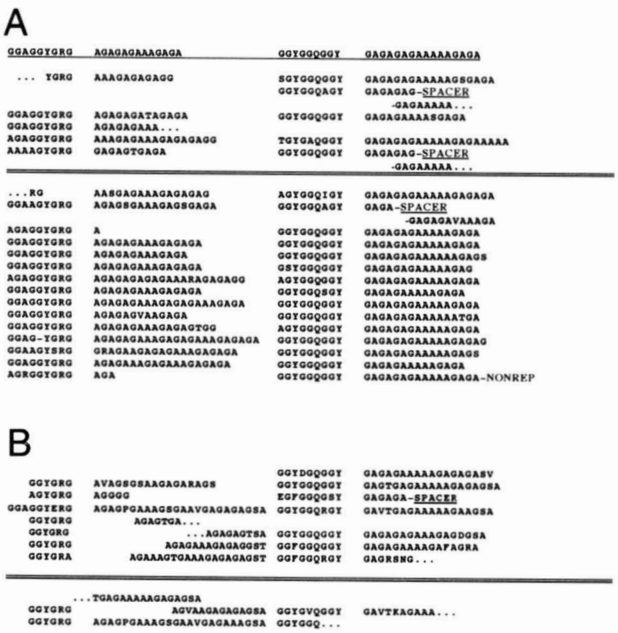

Fig. 1. MiSp spacer and repetitive regions. Amino acid sequences for the repetitive regions of MiSp 1 (**A**) and MiSp 2 (**B**) were deduced from cDNA sequences. ORFs were established based on peptide sequence data. Underlined repetitive sequences are arranged to illustrate conserved residues. Breaks in sequence information are shown with "...", otherwise the sequences are continuous. "SPACER" and "NONREP" represent spacer regions and carboxy-terminal nonrepetitive regions, respectively. The double-underlines separate different cDNAs.

These cDNAs were initiated from both the GC-rich hexanucleotide and oligo(dT) primers. The DNA sequences from multiple clones indicated two different proteins, minor ampullate silk proteins 1 and 2 (MiSp 1 and 2) were present. When translated, the proteins contained some repetitive sequences which were similar to the few known spider silk proteins. Figure 1 shows representative repetitive sequences from both proteins. Northern blotting for the MiSp 1 and MiSp 2 mRNAs, using multiple specific probes, indicated two transcripts of about 9.5 kb and 7.5 kb, respectively (Fig. 2) with the spacer probe hybridizing to both transcripts (Fig. 2B). No cross hybridization was observed between major and minor ampullate mRNA with the probes used. Based on the these data, we are confident the majority of MiSp 1 has been sequenced from the multiple clones.

Southern blot analysis of genomic DNA confirmed the presence of two MiSp genes and indicate that there are only two MiSp genes that have these sequences. The spacer, repetitive, and consensus region probes only hybridize to genomic *Pvu* II or *Pst* I restriction fragments 600 bp or smaller. *Pvu* II and *Pst* I cleave at least 30 regularly, nearly evenly spaced sites within a single 1200 bp repetitive region but none within the spacer region. As a result, the 600 bp *Pvu* II or *Pst* I genomic fragments correspond to the uncleaved spacer regions. A *Bcl* I site is present in each spacer region, which is consistent with the several small (<3 kb) genomic *Bcl* I fragments on the Southern blots. None of the other six base recognition restriction enzymes used in Southern blotting experiments could produce hybridizing fragments smaller than 9 kb and 7 kb for MiSp 1 and 2, indicating there are no sequences intervening within the majority of the protein coding region. More precise

