J. Biochem. **137**, 721–729 (2005)
DOI: 10.1093/jb/mvi090

Design, Expression and Solid-State NMR Characterization of Silk-Like Materials Constructed from Sequences of Spider Silk, *Samia cynthia ricini* and *Bombyx mori* Silk Fibroins

Mingying Yang and Tetsuo Asakura*
Department of Biotechnology, Tokyo University of Agriculture and Technology, Koganei, Tokyo 184-8588

Received November 19, 2004; accepted March 28, 2005

Silk has a long history of use in medicine as sutures. To address the requirements of a mechanically robust and biocompatible material, basic research to clarify the role of repeated sequences in silk fibroin in its structures and properties seems important as well as the development of a processing technique suitable for the preparation of fibers with excellent mechanical properties. In this study, three silk-like protein analogs were constructed from two regions selected from among the crystalline region of *Bombyx mori* silk fibroin, $(GAGSGA)_2$, the crystalline region of *Samia cynthia ricini* silk fibroin, $(Ala)_{12}$, the crystalline region of spider dragline silk fibroin, $(Ala)_6$, and the Gly-rich region of spider silk fibroin, $(GGA)_4$. The silk-like protein analog constructed from the crystalline regions of the spider dragline silk and *B. mori* silk fibroins, $(\underline{A_6}\underline{SCS})_8$, that constructed from the crystalline regions of the *S. c.ricini* and *B. mori* silk fibroins, $(\underline{A_{12}}\underline{SGS})_4$, that constructed from and the crystalline region of *S. c.ricini* silk fibroin and the glycine-rich region of spider dragline silk fibroin, $(\underline{A_{12}}\underline{SGS})_4$,were expressed their molecular weights being about 36.0 kDa, 17.0 kDa and 17.5 kDa, respectively in *E. coli* by means of genetic engineering technologies. $(\underline{A_{12}}\underline{SCS})_4$ and $(\underline{A_{12}}\underline{SGS})_4$ undergo a structural transition from α-helix to β-sheet on a change in the solvent treatment from trifluoroacetic acid (TFA) to formic acid (FA). However, $(\underline{A_6}\underline{SCS})_8$ takes on the β-sheet structure predominantly on TFA treatment and FA treatment. Structural analysis was performed on model peptides selected from spider dragline and *S. c.ricini* silks by means of ^{13}C CP/MAS NMR.

Key words: *Bombyx mori* silk, ^{13}C CP/MAS NMR, genetic engineering, *Samia cynthia ricini* silk, silk-like protein analog, spider dragline silk.

The silk spun from domestic silkworm *Bombyx mori* silk, has been used commercially for biomedical sutures for decades, and for textile production for at least 2,500 years. The complete primary structure of *B. mori* silk fibroin has been determined by Mita *et al.* (*1*), and more recently by Zhou *et al.* (*2*). The *B. mori* silk fibroin is mainly composed of repeated motifs, $[GAGSGA]_n$ and $[GAGXGA]_n$, where X = Y or V. Motif $[GAGSGA]_n$ forms crystalline domains in *B. mori* silk fibroin, which has been considered to contribute to the strength of *B. mori* silk fibers.

On the other hand, spider silk has attracted the interest of scientists of various disciplines for a long time. A balanced combination of high tensile strength, stiffness, and elasticity makes these fibers one of the toughest, if not the toughest, materials of its kind known. Spider dragline silk is thought to be composed mainly of two proteins, spidroins I and II (*3*, *4*). These proteins can be described as block copolymers with alternating polyalanine and glycine-rich blocks. The polyalanine domains within the fully formed silk have previously been characterized by NMR and X-ray analysis and shown to form anti-parallel β-sheets (*5*–*8*). It is believed that the β-sheet crystals contribute to the high strength and stiffness. However, the glycine-rich regions were originally described as amorphous (*8*) or rubber-like (*9*). More recently, NMR spectra have suggested that 3_1-helix may be present in the glycine-rich regions of spidroin I (*10*). The 2D spin-diffusion ^{13}C NMR spectrum of the sequence of $(GGA)_{10}$ provided further forceful evidence of 3_1-helix formed in the glycine-rich region (*11*).

By contrast, for *Samia cynthia ricini*, the primary structure follows in general the design principles of spider dragline silk with polyalanine and glycine-rich regions. However, the polyalanine region (11–13 residues) is twice the length of the equivalent blocks (4–6 alanine residues) in spider dragline silk (Yukuhiro, K., personal communication). The polyalanine regions in *S. c.ricini* undergo a structural change from α-helix to β-sheet in the spinning process as the silk protein passes from the gland to the final fiber (*12*). The structural α-helix to β-sheet transition could also be driven by stretching of the protein in the silkgland or by treatment with methanol. It is reasonable to suggest that, as in spider dragline silk, the polyalanine β-sheet crystals contribute to the strength of *S. c.ricini* silk.

NMR spectroscopy provides direct evidence regarding the local structure and dynamics of a silk fibroin, as mentioned above. However, difficulties encountered in structural determination are due to the heterogeneity of the

*To whom correspondence should be addressed. Tel/Fax: +84-42-383-7733, E-mail: asakura@cc.tuat.ac.jp

Table 1. **The primary structures of silk-like protein analogs produced from *E. coli* and model peptides synthesized by the solid phase method.**

(1) {DGG(A)$_6$GGAASGAGYGA(GAGSGA)$_2$GAGYGA}$_8$ (A$_6$SCS)$_8$
(2) {DGG(A)$_{12}$GGAASGAGYGA(GAGSGA)$_2$GAGYGA}$_4$ (A$_{12}$SCS)$_4$
(3) {DGG(A)$_{12}$GGAASGAGYGA(GGA)$_4$GAGYGA}$_4$ (A$_{12}$SGS)$_4$
(4) DGG(A)$_6$GGA (A$_6$)
(5) DGG(A)$_9$GGA (A$_9$)
(6) DGG(A)$_{12}$GGA (A$_{12}$)
(7) GAGYGA(GAGSGA)$_2$GAGYGA (SCS)
(8) GAGYGA(GGA)$_4$GAGYGA (SGS)
(9) (GAGSGA)$_2$(AGYGAG)(AGYGAG)(AGYGAG) (CS3)
(10) GGLGGQGAG(A)$_6$GGAGQGGYGGLGSQGAGRGGQG(A)$_6$GGAGQG (SPM)
(11) GGAGGGYGGDGG(A)$_{12}$GGAGDGYGAG (SCM)

repeated sequences of the silk fibroins. Therefore, model peptides with defined primary structures selected from among the native silk sequences have been used to determine the local structures of the repeated sequences because they avoid the large variations in structural distribution resulting from the heterogeneity in the primary structure. Solvent treatment prior to the NMR measurements induces structural changes of these model peptides and thus provides a model for the conformation changes in the formation of silk fibers. A model peptide containing poly-Ala regions of *S. c.ricini* silk fibroin, GGAGGGYGGDGG(A)$_{12}$GGAGDGYGAG, was shown to take on an α-helical conformation after trifluoroacetic acid (TFA) treatment corresponding to the pre-spun state (*13*, *14*). In contrast after dissolution in formic acid (FA) and then drying in air, it adopted a mainly β-sheet structure, corresponding to that of the final silk fibers. Furthermore, a model peptide (AG)$_{15}$ after FA treatment also adopts a β-sheet structure corresponding to the silk II form of *B. mori* silk thread (*15*, *16*).

