Secondary Structural Studies of Biosynthetic Polypeptides with a Repeating Sequence of Glycine-Rich Sequence of Spider Dragline Silk

Yasumasa Fukushima* and Hiroshi Nakajima
Research & Development Center, Unitika Ltd., 23 Kozakura, Uji, Kyoto 611

(Received July 14, 1997; CL-970551)

Polypeptides with a repeating sequence of glycine-rich sequence of spider dragline silk were synthesized in *E. Coli.* The polypeptide in the solid state formed a β-sheet structure which exists in crystalline region of spider silk.

Spider dragline silk is a high-performance fibrous protein with unique properties of high tensile strength and elasticity. Structural studies indicate that the dragline silk is semicrystalline protein composed of an antiparallel pleated β-sheet structure interspersed with elastic unordered segments.[1] The primary sequence of dragline silk of the spider *Nephila clavipes* was determined from partial complementary DNA clones, so it turned out that the silk comprised of predominantly two kinds of the repetitive amino acid s equences.[1] One is a 13 amino acid segment containing a polyalanine sequence, and the other is a 15 amino acid highly conserved segment composed of glycine-rich sequence with Ala, Tyr, Gln, and Leu.

It was presumed that the glycine-rich sequence formed a β-sheet secondary structure from diffraction patterns of transmission electron microscopy of the dragline silk[2] On the other hand, it was reported that the polyalanine sequence existed as a β-sheet conformation from solid state NMR measurements.[3] Therefore, it is not clear that which repeating sequence adopts the β-sheet structure of the spider silk.

Several groups have succeeded the syntheses of proteins with a repeating sequence of naturally occurring structural proteins such as silk, elastin, and collagen in *E. coli.*[4] They have described the general idea of the tandem gene production by self-ligation of repeating gene bearing nonpalindromic termini. We report here a successful biosynthesis of polypeptides with a repeating sequence of the glycine-rich sequence **1** in the similar manner and the secondary conformational analyses in the solid state.

GlyLeuGlyGlyGlnGlyGlyGlyAlaGlyGlnGlyGly
(1)

The oligonucleotide duplexes **2** and **3** were synthesized chemically employing β-cyanoethyl phosphoramidite method. The coding sequence was designed by selecting optimal codons for each amino acid in *E. coli.*[5] and *Ban*II site as nonpalindromic termini. After the duplex **2** which encodes sequence **1** was ligated between the *Eco*RI and *Bam*HI sites of cloning plasmid pUC18,[6] *E. coli* JM109 cells were transformed and screened by double-stranded sequencing using Pharmacia user protocol.

The 45-base pair monomer obtained by *Ban*II digestion of the recombinant plasmid was purified on a 8% polyacrylamide gel. The monomer was self-condensed enzymatically in head-to-tail fashion due to its nonpalindromic termini and a population of multimers with a chain length distribution in the range of 10 to 80 repeats was observed on the 1.5% agarose gel. The multimers were inserted into the unique *Ban*II site of pUC18-ADAPT plasmid constructed by ligating adapter sequence **3** between the *Eco*RI and *Hind*III sites of pUC18 and *E. coli* HB101 cells were transformed with the recombinant pUC18-ADAPT. The methionine residue at each adjacent side of the *Ban*II site is allowed to cleave expressed fusion proteins with cyanogen bromide in order to obtain the repetitive proteins.

```
        Gly Leu Gly Gly Gln Gly Gly Gly
AATTC ATAA GGG CTC GGT GGC CAG GGT GGC GGT
    G TATT CCC GAG CCA CCG GTC CCA CCG CCA
EcoRI       BanII

Ala Gly Gln Gly Gly Tyr Gly Gly Leu
GCT GGT CAG GGC GGT TAT GGT GGG CTC G
CGA CCA GTC CCG CCA·ATA CCA CCC GAG CCTAG
                          BanII  BamHI
```

