Isolation of a Clone Encoding a Second Dragline Silk Fibroin

NEPHILA CLAVIPES DRAGLINE SILK IS A TWO-PROTEIN FIBER*

(Received for publication, March 16, 1992)

Michael B. Hinman and Randolph V. Lewis†

From the Department of Molecular Biology, University of Wyoming, Laramie, Wyoming 82071

Spider dragline silk is a unique protein fiber possessing both high tensile strength and high elasticity. A partial cDNA clone for one dragline silk protein (Spidroin 1) was previously isolated. However, the predicted amino acid sequence could not account for the amino acid composition of dragline silk. We have isolated a partial cDNA clone for another dragline silk protein (Spidroin 2), demonstrating that dragline silk is composed of multiple proteins. The amino acid sequence exhibits an entirely different repetitive motif than Spidroin 1. Spidroin 2 is predicted to consist of linked β-turns in proline-rich regions which alternate with β-sheet regions composed of polyalanine segments. This structure for Spidroin 2 provides a model for dragline silk structure and function.

Spiders can generate several protein silks from specialized glands designed for diverse functions, unlike insects which produce only one type of silk. Nephila clavipes, an orb-web spinning spider, has six different types of glands (1, 2), each producing a different silk. The presence of multiple silks within a single organism presents a unique opportunity to examine protein structure and function. These silks are related by a predominance of alanine, serine, and glycine in their amino acid compositions. They are synthesized in specialized cells at the tail of their respective glands, secreted into a glandular lumen, and finally extruded on demand through a duct and valve system ending in a spinneret. However, the different silks vary widely in mechanical properties, implying disparate amino acid sequences. Several factors contributed to the choice of dragline silk from the major ampullate gland of N. clavipes as the first silk to be studied, including: (i) its singular combination of high tensile strength and high elasticity (3–6); (ii) ease of gathering large quantities of the silk (7); (iii) the size and accessibility of the gland; and (iv) previous extensive physical and chemical characterization (2, 8–11). Most structural models of dragline silk suggest a pseudocristalline protein having significant proportions of antiparallel β-sheet interspersed with regions of undefined structure which are thought to be responsible for the elastic properties (7, 12, 13). The amino acid sequence, which would give a great amount of information with regard to possible structure, was exceedingly difficult to obtain due to the solvent- and enzyme-resistant nature of dragline silk (10, 11). We made cDNA clones from the messenger RNA of the major ampullate glands in order to determine the protein’s primary structure and gain insight into possible higher order structures.

The first clone isolated encoded a protein, Spidroin 1, containing repetitive elements which were not rigorously conserved in terms of sequence but did preserve a structural identity (14). Each structural repeat was composed of two segments, a polyalanine segment of 6–9 amino acids followed by a (Gly-Gly-Xaa), segment, where n ranged from 5 to 11. However, the sequence contained virtually no proline, known to be 3.5% of the amino acid composition, and only two-thirds of the tyrosine found in N. clavipes dragline silk (11, 15). A pentapeptide, Gly-Tyr-Gly-Pro-Gly, purified from the acid-cleaved silk at the same time as those used as the basis for cloning Spidroin 1 was not present in the predicted sequence of Spidroin 1.

Materials and Methods

Isolation of the Spidroin 2 Clone—Twelve cDNA clones, in Bluescript (Stratagene), were isolated from a previously constructed library (14) using the oligonucleotide 5′-CCNGGNNCATANCC-3′ as a probe. The probe was labeled with 32P by kinase (16) and hybridized to picked colonies (14) using the method of Wood et al. (17) employing tetramethyl ammonium chloride in the final 47°C wash. Plasmid DNA was prepared,1 digested with EcoRI and BamHI, and subjected to agarose gel electrophoresis on a 1% agarose (Bethesda Research Laboratories) gel. The DNA was transferred to Zetaprobe (Bio-Rad) using a Vacublot technique (18). The resulting Southern blot was hybridized (17) with the same probe under the same conditions as the original colony screening. Four clones showed reasonable hybridization and pMASS2 was chosen because it contained the largest insert, approximately 2 kilobases.