In this study, three silk-like protein analogs that consist of two regions selected from among the crystalline region of *B. mori* silk fibroin, (GAGSGA)$_2$, the crystalline region of *S. c.ricini* silk fibroin, (Ala)$_{12}$, the crystalline region of spider dragline silk fibroin, (Ala)$_6$, and the Gly-rich region of spider silk fibroin, (GGA)$_4$, were prepared, and their structures were characterized by means of solid state NMR. Before designing the analogs, NMR characterization were attempted for sequential model peptides containing repeated sequences of the crystalline domains in several silk fibroins. TFA and FA treatment of these samples was performed to produce the structures in the pre-spun state and in the final silk fibers, respectively. Fiber formation from silk-like protein analogs was attempted.

EXPERIMENT PROCEDURES

Design of Silk-Like Protein Analogs—The three silk-like protein analogs prepared here are listed in Table 1. The analogs were

(1) {DGG(A)$_6$GGAASGAGYGA(GAGSGA)$_2$GAGYGA}$_8$ (A$_6$SCS)$_8$,

(2) {DGG(A)$_{12}$GGAASGAGYGA(GAGSGA)$_2$GAGYGA}$_4$ (A$_{12}$SCS)$_4$,

(3) {DGG(A)$_{12}$GGAASGAGYGA(GGA)$_4$GAGYGA}$_4$ (A$_{12}$SGS)$_4$

where A$_6$ is the crystalline region of spider dragline silk fibroin, DGG(A)$_6$GGA, A$_{12}$ is the crystalline region of *S. c.ricini* silk fibroin, DGG(A)$_{12}$GGA, and C is the crystalline region of *B. mori* silk fibroin, (GAGSGA)$_2$. Here additional sequences, DGG in the N-terminal and GGA in the C-terminal, were introduced to stabilize the α-helix structure of the polyalanine region before spinning (*13*, *14*). The sequences GAGYGA and (GGA)$_4$ are denoted as S and G, respectively. The sequence GAGYGA is known to adopt a combination of distorted β-sheet and distorted β-turn conformations before and after spinning. Its hydrophilic Y residue probably contributes to the solubility of *B. mori* silk fibroin (*17*). It was hoped that the introduction of this hexapeptide would improve the solubility and flexibility of protein analogs compared with the relatively stiff rods produced with the combination of crystalline blocks, polyalanine and GAGSGA. The sequence (GGA)$_4$ is present in the glycine-rich region of spider dragline silk fibroin and is considered to take on the 3$_1$-helix conformation (*11*).

Gene Construction—*E. coli* strain DH5α was used for the propagation and construction of plasmids, and *E. coli* strain BL21(DE3)pLysS for the production of proteins. Synthesized DNA fragments having undergone phosphorylation were purchased from Sigma Genosys, Japan. Restriction enzymes and ligase were purchased from Takara Shuzo. Plasmid pUC118 was obtained from Takara Shuzo, and pET30a Novagen. All other chemicals were of analytical grade. Bacteria were grown in an enriched medium, and DNA manipulations were performed. The transformation conditions were as described by Sambrook *et al.* (*18*). DNA sequencing was performed with an ABI PRISM™ 377 Auto-sequencer according to the user's manual. *E. coli* cell density was measured at λ = 600nm with a Hitachi U-3200 spectrophotometer in quartz cuvettes with a path length of 1 cm. Batch cultures were performed in a MD-6C Fermentor (B.E. Marubishi Co), with a 1.2 litter vessel.

In Fig. 1, The oligonucleotide fragments encoding A$_6$ and A$_{12}$, and those encoding the combination of the semi-crystalline and crystalline regions of *B. mori* silk fibroin, GAGYGA(GAGSAG)$_2$GAGYGA (SCS), and the combination of the semi-crystalline region of *B. mori* silk fibroin and the sequence of (GGA)$_4$ from the Gly-rich region of spider dragline silk, GAGYGA(GGA)$_4$GAGYGA (SGS) are summarized. The complementary oligonucleotides were annealed by heating at 95°C and cooling slowly to room temperature over about 3 h. To construct the monomers, the duplex DNAs encoding the A$_6$SCS, A$_{12}$SCS and A$_{12}$SGS were ligated into *Bam*HI-digested pUC118. Multimers were obtained by using previously reported strategies involving head-to-tail ligation and orientation for *Nhe*I and *Spe*I sites (*19*). Multimerized DNA fragments encoding these three recombinant proteins were inserted into *Bam*HI and *Hin*dIII–digested pET30a. The ligated mixture was used to transform the expression host strain, BL21(DE3)pLysS. Identification of recombinant clones and expression plasmids was accomplished by restriction analysis with appropriate enzymes and sequencing.

Protein Expression and Purification—Expression vectors containing the verified genetic constructs were trans-