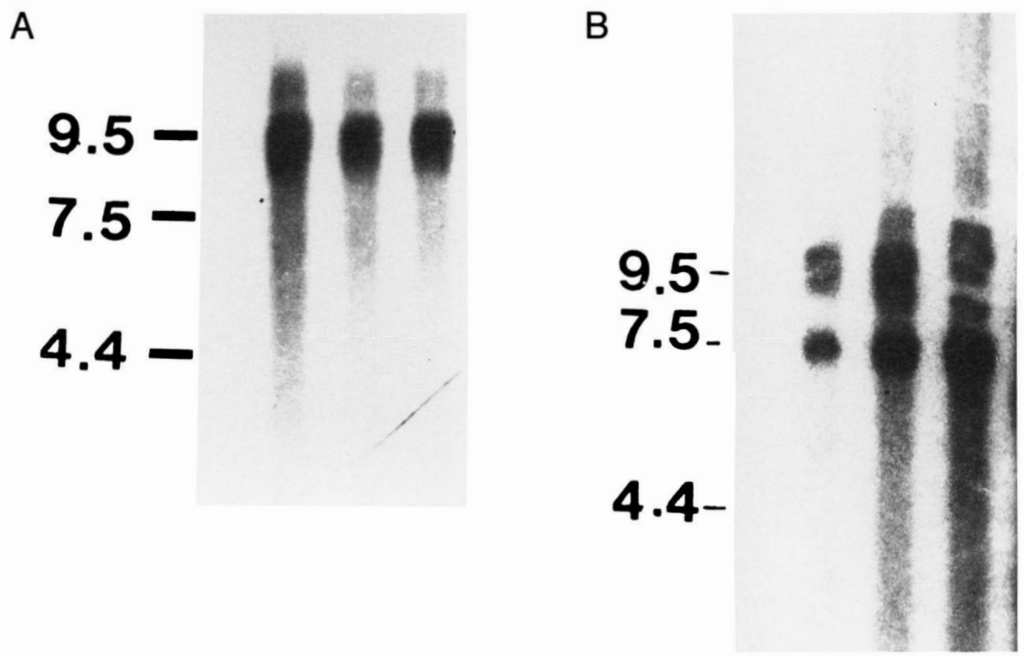

Fig. 2. A: Northern blot for MiSp 1. Total minor ampullate RNA (Lane 1: 0.5 μg, Lane 2: 0.1 μg) was electrophoresed on a 0.6% agarose gel and transferred to nylon membrane. MiSp 1 repetitive region probe was used for hybridization. The size markers are shown in kilobases. **B:** Northern blot for MiSp 1 and MiSp 2. Total minor ampullate RNA (Lane 1: 0.1 μg, Lane 2: 0.5 μg) was electrophoresed on a 1% agarose gel and transferred to nylon membrane. The MiSp spacer region probe was used for which hybridizes to both MiSp 1 and MiSp 2.

restriction mapping could not be performed because restriction enzymes either have no sites or many sites in the coding regions.

MiSp 1 and 2 repetitive regions

A typical repetitive region has ten repeat units. There are two types of motifs within a single repeat unit of MiSp 1, GGXGGY (X = Q or A) motifs alternate with $(GA)_y(A)_z$ motifs, where $y = 3–6$ and $z = 2–5$ (Fig. 1A). The first GGXGGY motif in the repeat unit is always followed by GRG and X is always alanine. The second GGXGGY motif is followed directly by the $(GA)_y(A)_z$ and X is always glutamine. Tyrosines are completely conserved in MiSp 1, but phenylalanine substitutions occur in MiSp 2 in only 3 of 29 cases (Fig. 1B). Although there are substitutions for other residues, there is typically no more than one substitution per repeat. When comparing the repeat units within either MiSp, the repeats are more highly conserved than those of either of the major ampullate silk proteins (Hinman et al., 1994). Repeat units of MiSp 2 are similar to MiSp 1, but are not as well conserved.