Nested Deletion Sequencing—pMASS2 was subjected to exonuclease III (Erase-a-Base kit, Promega) and used in a nested deletion strategy for sequencing. DNA from each time point was prepared using a quick plasmid preparation method1 and subjected to double-stranded sequencing with Sequenase (United States Biochemicals) utilizing the 7-deaza-GTP sequencing kit (19). Each strand was sequenced at least two times, some sections as many as five times. This strategy helped resolve compressions seen during coding strand sequencing and aided in organizing the sequenced deletions, as the pMASS2 insert was highly repetitive at the nucleotide level as well as the peptide level.

Computer Predictions of Hydrophilicity and Secondary Structure—Hydrophilicity profiles and secondary structure predictions were generated using MacVector, version 3.5 (International Biotechnologies, Inc.) on a Macintosh IIfx (Apple Computer, Inc.) personal computer. Hydrophilicity was estimated using Kyte-Doolittle methods (20) with a window size of 7 to see smaller structures without background. Secondary structure was predicted using Chou-Fasman (21) algorithms and Robson-Garnier (22) algorithms separately, then generating a consensus prediction.

* This work was supported by Grant 28457-LS from the Army Research Office. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Molecular Biology, University Station, Box 3944, University of Wyoming, Laramie, WY 82071.

1 Y. Kadokami and R. V. Lewis, manuscript in preparation.
predicted amino acid sequence. The start of the 16-base long
There is a relatively short 3' noncoding region with the stop
codon (nucleotide 1882) occurring 97 bases before the poly(A)-
tail. The frequency of codon usage in Spidroin 2 (Table 1) is
very similar to that of Spidroin 1 (14). For example, glycine,
the most prevalent amino acid in both proteins, shows an
89% preference for A or T as the third nucleotide in dragline
silk can proceed more efficiently if the fibroin mRNA is heated
to be explained by considering the secondary structure of the
loops formed between nearby polyalanine coding regions if
the third base were also G or C instead of A.

The DNA sequence is shown in Fig. 1 along with the
putative polyadenylation signal site and the beginning of the poly(A)-
tail. The frequency of codon usage in Spidroin 2 (Table 1) is
very similar to that of Spidroin 1 (14). For example, glycine,
the most prevalent amino acid in both proteins, shows an
89% preference for A or T as the third nucleotide in dragline
silk can proceed more efficiently if the fibroin mRNA is heated
to be explained by considering the secondary structure of the
loops formed between nearby polyalanine coding regions if
the third base were also G or C instead of A.
elements, shown in Fig. 1, consist of a highly conserved region and divergent COOH-terminal tail. The dashes represent deletions, allowing the elements to be arranged for maximum identity.

100 °C and quickly cooled, suggesting that a G/C-rich message may already have a great degree of secondary structure which would be composed of having G or C in the third position of the codons.

The next 65 residues show less conservation, and the last 98 residues are widely divergent. The pentapeptide on which our probe was based, Gly-Tyr-Gly-Gly-Ser, is very highly conserved and leads into a series of 6

Table 1: Codon frequencies for Spidroin 2 cDNA

| Codon (Amino acid) | Count | Codon (Amino acid) | Count | Codon (Amino acid) | Count |
|--------------------|-------|--------------------|-------|--------------------|-------|
| TTT (F)            | 1     | CCA (P)            | 59    | AGT (S)            | 17    |
| TTC (F)            | 1     | CCG (P)            | 0     | AGC (S)            | 3     |
| TTA (L)            | 2     | CAT (H)            | 0     | AGA (R)            | 2     |
| TTG (L)            | 3     | CAC (H)            | 0     | AGG (R)            | 0     |
| TCT (S)            | 2     | CGA (Q)            | 72    | GTT (V)            | 7     |
| TCC (S)            | 3     | CAG (Q)            | 2     | GTC (V)            | 2     |
| TCA (S)            | 10    | CGT (R)            | 2     | GTA (V)            | 3     |
| TCG (S)            | 0     | CGC (R)            | 0     | GTG (V)            | 1     |
| TAT (Y)            | 30    | CGA (R)            | 0     | GCT (A)            | 23    |
| TAC (Y)            | 1     | CGG (R)            | 0     | GCC (A)            | 32    |
| TAA (*)            | 1     | ATT (I)            | 3     | GCA (A)            | 75    |
| TAG (*)            | 0     | ACG (T)            | 4     | GGG (G)            | 2     |
| TGT (C)            | 1     | ATC (I)            | 0     | GAT (D)            | 2     |
| TGC (C)            | 1     | ATG (M)            | 0     | GAC (D)            | 0     |
| TGA (*)            | 0     | ACT (T)            | 1     | GAA (E)            | 2     |
| TGG (W)            | 0     | ACC (T)            | 1     | GAG (E)            | 0     |
| CTI (L)            | 1     | ACA (T)            | 0     | GGT (G)            | 20    |
| CTG (L)            | 3     | ACG (T)            | 0     | GGC (G)            | 19    |
| CTA (L)            | 2     | AAT (N)            | 0     | GGA (G)            | 143   |
| CGT (P)            | 22    | AAC (N)            | 2     | GGG (G)            | 2     |
| CCC (F)            | 5     | AAG (K)            | 0     | GAG (G)            | 0     |