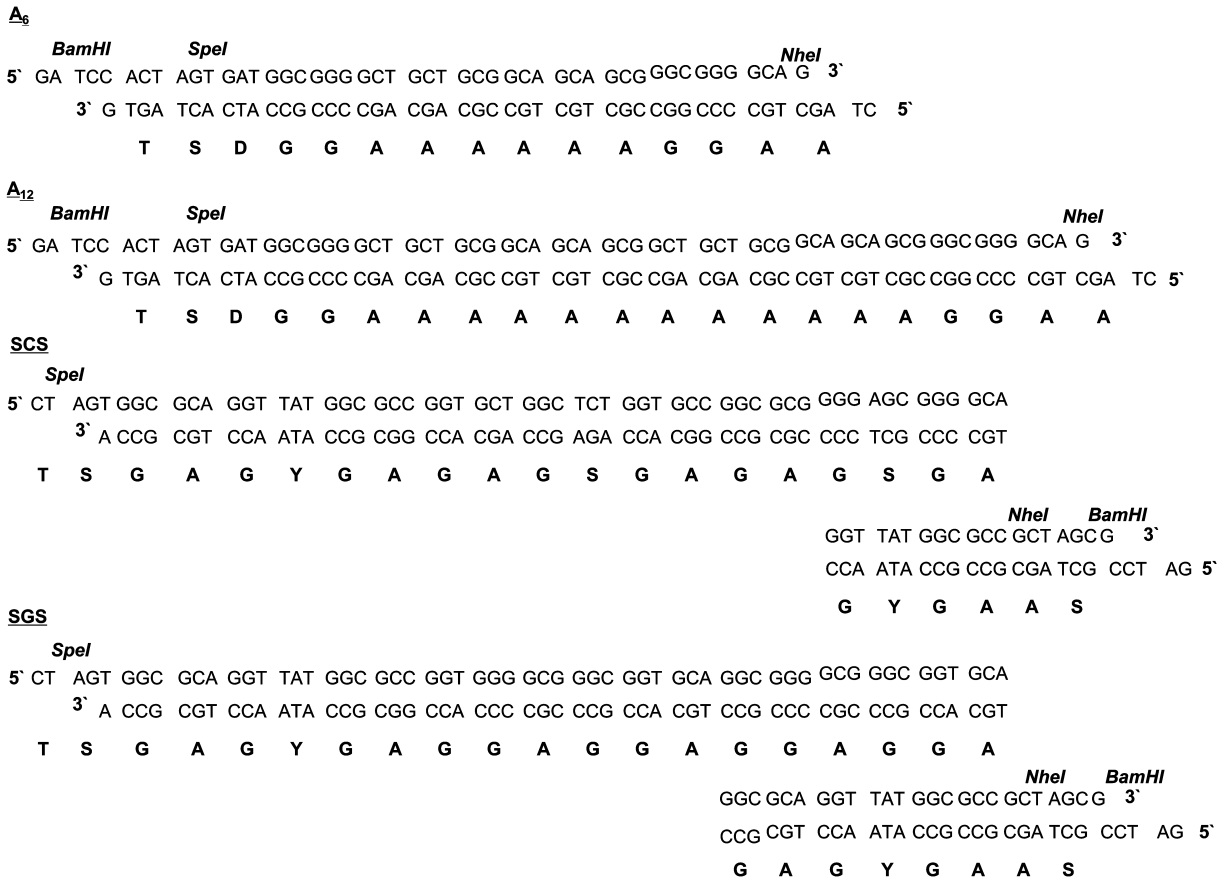

Fig. 1. **The designed oligonucleotide sequences for $\underline{A_6}$, (DGG$(A)_6$GGA), $\underline{A_{12}}$, (DGG$(A)_{12}$GGA), $\underline{SCS}$, (GAGYGA(GAGSGA)$_2$GAGYGA), and $\underline{SGS}$, (GAGYGA$(GGA)_4$GAGYGA) in three silk-like protein analogs, $(\underline{A_6}\underline{SCS})_8$, $(\underline{A_{12}}\underline{SCS})_4$ and $(\underline{A_{12}}\underline{SGS})_4$.**

formed into *E. coli* strain BL21(DE3)pLysS, and subsequent cultures were grown in Luria-Bertain broth (TB) containing chloramphenicol (25 μl/ml) and kanamycin (25 μl/ml). Batch cultures of 1 liter in volume were allowed to grow at 37°C to an optical density between 0.8–1.5, as determined from the optical absorbance at 600 nm (OD_{600}). Protein expression, under the control of a bacteriophage T7 promoter, was induced by the addition of IPTG to a final concentration of 1 mM. Expression was continued for 1–4 h before harvesting *via* centrifugation (8,500 rpm, 40 min, 4°C). The cells were collected and stored at –70°C before purification. Protein expression and the level of expression were confirmed by Western-blot analyses, a T7-tag antibody–horseradish peroxidase conjugate being used for visualization.

Protein Purification—Frozen cells were thawed on ice and resuspended in buffer B (100 mM NaH_2PO_4, 10 mM Tris·Cl, 8 M urea, pH 8.0) at 5 ml per gram wet weight, and then the cells were lysed by gently vortexing until the solution became translucent. Cellular debris was removed by centrifugation (14,500 rpm, 30 min, 4°C). To purify the fusion protein using His-tags, the supernatant was loaded onto a nickel chelate affinity column which has been charged with Ni^{2+}. The column was washed with buffer C (100 mM NaH_2PO_4, 10 mM Tris·Cl, 8 M urea, pH 6.3). The protein was eluted with buffer E (100 mM NaH_2PO_4, 10 mM Tris·Cl, 8 M urea, pH 4.5). The eluted protein was dialyzed against distilled water for 3 days and then lyophilized.

Cyanogen bromide cleavage of fusion proteins was performed by the method described by Fukushima *et al.* (*20*). Approximately 60mg of fusion protein was dissolved in 10ml of 99% formic acid and then diluted to 70% formic acid with distilled water. 100 mg of CNBr crystals was then added and the solution was stirred at room temperature for 24 h before transfer to a dialysis membrane (MWCo 3,500 Da). The solution was dialyzed against deionised water for 3 days, and then the cleaved protein was recovered by lyophilization. The purified proteins and cleavage fragments were confirmed by SDS–polyacrylamide gel electrophoresis.

Solid Phase Synthesis of Model Protein Analogs—Polyalanine peptides with different lengths selected from among the silk fibroin sequences: $\underline{A_6}$, $\underline{A_9}$ and $\underline{A_{12}}$, listed in Table 1 were synthesized by means of solid phase F-moc chemistry with a fully automated Pioneer Peptide Synthesis system (Applied Biosystems Ltd.), as described previously(*13*, *14*). Similarly, the longer sequential model peptide, GGLGGQGAG$(A)_6$GGAGQGGYGGLGSQGAGRGGQG$(A)_6$GGAGQG denoted as $\underline{SPM}$, derived from the sequence of spider dragline silk, and the longer peptide, GGAGGGYGGDGG$(A)_{12}$GGAGDGYGAG denoted as $\underline{SCM}$, derived from the sequence of *S. c.ricini* silk fibroin were synthesized (Table 1). The crude fractions were dissolved

Table 2. **Spinning coditions for fiber formation of silk-like protein analog {DGG(A)$_6$ GGAASGAGYGA(GAGSGA)$_2$GAGYGA}$_8$.**

Solvent	Concentration (%)	Coagulant	Fibers
FA	18	methanol	–
FA + 15% LiBr	25	90% methanol + 10% H_2O	–
HFA	10	methanol	–
60% FA + 40% HFA	25	90% methanol + 10% H_2O	+
50% FA + 50 HFA	20	40% methanol + 60% ethanol	+
DCA	12.5	methanol	++

–: fibers cannot be formed; +: fibers can be formed; ++: fibers can be formed and stretched in methanol.

in 9M LiBr and dialyzed extensively against distilled water, and then the peptides collected by lyophilization.