(2)

```
   Met Ser Arg Thr Ala Val Ser Met Gly Leu
AATTCT ATG TCT AGA ACT GCA GTA TCC ATG GGG CTC
    GA TAC AGA TCT TGA CGT CAT AGG TAC CCC GAG
EcoRI                                  BanII

Met Gly Gly Met End End Ala Cys Pro Gly Ser
ATG GGC GGC ATG TAA TAG GCA TGC CCG GGA TCC A
TAC CCG CCG TAC ATT ATC CGT ACG GGC CCT AGG TTCGA
                                        HindIII
```

(3)

Seven plasmids with different repeating numbers (n = 4, 5, 6, 7, 11, 12, and 13) were isolated, and each fragment digested by *Pst*I and *Hind*III were inserted into the expression plasmid pRSETB which is characterized by T7 promoter driving high-level expression of inserted gene in the presence of T7 RNA polymerase and six consecutive histidine sequence for nickel chelating affinity purification. The fusion proteins were comprised of the repeating sequence and N- and C-terminal extensions as shown in sequence **4**.[7]

$$\mathrm{MRGSH_6GMASMTGGQQMGRDLYDDDDKDPSSRSAVSM(GLGGQGGGAGQGGYG)_nMGGM}$$
(4)

Each recombinant pRSETB plasmid containing the different multimer was expressed in *E. coli* JM109 by an addition of isopropyl-β-D-thiogalactopyranoside (IPTG) and an infection of M13 phage containing the T7 RNA polymerase gene. The fusion proteins were produced as soluble proteins. Figure 1 shows SDS-polyacrylamide gel of whole cell lysates stained with Coomassie Blue R-250 for estimating the production levels of the

fusion proteins ([SCAP]n is a polypeptide with sequence **4**). The production levels of the fusion proteins slightly increased as repeating number n decreased from 13 to 6 and were almost the same between n = 6 and 4. The electrophoretic mobilities of the proteins were consistent with those of expected molecular weight on SDS-polyacrylamide gel.

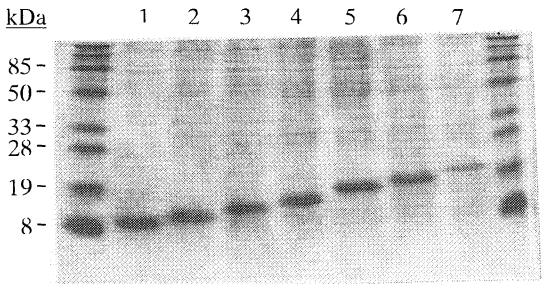

Figure 1. Production of sequence **4** fusion proteins in E. coli.as analyzed by 15% SDS-polyacrylamide gel of whole cell lysates stained with Coomassie Blue R-2 50. Lanes: (1) $[SCAP]_4$, (2) $[SCAP]_5$, (3) $[SCAP]_6$, (4) $[SCAP]_7$, (5) $[SCAP]_{11}$, (6) $[SCAP]_{12}$, (7) $[SCAP]_{13}$.

Each fusion protein was purified by nickel chelating affinity chromatography using protein purification protocol supplied by Invitrogen Inc. The purified proteins were obtained as a single band on SDS polyacrylamide gel (data not shown). Typical yields for fusion proteins $[SCAP]_{13}$, $[SCAP]_{12}$, $[SCAP]_{11}$, $[SCAP]_7$, and $[SCAP]_6$ were 1.2, 1.7, 3.8, 4.6, and 5.2 mg of lyophilized powder per liter of fermentation medium. The repeating portions were cleaved from the fusion constructs with cyanogen bromide and the fusion constructs were removed by dialysis. ^1H NMR spectrum was consistent with sequence **1**. The amino acid ratios found of the cleaved $[SCAP]_{13}$ were $Ala_{6.8}$ $Gly_{63.7}$ $Gln_{12.7}$ $Leu_{6.2}$ $Tyr_{6.1}$. (Anal. $Ala_{6.7}$ $Gly_{66.7}$ $Gln_{13.3}$ $Leu_{6.7}$ $Tyr_{6.7}$).