ASASASSRSLSPCASAASVSAVSMLVSSGTPSNASALSTISNVSQIGAS Spidroin 1
ASASASSRSLSPCASAASVSAVSMLVSSGTPSNASALSTISNVSQIGAS Spidroin 2

FIG. 2. Predicted amino acid sequence for the Spidroin 2 protein, rearranged to show repetitive elements. The most repetitive protein elements have been arranged from the amino-terminal through the highly conserved region, followed by the less conserved region and divergent COOH-terminal tail. The dashes represent deletions, allowing the elements to be arranged for maximum identity.

While the repetitive regions of Spidroin 1 and Spidroin 2 show very little homology, there is a 49-amino-acid region in the COOH-terminal domain of Spidroin 2 with 80% identity to a corresponding region in Spidroin 1 (Fig. 3). These regions both follow the highly repetitive regions of their respective proteins.

DISCUSSION

We believe Spidroin 1 and Spidroin 2 are the only two major protein subunits of dragline silk from N. clavipes. All of the peptides reported in a previous paper (14), the peptide used for designing our probe (Gly-Tyr-Gly-Pro-Gly), and over 20 smaller peptides produced by partial acid digestion of dragline silk are found within the predicted protein sequences of Spidroin 1 and Spidroin 2. The total amino acid composition of dragline silk can be accounted for by a combination of the amino acid compositions of the two Spidroins. Using the cloned cDNAs as probes, each Spidroin exhibits only a single hybridizing mRNA band in extracts from the major ampullate gland, indicating a unique mRNA for each protein.

The overall 3.5% proline content of dragline silk can be accounted for by averaging the high level of proline (13.5%) in Spidroin 2 with the negligible amount of proline in Spidroin 1. Using their respective cDNAs as probes, the sizes of the mRNAs for Spidroin 1 and Spidroin 2 have recently been determined as 5.6 and 3 kilobases, respectively. We calculated the amino acid contribution of the two Spidroins to the total amino acid composition of dragline silk by assuming that: 1) the length of each protein was proportional to the size of its cDNA; and 2) the amino acid composition of dragline silk can be accounted for by a combination of the amino acid compositions of the two Spidroins.

Fig. 3. Comparison of a conserved sequence found in the COOH-terminal regions of Spidroin 1 and Spidroin 2. Amino acids 646-694 of Spidroin 1 (from 14) are compared with amino acids 526-574 of Spidroin 2. Amino acids conserved in both sequences are in bold type.

10 alanines with an occasional conservative serine substitution.

\[^{2}\text{M. Xu and R. V. Lewis, unpublished results.}\]

\[^{3}\text{Y. Kadokami and R. V. Lewis, manuscript in preparation.}\]
mRNA; 2) the repetitive regions of both proteins extended into their respective uncharacterized amino-terminal regions; and that 3) the molecular ratio of Spidroin 1 to Spidroin 2 is about 3 to 2. The calculations indicate that Spidroin 1 and Spidroin 2 are sufficient to account for the total amino acid composition. It should be noted that within a single strand of silk from one spider, the amino acid composition varies from site to site (11). Varying amounts of proline have also been found in dragline silk from a single spider over its lifetime (24). These variations in amino acid composition can be accounted for if the expression of the two Spidroin proteins is independently regulated.