Like MiSp 1, the repeat unit of MiSp 2 has $(GGX)_n$ [X = Y, Q, or A; $n = 1–3$] domains that alternate with GAGA motifs. The first $(GGX)_n$ motif is shorter than the same motif in MiSp 1 since MiSp 2 lacks the GGA present before the GGYGRG sequence. Four residues (Y, Q, R, and S) in the MiSp 2 repeat are highly conserved. One of the conserved residues which distinguishes MiSp 1 from MiSp 2 is the serine (underlined) in the GAGAGAAAAA GAGSA domain of MiSp 2.

Nonrepetitive spacer regions

The repetitive regions of both MiSp proteins are interrupted by serine-rich spacer regions (Fig. 3). The 137 amino acid spacer regions are virtually identical in the two MiSps and show no identity to any other known spider silk sequence. MiSp 1 spacers (137 amino acids) are separated by 10 repeat units with an average of about 140 residues. Serine (24%) and alanine (18%) are the most abundant residues in this region. Leucine (8%), asparagine (7%), threonine (7%), and valine (7%) are also present in higher percentages than in the repetitive regions. The glycine content is lower in the spacer (10%) than in the repetitive regions (17%). The spacer region is not internally repetitive, with the exception of the sequence, SSAA, which is present as a single tandem repeat.

The spacers are similar in serine composition to the *Bombyx mori* amorphous regions, although the silkworm fibroin amorphous regions are only about 30 amino acids (Mita et al., 1994). Spacer regions are also similar in composition to the spider silk protein carboxy-terminal consensus region but apparently nonhomologous. Even with the >90% sequence identity among spacers from both MiSp proteins, there is no evidence to suggest the presence of spacers or spacer-like sequences in proteins from major ampullate silks (M.B. Hinman & R.V. Lewis, unpubl. data). Genomic restriction analysis and Southern blotting showed that the spacer probes only hybridized to the MiSp genes (data not shown). However, it is possible that silk proteins from other glands have spacers with sequences that differ sufficiently from the MiSp spacers to prevent hybridization.

MiSp Spacer Region

^{1}GGSSAGNAFAQSLSSNLLSSGDFVQMISSTTSTDHAVSVATSVAQN

VGSQLGLDANAMNNLLGAVSGYVSTLGNAISDASAYANALSSAIGN

VLANSGSISESTASSAASSAASSVTTTLTSYGPAVFYAPSASSGG^{137}

Fig. 3. MiSp spacer region consensus sequence. Amino acid sequences for the nonrepetitive spacer regions were deduced from cDNA sequences.

MiSps are highly organized

Both minor ampullate silk proteins have a four-level hierarchy of organization that begins with the repeating tracts of glycylalanine (GA) and alanylalanine (A) (Fig. 4A). The next level is the organization of motifs within the repeat unit: $(GGX)_n$ motifs alternate with $(GA)_y(A)_z$ motifs (Fig. 4B). A typical repetitive region is composed of ten repeat units, $[(GGX)_n(GA)_y(A)_z]10$ (Fig. 4C). The highest level of organization is the alternating array of repetitive and spacer regions (Fig. 4D). MiSp organization shows some similarity to that found in *B. mori* silkworm silk heavy chain