Solid State ^{13}C CP/MAS NMR Observation—TFA and FA treatment prior to NMR measurement was performed for the peptides and silk-like protein analogs to determine the conformations present before and after spinning. ^{13}C CP/MAS spectra were acquired with a Chemagnetics CMX-400 spectrometer operating at 100 MHz, with a CP contact time of 1.5 ms, TPPM (two-pulse phase-modulated) decoupling, and magic angle spinning at 5 kHz. A total of 10,000–20,000 scans were collected over a spectral width of 80 kHz, with a recycle delay of 0.5 s. Chemical shifts are reported in ppm relative to TMS as a reference.

Fiber Formation—In order to prepare fibers from the silk-like protein analogs, several dope solutions were prepared by dissolving the silk-like protein analogs in several solvents, as summarized in Table 2. The silk solutions were loaded into a 1ml syringe and then extruded at room temperature with moderate force through an internal 0.45 mm diameter hypodermic needle through an air gaps of 10mm. The coagulation baths are also summarized in Table 2. Attempts were made to post-draw the fibers in methanol before drying in air.

RESULTS AND DISCUSSION

Gene Construction—Electrophoresis of *Bam*HI-digested recombinant cloning vectors pUC118-($\underline{A_6}$SCS)$_8$, pUC118-($\underline{A_{12}}$SCS)$_4$, and pUC118-($\underline{A_{12}}$SGS)$_4$ against PCR markers confirmed the existence of DNA's of 1,056 bp, 600 bp, and 600 bp, respectively, as shown in Fig. 2. The codon for each amino acid was selected using codon usage in *E. coli* as previously reported. The 5′-*Bam*HI site at $\underline{A_6}$ or $\underline{A_{12}}$, and the 3′-*Bam*HI site at $\underline{\text{SCS}}$ or $\underline{\text{SGS}}$ are positioned at each end for insertion of the DNA into the *Bam*HI site of cloning plasmid pUC118. The *Spe*I and *Nhe*I restriction sites encoding Thr-Ser and Ala-Ser were inserted between the two *Bam*HI sites in order to isolate the DNA fragments from the recombinant plasmids. Multimerization could be achieved by insertion of the *Nhe*I–*Spe*I fragments into the *Nhe*I- or *Spe*I-digested vectors by taking advantage of the directional linkages produced by these two enzymes. These enzymes generate identical cohesive ends, and when the ligation joins *Nhe*I and *Spe*I sites together in a "head-to-tail" fashion (as in the self-condensation strategy), all internal restriction sites are destroyed. Subsequent digestion with *Nhe*I and *Spe*I will excise any inverted repeats within the multimerized fragments, ensuring correct maintenance of the protein reading frames. These multimers could also be inserted into either *Nhe*I- or *Spe*I-digested recombinant vectors to obtain larger genes of any desired degree of multimerization in a tightly controlled manner. As shown in Fig. 2, pUC118-($\underline{A_6}$SCS)$_8$, pUC118-($\underline{A_{12}}$SCS)$_4$, and pUC118-($\underline{A_{12}}$SGS)$_4$ appear to have been successfully constructed.

Protein Expression and Purification—Protein expression efforts involved the commercially available expression vector pET30a, which contains a [His]$_6$ sequence at the N-terminus and C-terminus repectively of the recombinant protein for purification by immobilized metal affinity chromatograghy. Therefore, the purified pUC118-($\underline{A_6}$SCS)$_8$, pUC118-($\underline{A_{12}}$SCS)$_4$, and pUC118-($\underline{A_{12}}$SGS)$_4$ DNA fragments were inserted into pET30a between the *Bam*HI and *Hin*dIII restriction sites, flanked by N- and C-terminal extensions of 53 and 19 amino acids, respectively. These terminal regions were derived from the transfer and expression vectors, and can be removed by CNBr cleavage at the flanking methionine residues. *E. coli* strain BL21(DE3)pLysS was used as the protein expression host. The pET30a-multimers were transformed into BL21(DE3)pLysS and the encoded proteins were expressed upon induction of IPTG. Figure 3 shows a SDS-page gel stained with Coomassie Blue R-250 for the [His]$_6$ tagged and cleaved proteins. Protein expressed from the pET30a vector can be purified from the Ni-NTA column through the [His]$_6$ sequence position at the N-

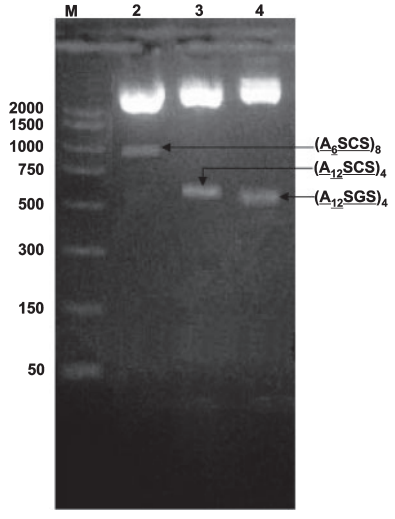

Fig. 2. ***Spe*I and *Nhe*I digestion analyses of multi-merized cloning vectors.** Lane 1, PCR markers; lane 2, pUC118-($\underline{A_6}$SCS)$_8$; lane 3, pUC118-($\underline{A_{12}}$SCS)$_4$; and lane 4, pUC118-($\underline{A_{12}}$SGS)$_4$.

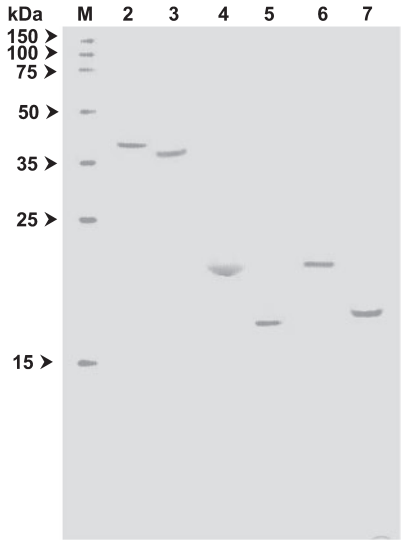

Fig. 3. **SDS-page analyses of purified proteins and cleaved proteins.** Lane 1, Perfect Protein™ Markers; lane 2, His-tagged $(\underline{A_6}SCS)_8$; lane3, $(\underline{A_6}SCS)_8$; lane 4, His-tagged $(\underline{A_{12}}SCS)_4$ (4); lane 5, $(\underline{A_{12}}SCS)_4$ (4); lane 6, His-tagged $(\underline{A_{12}}SGS)_4$; and lane 7, $(\underline{A_{12}}SGS)_4$.