Figure 2 shows the fourier transform infrared spectra of the cyanogen bromide cleaved $[SCAP]_{13}$ in the solid state in the amide I and II regions which have been used for characterization of the secondary structure of many proteins.[8] The IR spectrum of the lyophilized form exhibits amide I and II vibrational modes at 1656 and 1539 cm^{-1}, respectively, characteristic of the disordered conformation. On the sample prepared as cast film from formic acid, the strong infrared vibration at 1626 and 1533 cm^{-1} can be assigned to the β-sheet conformation, and the weak amide I component at 1696 cm^{-1} can be characteristic of the regular alternation of the β-sheet chain direction. Although the vibration observed at 1650, 1661, and 1545 cm^{-1} can be attributed to secondary structures other than the β-sheet conformation, the polypeptide would form predominantly the β-sheet conformation by the treatment with formic acid.

These results indicate that a biosynthesis of a polypeptide with a repeating sequence is a effective method as one of

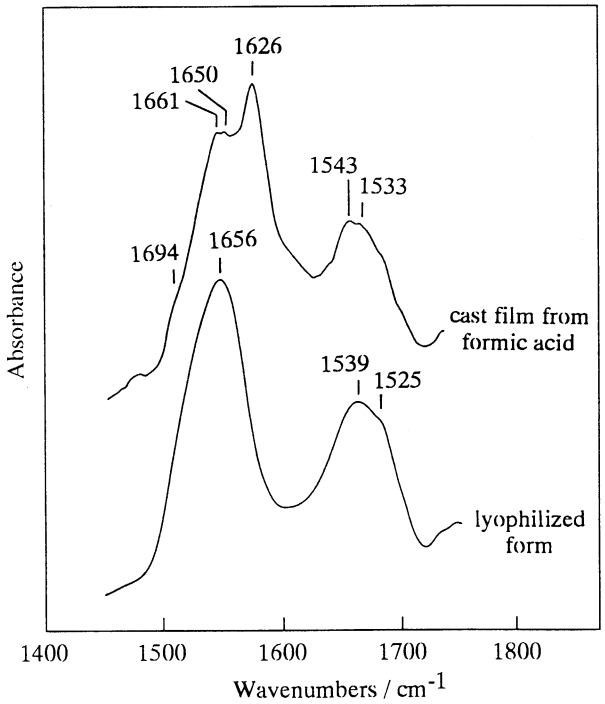

Figure 2. Fourier transform infrared spectra in the amide I and II regions on the samples of the cyanogen bromide cleaved $[SCAP]_{13}$ prepared in lyophilized form and as cast film from formic acid.

precise polymerization. The IR measurements of the synthetic repetitive polypeptide might suggest that the glycine-rich sequence of the spider dragline silk adopted predominantly the β-sheet conformation.

We thank Dr. M. Hirami of Unitika Ltd. for many helpful discussions and valuable comments on this work.

References

1 M. Xu and R. V. Lewis, *Proc. Natl. Acad. Sci. USA*, **87**, 7120 (1990).
2 B. L. Thiel, D. D. Kunkel, and C. Viney, *Biopolymers*, **34**, 1089 (1994).
3 A. H. Simmons, C. A. Michal, and L. W. Jelinski, *Science*, **271**, 84 (1996).
4 K. P. McGrath, M. J. Fournier, T. L. Mason, and D. A. Tirrell, *J. Am. Chem. Soc.*, **114**, 727 (1992).
5 S. Aota, T. Gojobori, F. Ishibashi, T. Maruyama, and T. Ikemura, *Nucleic Acids Res.*, **16**, r315 (1988).
6 C. Yanisch-Perron, J. Vieira, and Messing, *Gene*, **33**, 103 (1985).
7 Abbreviations for amino acid residues are as follows: A, Ala; D, Asp; G, Gly; H, His; K, Lys; L, Leu; M, Met; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; and Y, Tyr.
8 T. Miyazawa and E. R. Blout, *J. Am. Chem. Soc.*, **83**, 7 (1961).