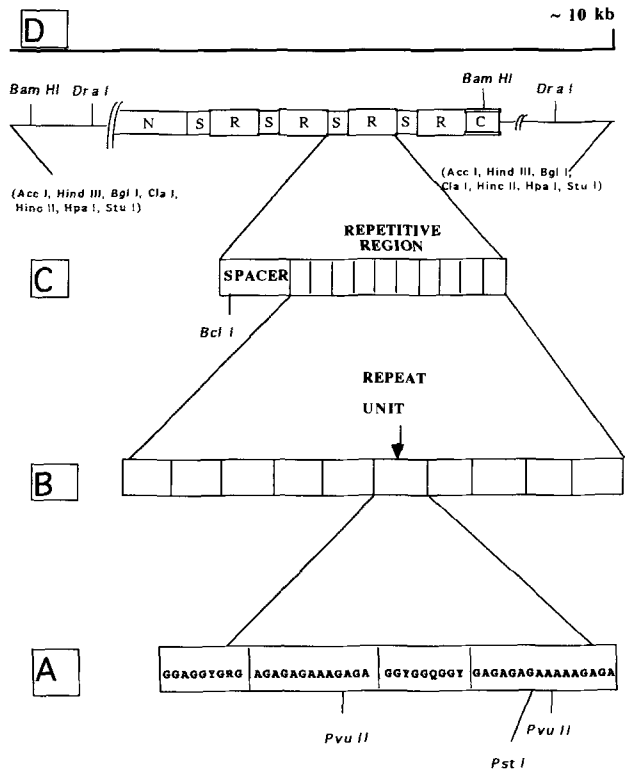

Fig. 4. Levels of organization of MiSp 1. The first level of repetitiveness is the tri- and dipeptide stretches, $(GGX)_{2-3}$ and $(GA)_{3-4}$, shown in boxes at the bottom. Two GGX and two GAGAG domains form the consensus repeat unit. About 10 units repeat to form the repetitive (R) regions. Spacers (S) alternate with R. N = nonrepetitive amino terminal region and C = carboxy-terminal consensus region. Restriction sites are shown to indicate the position of the corresponding nucleotide sequence and are based on sequence and genomic restriction mapping information. Sites in parenthesis exist outside the Bam HI and Dra I sites, but their relative locations have not been determined. MiSp 2 is organized in a similar manner; however, the consensus repeat has a slightly different content and length.

fibroin in which crystalline (repetitive) and amorphous (spacer) regions alternate throughout a 15 kb coding region (Tsuijimoto & Suzuki, 1979a, 1979b).

C-terminal spider silk consensus sequences

A 99 amino acid carboxy-terminal consensus region (Fig. 5) is present in the four known *Nephila clavipes* silk proteins. The MiSp 1 consensus sequence is more closely related to that of MiSp 2 (99% identical) than to that of either MaSp 1 (50% identical) or MaSp 2 (49% identical). Sequence conservation in this carboxy-terminal region was shown in silk proteins from two other orb-weaving species, *Araneus bicentenarius* and *Araneus diadematus* (Beckwitt & Arcidiacono, 1994; Guerette et al., 1996). There is no sequence similarity between the carboxy-terminal nonrepetitive regions of *B. mori* silk fibroin and spider silk (Mita et al., 1994). The function of the carboxy-terminal consensus region is unknown and it has not been determined whether the region is present in the fiber. Regardless of function, this high sequence conservation suggests an evolutionary significance.

No amino-terminal sequence information is available for any of the spider silk proteins because the transcripts are long (average 10 kb) and presumably have high secondary structure. The repetitive and spacer regions alternate throughout the majority of the two MiSps. Results of sequencing, PCR, and genomic restriction mapping indicate that the vast majority of each protein is highly repetitive and is organized as shown in Figure 3. We base this on the fact that restriction mapping shows no detectable introns in the repetitive regions of either of the MiSp genes and that alternating spacer-repetitive regions comprise the majority of the coding regions. This lack of introns has also been noted for *Chironomus* (Wieslander, 1994) and *Bombyx* silk genes (Tsuijimoto & Suzuki, 1979a). This is based on the restriction mapping results which could detect no introns and showed the restriction pattern to be consistent with the cloned sequences for nearly the entire length of the gene.

```
1  SRLSSAEASSRISSAASTLVSSGGYLNTAALPSVISDLFAQVGASSPGVS
2  --------C----A----------S-----------------S-------
3  -----PQ----V---V-N--A-- PT-S---S-T--NVVS-I---N--L-
4  ---A-PDSGA-VA--V-N----- PTSS---S----NAVS-I---N--L-
5  V-----G-A--V--     N-AAIASAGA----N---NIYSG-LS-  ---

1  DSEVLIQVLLEIVSSLIHILSSSSVGQVDFSSVGSSAAAVGQSMQVVMG*
2  GN-------------------------------------------------
3  GCD----A------ACVT------I---NYGAASQF-QV----VLSAF-
4  GCD----A---V--A--Q--G---I---NYG-A-QATQI----VYQAL--
5   S--A---A---VI-A---V-G-A-I-N-  ----VNS-LNAVQNA-GAYAG
```