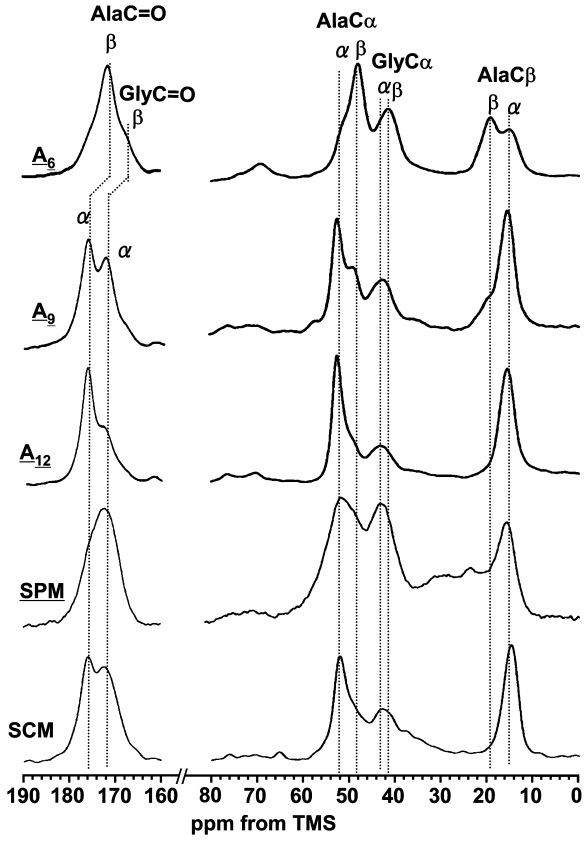

Fig. 4. **^{13}C CP/MAS NMR spectra of model peptides $\underline{A_6}$: DGG$(A)_6$GGA, $\underline{A_9}$: DGG$(A)_9$GGA, $\underline{A_{12}}$: DGG$(A)_{12}$GGA, $\underline{SPM}$: GGLGGQGAG$(A)_6$GGAGQGGYGGLGSQGAGRGGQG$(A)_6$GGAGQG, and $\underline{SCM}$: GGAGGGYGGDGG$(A)_{12}$GGAGDGYGAG after TFA treatment (α and β denote α-helix and β-sheet, respectively).**

Table 3. **^{13}C CP/MAS chemical shifts (in ppm) of model peptides synthesized by the solid phase method and silk-like protein analogs produced from *E. coli* after trifluoroacetic acid (TFA) and formic acid (FA) treatment.**

Carbons	TFA treatment							FA treatment						Random coil	α-helix	β-sheet (silk II)
	$\underline{A_6}$	$\underline{A_9}$	$\underline{A_{12}}$	SPM	SCM	$(\underline{A_6}SCS)_8$	$(\underline{A_{12}}SCS)_4$	$(\underline{A_{12}}SGS)_4$	$\underline{A_6}$	$\underline{A_9}$	$\underline{A_{12}}$	SPM	$(\underline{A_6}SCS)_8$	$(\underline{A_{12}}SCS)_4$	$(\underline{A_{12}}SGS)_4$			$(\underline{AG})_{15}$
Ala Cβ	15.5	15.7	15.7	15.7	15.3	16.0	15.4	15.5	16.8				16.4	16.3	16.5	16.6	15.7	16.7
	19.4	20.0				19.8	18.5	18.9	19.6	20.9	20.2	20.6	19.6	20.2	20.2			19.6
									22.7	22.6	22.7		22.8					22.2
AlaCα	48.0	48.6				48.9	49.0	49.3	48.7	48.9	48.6	48.9	49.1	49.7	48.8	50.0		48.7
	52.5	52.5	52.5	52.1	52.3												52.5	
Ala C=O	171.3					172.0	172.6	172.7	171.9	172.1	171.8	171.9	172.3	172.4	172.1	175.5		
		176.2	176.1	176.5	175.9												176.5	171.8
Gly Cα	41.4					42.5	42.3	42.6	42.3	42.3	42.2	43.2	42.6	42.6	42.3	42.7		42.4
		44.0	43.3	43.0	43.1												44.0	
Gly C=O	168.8					172.0			168.8	168.8	168.6		–	–	–	171.3		169.1
		172.5	172.5	171.8	172.3		172.6	172.7									172.3	

$\underline{A_6}$: DGG$(A)_6$GGA. $\underline{A_9}$: DGG$(A)_9$GGA. $\underline{A_{12}}$: DGG$(A)_{12}$GGA. $\underline{SPM}$: GGLGGQGAG$(A)_6$GGAGQGGYGGLGSQGAGRGGQG$(A)_6$GGAGQG. $\underline{SCM}$: GGAGGGYGGDGG$(A)_{12}$GGAGDGYGAG. $(\underline{A_6}SCS)_8$: [DGG$(A)_6$GGAASGAGYGA(GAGSGA)$_2$GAGYGA]$_8$. $(\underline{A_{12}}SCS)_4$: [DGG$(A)_{12}$GGAASGAGYGA(GAGSGA)$_2$GAGYGA]$_4$. $(\underline{A_{12}}SGS)_4$: [DGG$(A)_{12}$GGAASGAGYGA(GGA)$_4$GAGYGA]$_4$.

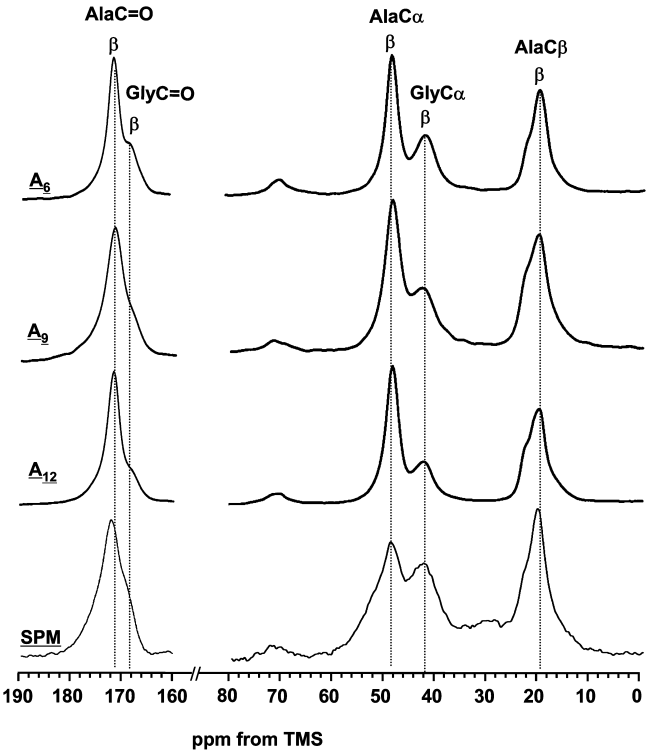

Fig. 5. **^{13}C CP/MAS NMR spectra of model peptides A$_6$: DGG(A)$_6$GGA, A$_9$: DGG(A)$_9$GGA, A$_{12}$: DGG(A)$_{12}$GGA, and SPM: GGLGGQGAG(A)$_6$GGAGQGGYGGLGSQGAGRGGQG(A)$_6$GGAGQG after FA treatment (α denotes β-sheet).**