The predominant use of A or T (U) as the third nucleotide in the codons for glycine and alanine is not as pronounced as that found for B. mori silk (25). However, it has been shown that the production of dragline protein from spider silk mRNA is subject to translational pauses (26, 27) in the same manner as B. mori silk (28) under both in vivo and in vitro conditions. Both organisms appear to produce gland specific pools of tRNAs for glycine and alanine as these amino acids represent the major structural components of the silk proteins (22, 29, 30). Efficient in vitro translation of the silk proteins from both organisms requires tRNA supplementation from appropriately conditioned glands (27, 28, 30). While the pattern of tRNA accumulation in B. mori is approximately proportional to the use of corresponding amino acids (31), alanine tRNA is the most abundant species in N. clavipes major amputall glands (30). The difference can probably be attributed to the different nature of the repetitive elements of the two silks. Since the alanines in both Spidroin 1 and Spidroin 2 are found in clusters of 6-10, as opposed to the dispersed nature of B. mori alanines (Gly-Ala-Gly-Ala-Gly-Ser), it may be necessary to have a larger pool of alanine tRNA during translation of spider silk to get through the areas of repetitive alanine usage.

The overall repetitive nature of the amino acid sequence of Spidroin 2 is emphasized by computer predictions of the hydrophilicity and secondary structure of the protein (Fig. 4). Computer predictions of secondary structure using Chou-Fasman (CF) and Robson-Garnier (RG) statistical methods predict a number of turns in the proline-rich regions and helical conformations for the polyalanine regions. The Robson-Garnier method predicts significant regions of β-sheet, especially in the polyalanine segments, but the Chou-Fasman method confines any sheet-like structure to the hydrophobic COOH-terminal tail. The consensus prediction (CF/RG) shows that there are numerous turns in the proline-rich regions alternating with polyalanine segments which are predicted to be helix-forming.

We originally believed that the polyalanine segments were likely to form α-helices (10), based on computer predictions for Spidroin 1 and Spidroin 2 and physical studies of peptides in aqueous solutions (10). Based on arguments which follow, we now believe that Spidroin 2 is composed of β-turn structures in the proline-rich regions alternating with β-sheet regions formed by the polyalanine segments. We also believe the polyalanine regions in Spidroin 1 are in the β-sheet conformation.

The β-turn structure, difficult to detect with x-ray diffraction, Fourier transform infrared spectroscopy, or Raman spectroscopy, is the most likely conformation for the pentapeptide repeats of Spidroin 2 to adopt (32-34). Gly-Pro-Gly-Xaa-Yaa repeats are known to form β-turns and a Gln in position 4 may indicate type II β-turns (54). The β-turn conformation has been demonstrated for similar proline-rich sequences in unrelated proteins such as gluten, synaptophysin, and elastin (32, 33).

The polyalanine regions of Spidroins 1 and 2 probably form the antiparallel β-sheets observed by many studies (3–8, 35). X-ray diffraction studies of dragline silk from our laboratory and others (36) show a 10.6Å spacing between stacks of pleated sheets in dragline silk. This corresponds to the spacing found for β-sheets formed by polyalanine (37). The Gly-Gly-Xaa repeats of Spidroin 1 have been shown to be unfavorable for β-sheet formation in silk (38). Studies on polyalanine peptide crystals containing low amounts of water demonstrate only the presence of β-sheet structures (39), and spider dragline silk contains less than 6% water, corresponding to less than one water molecule per 3 amino acids.

Under tension, the linked β-turns in Spidroin 2 may form a β-turn spiral or extend an already present spiral, as β-turns are known to have a degree of structural flexibility (32, 33). This would be similar to an elastic mechanism proposed for elastin (39) with entropic forces generated by the disruption of β-turns driving the retraction of the β-turn spiral. The high tensile strength of the fiber could be derived from the high proportion of β-sheet (oriented along the axis of the fiber). Elucidation of the primary structure (and predicted secondary structure) of Spidroin 2, and demonstration of the two protein subunit nature of dragline silk, has allowed disparate results from chemical and biophysical studies to be incorporated within a single model.

Acknowledgment—We thank Dr. Yoichi Kadokami for valuable discussions and technical assistance.