1 = Minor Ampullate Spidroin 1
2 = Minor Ampullate Spidroin 2
3 = Major Ampullate Spidroin 1
4 = Major Ampullate Spidroin 2
5 = *A. diadematus* ADF-1

Fig. 5. Spider silk carboxy-terminal consensus region. Amino acid sequence of the silk consensus region downstream from the repetitive regions found in all known spider silks. All sequences were deduced from cDNA sequences. Dashes indicate identical residues. MiSp 1 = 1, MiSp 2 = 2, MaSp 1 = 3, MaSp 2 = 4, *A. diadematus* ADF-1 = 5.

Discussion

The sequence, GAGAGA, is common among the MiSps and *B. mori* silks (Mita et al., 1994), but not present in either of the dragline silk proteins. $(GGX)_n$ motifs are present in MiSp 1, MiSp 2, and MaSp 1. Each MiSp protein repeat has two $(GGX)_3$ motifs and MaSp 1 has one $(GGX)_8$ motif per repeat. MaSp 2 has a glycine-rich motif that differs from the other spider silk proteins in that proline occupies every fifth residue in the MaSp 2 glycine-rich motif. All four spider silk proteins have regions of repeating alanine residues which vary in length from ten alanines (MaSp 2) to as few as three (MiSps). Silkworm silk fibroin has an occasional nonconserved tri-alanine sequence as well. Alanine regions of the MiSp proteins are always flanked by stretches of GA. ADF-1, a presumed minor ampullate silk protein from *A. diadematus*, is similar to the MiSps (Guerette et al., 1996). The 174 amino acid ADF-1 sequence reported consists of two repeating domains: GGYGQGY and $(GA)_y(A)_z$. However, unlike the MiSps, tyrosines and glutamines are not as highly conserved and the length of each repeat varies. The conserved carboxy-terminal region of ADF-1 shows high sequence conservation with the *Nephila clavipes* protein sequences (Fig. 5). While there are 6 sequence differences between MiSp1 and MiSp2, ADF-1 shows 46 substitutions and 3 gaps. No ADF-1 spacer region was described; however, the sequence data reported in Genbank for ADF-1 include a possible spacer region at the 5′ end of the sequence. This non-repetitive region differs in sequence from our sequence but a 13 amino acid region is identical (ESTASSAASSASS) suggesting this region is a spacer region. A flagelliform silk protein also contains highly conserved spacer regions which differ greatly in sequence from the minor ampullate silk protein spacers (Hayashi & Lewis, 1998).

Few structural analyses have been performed on minor ampullate silk primarily due to the lack of sequence information. The only secondary structure detected by FTIR analysis and fiber X-ray diffraction is β-sheet oriented parallel to the fiber axis (Dong, 1992; Parkhe et al., 1997). Silkworm and dragline silks also have a predominance of β-sheet that are formed by GAGAGA and AAAAA domains, respectively. When assigning secondary structure to specific silk sequences, it is useful to examine the β-sheet side chain spacing of known sequences measured by X-ray diffraction. Early diffraction studies showed that the side chain spacing for *B. mori* silk fibroin (GAGAGSGA) and polyalanine (β-sheet form) are 9.3 Å and 10.7 Å, respectively (Brown & Trotter, 1960; Warwicker, 1960). More recently, side group spacing was defined for a synthetic GAGAGA β-sheet (8.9–9.2 Å) and for major (10.77 Å) and minor (10.45 Å) ampullate silks (Fossey et al., 1991; Parkhe et al., 1997). Neither spacer regions nor GGX domains (X = Y, Q, or R) participate in β-sheets because the side chain spacing eliminates the inclusion of larger residues in the sheets, which would require at least 5 Å of additional space. Although the MiSps have short alanine regions, the flanking glycine-alanine stretches also likely form β-sheets. The GAGA domains are similar to silkworm fibroin crystalline regions. Minor ampullate silk side group spacing (10.45 Å) is consistent with the side chain spac-