and C-terminus. The targeted [His]$_6$ tagged protein was obtained after purification. The yields of these three silk-like protein analogs, *i.e.*, (A$_6$SCS)$_8$, (A$_{12}$SCS)$_4$, and (A$_{12}$SGS)$_4$, were 40.8 mg, 20.5 mg and 28.6 mg, respectively, after removal of the His-tag through cleavage. The purity of the cleaved protein was indicated by the presence of a single band on SDS-PAGE. The molecular weights of the cleaved proteins, (A$_6$SCS)$_8$, (A$_{12}$SCS)$_4$, and (A$_{12}$SGS)$_4$, were estimated from the SDS-PAGE bands to be about 36.0 kDa, 17.0 kDa, and 17.5 kDa, respectively, which agree with the theoretically predicted molecular weights of about 34.0 kDa, 16.5 kDa, and 16.5kDa.

Structural Characteristics of Polyalanine Model Peptides after TFA Treatment—Figure 4 shows the ^{13}C CP/MAS spectra of model peptides A$_6$, A$_9$ and A$_{12}$ after TFA treatment. The ^{13}C chemical shifts of the main peaks of Ala and Gly residues are summarized in Table 3, as well as the chemical shifts of typical conformations. Structural assignment was performed taking into account conformation-dependent chemical shifts for all carbons of Ala and Gly residues. In the spectrum of A$_6$, there are two peaks in the Ala C$_\beta$ region at 15.5 ppm and 19.4 ppm together with ones of an Ala carbonyl carbon at 171.3 ppm, and Ala Cα at 48.0 ppm and 52.5 ppm, indicating that A$_6$ adopts a predominantly β-sheet conformation with a significant amount of α-helix. In contrast, in the spectrum of A$_9$, a main peak at about 15.7 ppm with a small peak at about 20.0 ppm was observed in the Ala C$_\beta$ region. A single peak at 15.7 ppm was observed for the spectrum of A$_{12}$. Thus, with an increase in the length of the polyalanine repeat from 6 through 9 to 12, the α-helix conformation becomes dominant. A similar tendency was observed for the Ala carbonyl and Ala Cα carbons; the peaks at 176.1 ppm and 52.5 ppm, respectively, become dominant with an increase in the length of polyalanine. The dominance of the α-helix was also indicated by the appearance of a shift at 172.5 ppm for the Gly carbonyl carbon.

The ^{13}C CP/MAS NMR spectra of the longer sequential model peptide, SPM, derived from the sequence of spider dragline silk, and the longer peptide, SCM, derived from the sequence of *S. c.ricini* silk fibroin, are also shown in Fig. 4. A sharp peak of the Ala C$_\beta$ carbon at 15.7 ppm was observed in the spectrum of SPM. In this resonance region, the contribution of the Leu carbon peak to the peak at 15.7 ppm is considered, but the content is too low and thus can be neglected. This indicates an increase in the α-helix content compared with in the case of A$_6$. There is a glycine-rich region between the two (Ala)$_6$ sequences in SPM as well as the N-terminal and C-terminal regions, and such a glycine-rich region contributes greatly to the change in the conformation of polyalanine region (Ala)$_6$. When the length of polyalanine becomes longer, such as in the case of (Ala)$_{12}$, the conformation does not change *i.e.*, the α-helix conformation, even if the glycine-rich region around (Ala)$_{12}$ changes greatly between A$_{12}$ and SCM. More detailed analysis should be performed for selective ^{13}C-labeled peptides.

Structural Characteristics of Polyalanine Model Peptides after FA Treatment—The structures of the polyalanine model peptides treated with FA were examined by means of ^{13}C solid state NMR (Fig. 5). The chemical shifts of the Ala carbonyl carbon, Ala C$_\alpha$ and Ala C$_\beta$ resonated at about 171.8 ppm, 48.6 ppm and 20.2 ppm, respectively, which are characteristics of a typical β-sheet structure, suggesting that the three peptides, A$_6$, A$_9$ and A$_{12}$, all take on a β-sheet conformation, despite their different polyalanine lengths. Thus, the structural α-helix to β-sheet transition occurs clearly on changing the solvent from TFA to FA. Furthermore, the Ala C$_\beta$ region of these polyalanine regions is broad and asymmetric. This indicates that the structures of these model peptides as a model of *S. c.ricini* silk fibroin and spider dragline silk fibroins after spinning are heterogeneous as well as in the case of silk fibers from *B. mori*. The ^{13}C CP/MAS NMR spectrum of SPM after FA treatment indicated that SPM takes on a β-sheet structure. Thus, even if there is a Gly-rich region between the two (Ala)$_6$ sequences, the structure of (Ala)$_6$ only becomes β-sheet.

^{13}C CP/MAS NMR Spectra of Silk-Like Protein Analogs Treated with TFA or FA—The effect of introduction of the crystalline region of *B. mori* instead of the Gly-rich region into polyalainine regions on the propensity of β-sheet formation was investigated. TFA and FA treatment was also performed to generate the structure of silk-like proteins before and after spinning. Figure 6 shows the ^{13}C CP/MAS NMR spectra of silk-like analogs, (A$_6$SCS)$_8$, (A$_{12}$SCS)$_4$, and (A$_{12}$SGS)$_4$ after TFA and FA treatment. The carbonyl carbon regions of these three silk-like proteins after either TFA or FA treatment are broad. The intense peak at about 172.0 ppm can be assigned to the β-sheet structure of Ala residues, with the possibility of a