REFERENCES
1. Lucas, F. (1984) Discovery 25, 20-26
2. Koev, J. (1987) EcologySpiders (Nentwig, W., ed) pp. 180-188, Springer-Verlag, Berlin
3. Denny, M. W. (1976) J. Exp. Biol. 85, 483-505
4. Denny, M. W. (1980) Mechanical Properties of Biological Materials (Vincent, L., and Curry, M., eds) pp. 247-272, Cambridge University Press, Cambridge
5. Work, R. (1985) J. Exp. Biol. 118, 379-404
6. Godfray, J. M., DeMont, M. E., and Denny, M. W. (1986) Endeavor 10, 37-41
7. Work, R. W., and Emerson, P. D. (1982) J. Arachnol. 10, 1-10
8. DeMont, M. E., Denny, M. W., and DeMont, M. E. (1986) Nature 322, 551-552
9. Tillinghast, E. R., and Townley, M. (1987) Arachnology of Spiders (Nentwig, W., ed) pp. 203-210, Springer-Verlag, Berlin
10. Dong, Z., Lewis, R. V., and Middaugh, C. R. (1991) Arch. Biochem. Biophys. 284, 53-57
11. Work, R. W., and Young, C. T. (1987) J. Arachnol. 15, 65-80

* K. Matsuno and R. V. Lewis, manuscript in preparation.
Spider Dragline Silk Is a Two-protein Fiber

12. Warwicker, J. O. (1960) J. Mol. Biol. 2, 350-362
13. Hepburn, H. R., Chandler, A. D., and Davidoff, M. R. (1979) Insect Biochem. 9, 69-77
14. Xu, M., and Lewis, R. V. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 7120-7124
15. Hinman, M., Dong, Z., Xu, M., and Lewis, R. V. (1992) Structure, Cellular Synthesis and Assembly of Biopolymers, Springer-Verlag, Heidelberg, in press
16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
17. Wood, W. I., Gitachier, J., Leaky, L. A., and Lawn, R. M. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 1588-1588
18. Kadokami, Y., and Lewis, R. V. (1992) BioTechniques 12, 13-14
19. United States Biochemical (1989) Reagent Kit for Sequencing with Sequenase and 7-deaza-GTP, Cleveland
20. Kyte, J., and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132
21. Chou, P. Y., and Fasman, G. D. (1978) Annu. Rev. Biochem. 47, 251-276
22. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) J. Mol. Biol. 120, 87-120
23. Chavancy, G., Marbaix, G., Huez, G., and Cleuter, Y. (1981) Biochimie 63, 611-616
24. Osaki, S. (1989) Acta Arachnol. 37, 69-75
25. Mita, K., Ishimura, S., Zama, M., and James, T. C. (1988) J. Mol. Biol. 203, 917-925
26. Candelas, G., Candelas, T., Ortiz, A., and Rodriguez, O. (1983) Biochem. Biophys. Res. Commun. 116, 1039-1038
27. Candelas, G. C., Ortiz, A., and Ortiz, N. (1989) Biochem. Cell Biol. 67, 173-176
28. Lizardi, P. M., Mahcavi, V., Shields, D., and Candelas, G. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 6211-6215
29. Sprague, K. U., Hagenbuahle, O., and Zuniga, M. C. (1977) Cell 11, 561-579
30. Candelas, G. C., Arroyo, G., Carrasco, C., and Domenciel, R. (1990) Dev. Biol. 140, 215-220
31. Sprague, K. U. (1975) Biochemistry 14, 925-931
32. Matsuoku, N., Greitzer, C. E., and Kretzinger, R. H. (1990) Proteins Struct. Funct. Genet. 7, 125-155
33. Wasserman, Z. R., and Salemme, F. R. (1990) Biopolymers 29, 1613-1631
34. Wilmut, C. M., and Thornton, J. M. (1988) J. Mol. Biol. 203, 221-232
35. Warwicker, J. O. (1960) Faraday Soc. Trans. 52, 554-557
36. Fraser, R. D. B., and MacRae, T. P. (1975) Conformation in Fibrous Proteins, pp. 253-343, Academic Press, New York
37. Arnott, S., Dover, S. D., and Elliott, A. (1967) J. Mol. Biol. 30, 201-208
38. Lotz, B., Brack, A., and Spach, G. (1974) J. Mol. Biol. 87, 199-203
39. Hempel, A., Camerman, N., and Camerman, A. (1991) Biopolymers 31, 187-192