ing of combined GAGAGA (8.9–9.3 Å) and AAAAA (10.7 Å) domains. Based on diffraction data and extrapolation of NMR data from *B. mori* and major ampullate silk, the $(GA)_y(A)_z$ domains of minor ampullate silk primarily form β-sheets that comprise the crystalline regions of the fiber. Structures of the spacer regions and GGX domains have not been determined because the only detected overall structure has been β-sheet.

A working model has been developed for minor ampullate silk based on deduced amino acid sequence (this work) and biophysical data (Parkhe et al., 1997). Multiple β-strands form sheets which are stacked to form the crystalline regions similar to those described for *B. mori* fibroin and dragline silk. These sheets involve both inter- and intra-molecular interactions. Each repeat unit has two domains of about 15 residues that participate in β-sheets. Each β-strand domain is separated by a GGX domain. Repetitive regions, each composed of about 20 β-sheet forming domains, are flanked by 137 amino acid spacer regions. There are about four large scale repetitive regions in MiSp 1, thus a large amount of the protein is in a β-sheet structure. The repetitive region is semicrystalline because of the large fraction of β-sheet within that region. The spacer-repetitive region unit is the highest level of repeating structure on the molecular level. Spacers may serve as a matrix for embedding the crystalline regions.

Silkworm fibroin, minor and major ampullate silks all have similar types of crystalline β-sheets and all have tensile strengths of the same order of magnitude. These high tensile strengths can be attributed to the additive effect of large numbers of hydrogen bonds formed by stacked stacks of β-sheets. The additional hydrophobic interactions of the alanine residues in major ampullate silk maybe responsible for its higher tensile strength. FTIR studies monitored major and minor ampullate silk fibers as they were stretched. Both silks were shown to have substantial β-sheet when unstretched (Dong, 1992). When axial tension was applied, the β-sheets of dragline silk remained constant, but the disordered regions appeared to diminish with a concomitant appearance of coiled structures. Minor ampullate silk only exhibited conformational changes just prior to breaking, when the β-sheet structure broke down and the fiber deformed. The minor ampullate silk model can account for this result because the flexibility provided by the GGX and spacer regions cannot reversibly withstand the same axial tension that dragline silk can since the turn structures in the MaSp 2 protein provide elasticity.

Minor ampullate silk is composed of two similar proteins that are predominately β-sheet. While there are some similarities among the silkworm and spider silk proteins, differences exist which are apparent in mechanical properties. Defining the structure and function relationship of the silk class of structural proteins is in its early stages because of the difficulties encountered with large, insoluble, repetitive proteins. β-sheet-forming glycine-alanine and possibly helical glycine-rich domains form the repetitive regions, which alternate with nonrepetitive spacer regions. The spacer-repetitive region unit repeats throughout the majority of the two proteins. The deformation upon stretch is probably due to the irreversible extension of the spacer regions. The large number of β-sheets formed by the interaction of many MiSp 1 and 2 proteins may account for the high tensile strength of the minor ampullate silk fiber.

Acknowledgments

We thank Dr. Holly Steinkraus for performing northern blotting experiments. Supported by U.S. Army Research Office Grant. Sequences have been deposited in GenBank, AF027735, AF027736, and AF207737.

References

Beckwitt R, Arcidiacono S. 1994. Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). *J Biol Chem 269*:6661–6663.

Brown L, Trotter IF. 1960. X-ray studies of poly-L-alanine. *Faraday Soc Trans 52*:545–551.