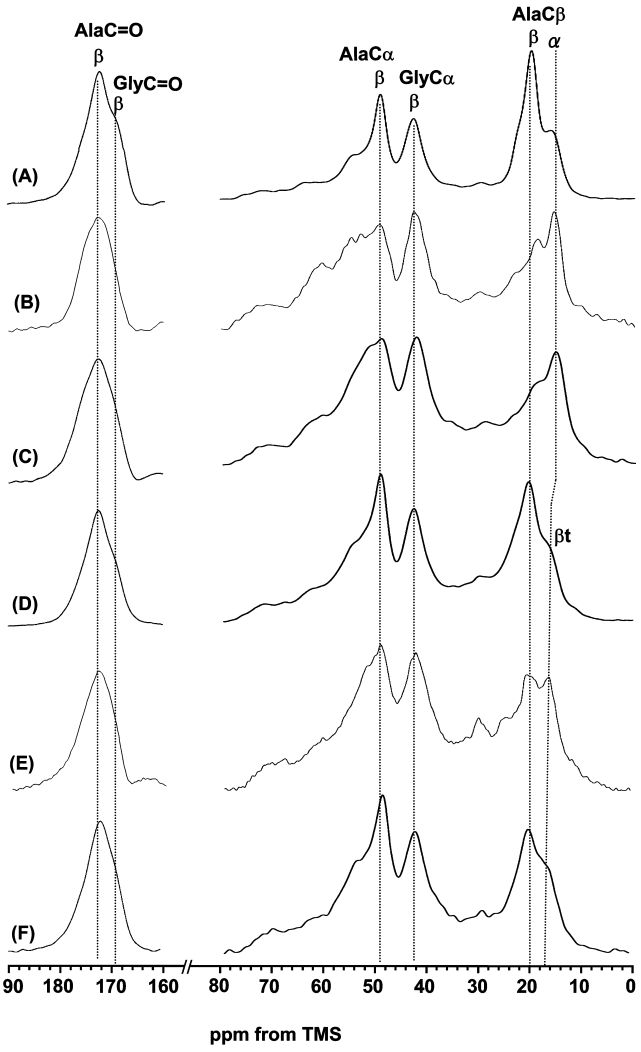

Fig. 6. **^{13}C CP/MAS NMR spectra of silk-like protein analogs (A) ($\underline{A_6SCS}$)$_8$, (B) ($\underline{A_{12}SCS}$)$_4$, and (C) ($\underline{A_{12}SGS}$)$_4$ after TFA treatment, and (D) ($\underline{A_6SCS}$)$_8$, (E) ($\underline{A_{12}SCS}$)$_4$, and (F) ($\underline{A_{12}SGS}$)$_4$ after FA treatment (*α*, β and βt denote α-helix, β-sheet and distorted β-turn).**

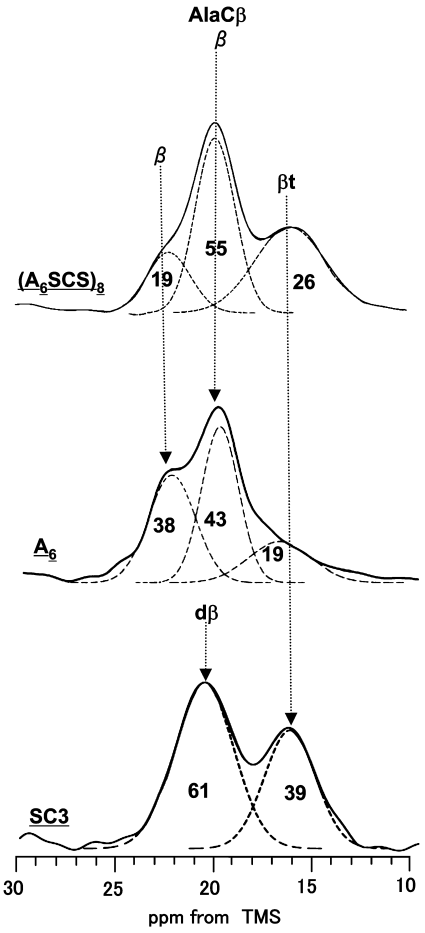

Fig. 7. **The decomposition of Ala C$_\beta$ peaks of ($\underline{A_6SCS}$)$_8$ and model peptides, $\underline{A_6}$: DGG(A)$_6$GGA and $\underline{CS3}$: (AGSGAG)$_2$(AGYGAG)(YGAGAG)(AGYGAG) (*17*) assuming Gaussian lineshapes (β, βt and dβ denote β-sheet, distorted β-turn and distorted β-sheet, respectively).**

random coil conformation of Gly residues. Therefore, it is difficult to assign the structures of these silk protein analogs clearly judging only from the chemical shift of the carbonyl carbons. However, the Ala C$_\beta$ spectrum provides a powerful tool for structural analysis because this region avoids overlapping of resonances from the other residues. Therefore, Ala C$_\beta$ was used to perform structural analysis of both its chemical shift and intensity. ($\underline{A_{12}SCS}$)$_4$ and ($\underline{A_{12}SGS}$)$_4$ after TFA treatment adopt an α-helical conformation, as judged from the intense peaks of Ala C$_\beta$ at 15.4ppm and 15.5 ppm, respectively, which are near the typical value of an α-helix, 15.7 ppm. However, the predominant Ala C$_\beta$ peak at 19.8 ppm suggests that ($\underline{A_6SCS}$)$_8$ after TFA treatment mainly adopts a β-sheet conformation similar to that of peptide $\underline{A_6}$, suggesting a high preference for β-sheet formation. Next, the structures of these silk-like analogs after FA treatment will be determined. FA treatment of the three silk-like protein analogs induces a predominantly β-sheet conformation, as judged from the Ala C$_\beta$ peak at about 20.2 ppm. The Ala C$_\beta$ peak of ($\underline{A_6SCS}$)$_8$ appears to be sharper than those of ($\underline{A_{12}SCS}$)$_4$ and ($\underline{A_{12}SGS}$)$_4$, implying that ($\underline{A_6SCS}$)$_8$ is packed more firmly into a β-sheet. Therefore, the addition of the crystalline region of *B. mori* silk fibroin (GAGSGA)$_2$ to the polyalanine crystalline region of spider silk GDGG(A)$_6$GGAG induced a strong propensity for β-sheet formation in silk-like protein analog ($\underline{A_6SCS}$)$_8$.

Detailed analysis of the Ala Cβ spectrum of ($\underline{A_6SCS}$)$_8$ Treated with FA—More detailed analysis of the structure of ($\underline{A_6SCS}$)$_8$ was involved based on expansion and decomposition of the Ala Cβ spectrum. Figure 7 shows the decomposition of the Ala Cβ spectra of ($\underline{A_6SCS}$)$_8$, and model peptides $\underline{A_6}$ and $\underline{CS3}$ (Table 1) after FA treatment assuming a Gaussian lineshape. The primary structure of silk-like protein analogs ($\underline{A_6SCS}$)$_8$ is [DGG(A)$_6$GGA-ASGAGYGA(GAGSGA)$_2$GAGYGA]$_8$. The underlined sequence containing the semi-crystalline and crystalline regions of *B. mori* silk fibroin in [DGG(A)$_6$GGAAS-$\underline{GAGYGA(GAGSGA)_2GAGYGA}$]$_8$ is very similar to that in the model peptide $\underline{CS3}$: (AGSGAG)$_2$(AGYGAG)(YGA-GAG)(AGYGAG). The detailed structure of latter peptide was determined by decomposing the Ala C$_\beta$ peak previ-