Dong Z. 1992. [Doctoral dissertation]. University of Wyoming. pp 50–52.

Dong Z, Lewis RV, Middaugh CR. 1991. Molecular mechanism of spider silk elasticity. *ArchBiochem Biophys 284*:53–57.

Fossey SA et al. 1991. Conformational energy studies of β-sheets of model silk fibroin peptides. I. Sheets of poly(Ala-Gly) chains. *Biopolymers 31*:1529–1541.

Gillespie DB, Viney C, Yager P. 1994. Raman spectroscopic analysis of the secondary structure of spider silk fiber. In Kaplan D, Adams W, Farmer B, Viney C, eds. *Silk polymers*. Washington, DC: American Chemical Society. pp 155–167.

Gosline JM, DeMont ME, Denny ME. 1986. The structure and properties of spider silk. *Endeavor 10*:37–43.

Guerette PA, Ginzinger DG, Weber BHF, Gosline JM. 1996. Silk properties determined by gland specific expression of a spider fibroin gene family. *Science 272*:112–115.

Hayashi CY, Lewis RV. 1998. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J Mol Biol 275*:773–784.

Herrmann BG, Frischauf AM. 1987. Isolation of genomic DNA. *Methods Enzymol 152*:180–192.

Hinman MB, Lewis RV. 1992. Isolation of a clone encoding a second dragline silk fibroin. *J Biol Chem 267*:19320–19324.

Hinman MB, Stauffer SL, Lewis RV. 1994. Mechanical and chemical properties of certain spider silks. In: Kaplan D, Adams W, Farmer B, Viney C, eds. *Silk polymers*. Washington, DC: American Chemical Society. pp 222–233.

Lewis RV. 1992. Spider silk: The unraveling of a mystery. *Acc Chem Res 25*:392–398.

Lucas F. 1986. Spiders and their silk. *Discovery 25*:20–26.

Mertz LM, Rashtchian A. 1994. Nucleotide imbalance and polymerase chain reaction: Effects on DNA amplification and synthesis of high specific activity radiolabeled DNA probes. *Anal Biochem 221*:160–165.

Mita K, Ichimura S, James TC. 1994. Highly repetitive structure and its organization of the silk fibroin gene. *J Mol Evol 38*:583–592.

Parkhe AD, Seeley SK, Gardner K, Thompson L, Lewis RV. 1997. Structural studies of spider silk proteins in the fiber. *J Molec Recog 10*:1–6.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*, second edition. New York: Cold Spring Harbor Laboratory Press.

Simmons A, Michal C, Jelinski L. 1996. Molecular orientation and two-component nature of the crystalline fracaation of spider dragline silk. *Science 271*:84–87.

Stauffer SL, Coguill S, Lewis RV. 1994. Comparison of physical properties of three silks from *Nephila clavipes* and *Araneus gemmoides*. *J Arachnol 22*:5–11.

Tsuijimoto Y, Suzuki Y. 1979a. Structural analysis of the fibroin gene at the 5′ end and its surrounding region. *Cell 16*:425–431.

Tsuijimoto Y, Suzuki Y. 1979b. The DNA sequence of *Bombyx mori* fibroin gene including the 5′ flanking, mRNA coding, entire intervening and fibroin protein coding regions. *Cell 18*:591–600.

Vollrath F. 1992. Spider webs and silks. *Scientific American 266*:70–76.

Warwicker JO. 1960. Comparative studies of fibroins II. The crystal structure of various fibroins. *Faraday Soc Trans 52*:554–557.

Wieslander L. 1994. The balbiani ring multigene family: Coding repetitive sequences and evolution of a tissue-specific cell function. *Prog Nucleic Acid Res Mol Biol 48*:275–313.

Xu M, Lewis RV. 1990. Structure of a protein superfiber: Spider dragline silk. *Proc Natl AcadSci USA 87*:7120–7124.