ously (*18*). The spectrum of CS3 indicates that there are two deconvoluted peaks due to the Ala C_β region in this peptide: that at 21.3 ppm (61%) is assigned to the distorted β-sheet, and the other at 16.3 ppm (39%) to distorted β-turns. In addition, the Ala C_β region of $DGG(A)_6GGAG$ was also deconvoluted. It contains two different β-sheet structures at 22.7 ppm (38%), and 19.6 ppm (43%), and a distorted β-turn with a broad peak at 16.8 ppm (19%). The Ala C_β region of $(A_6SCS)_8$ was deconvoluted by three components: 22.8 ppm (19%), 19.6 ppm (55%), and 16.4 ppm (26%). Taking the deconvoluted spectra of A_6 and CS3 as a reference, the peaks at 19.6 ppm and 22.8 ppm assigned to the β-sheet structure, the peak at 16.4 ppm is assigned to the distorted β-turn. The distorted β-sheet structure in peptide CS3 disappeared in the $(A_6SCS)_8$ due to the absence of a peak at 21.3 ppm. This is probably because introduction of peptide A_6 improved the arrangement of the β-sheet in SC3. On the other hand, compared with the deconvoluted Ala C_β region of $DGG(A)_6GGAG$, the content of β-sheet (22.8 ppm) in $(A_6SCS)_8$ decreased from 38% to 19%, however, the content of β-sheet at 19.6 ppm increased from 43% to 55%. In addition, the content of distorted β-turns increased to 26% in $(A_6SCS)_8$. The decrease in β-sheet (at 22.8 ppm) coupled with the increase in distorted β-turn in $(A_6SCS)_8$ is probably because Tyr can destabilize the β-sheet structure, and consequently the flexibility was improved compared with the relatively stiff rods produced on the combination of crystalline blocks $DGG(A)_6GGA$ and GAGSGA.

Fiber Formation of Silk-Like Protein Analogs—In order to determine the potential of these three silk-like protein analogs, it is required to spin these protein analogs into fibers. Several studies have focused on the spinning of native silks and silk-like protein into fibers in vitro (*21–23*). From these studies, it is clear that the choice of solvent for the production of the dope solution and coagulant bath is important as it can affect the properties and morphology of the fibers. Therefore, we sought to optimize the conditions for extruding fibers comprising our silk-like protein analogs. Table 2 lists the choices of solvent and coagulant bath for spinning the silk-like protein analog $(A_6SCS)_8$. Solvents including FA, HFA and dichloroacetic acid (DCA) were tried to prepare the silk dope solution, however, the best results were obtained by extruding the silk solution prepared with a DCA into coagulant methanol. These filaments could be post-drawn in the methanol. Spinning of silk-like protein analogs $(A_{12}SCS)_4$ and $(A_{12}SGS)_4$ was performed by extruding a silk solution prepared with DCA into coagulant methanol. The fibers produced from silk-like protein analogs were birefringent but brittle on drying. The fibers were deemed to be too fragile for tensile testing. We suggest that the scarcity of hydrophilic residues in our silk analogs may contribute to the fragility of the fibers because high hydrophobicity results in rapid and complete coagulation of the fibers, precluding the possibility of improving the orientation through post-drawing. Research is underway regarding the improvement of fiber quality by increasing the hydrophilicity of silk protein analogs or mixing native silk proteins or with synthetic peptides.

SUMMARY

The preference for β-sheet formation was investigated for two families of model peptides, one containing polyalanine- and Gly-rich regions of *N. clavipes* spider dragline or *S. c.ricini* silk, and the other with changing the length of the polyalanine region. It was found that a short polyalanine length, $(Ala)_6$, has a high preference for β-sheet formation, implying high strength of the spider silk, however, it is easily interrupted by a glycine-rich region. In an attempt to develop new silk-like materials of high strength, three silk-like protein analogs, $(A_6SCS)_8$, $(A_{12}SCS)_4$, and $(A_{12}SGS)_4$, were produced by adding the crystalline region of *B. mori*, GAGSGA, in the polyalanine region, $(Ala)_6$, or $(Ala)_{12}$. Depending on the solvent treatment, $(Ala)_{12}$ in $(A_{12}SCS)_4$ and $(A_{12}SGS)_4$ undergo the structural transition between α-helix and β-sheet seen in the native silks. However, $(A_6SCS)_8$ prepared from TFA or FA forms a β-sheet structure due to the combination of the shorter Ala length of polyalanine, $(Ala)_6$, and the sequence GAGSGA, indicating the high preference for β-sheet formation. Thus, silk-like protein analog $(A_6SCS)_8$ was deduced to be of high strength due to the combination of the crystalline regions of spider and *B. mori* silks. $(A_6SCS)_8$ exhibits potential for the formation of useful materials as it can adopt the β-sheet conformation independent of the prior solvent treatment. Further research is underway regarding the improvement of spinning technology to evaluate the potential of these and other silk-like proteins. The design and production of silk protein analogs coupled with NMR structural determination provides a potentially powerful tool for studying the relationship between sequence and properties, and the development of potentially useful materials.

TA acknowledges the support of the Insect Technology Project, Japan, and the Agriculture Biotechnology Project, Japan. The authors also thank Dr. David Knight at Oxford University for the useful discussions.

REFERENCES

1. 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
2. Zhou, C.Z., Confalonieri, F., Medina, N., Zivanovic, Y., Esnault, C., Yang, T., Jacquet, M., Janin, J., Duguet, M., Perasso, R., and Li, Z.G. (2000) Fine organization of *Bombyx mori* fibroin heavy chain gene. *Nucleic Acids Res.* **28**, 2413–2419
3. Xu, M. and Lewis, R.V. (1990) Structure of a protein superfiber: spider dragline silk. *Proc. Natl Acad. Sci.USA* **87**, 7120–7124
4. Hinman, M.B. and Lewis, R.V. (1992) Isolation of a clone encoding a second dragline silk fibroin. *J. Biol. Chem.* **267**, 19320–19324
5. Parkhe, A.D., Seeley, S.K., Gardner, K., Thompson, L., and Lewis, R.V. (1997) Structural studies of spider silk proteins in the fiber. *J. Mol. Recognit.* **10**, 1–6
6. Simmons, A., Ray, E., and Jelinski, L.W. (1994) Solid-state ^{13}C NMR of *Nephila clavipes* dragline silk establishes structure and identity of crystalline regions. *Macromolecules* **27**, 5235–5237
7. Simmons, A.H., Michal, C.A., and Jelinski, L.W. (1996) Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science* **271**, 84–87
8. Grubb, D.T. and Jelinski, L.W. (1997) Fiber morphology of spider silk: the effects of tensile deformation. *Macromolecules* **30**, 2860–2867

9. Gosline, J.M., Denny, M.W., and DeMont, M.E. (1984) Spider silk as a rubber. *Nature* **309**, 551–552
10. Kümmerlen, J., van Beek, J.D., Vollrath, F., and Meier, B.H. (1996) Local structure in spider dragline silk investigated by two-dimensional spin-diffusion nuclear magnetic resonance. *Macromolecules* **29**, 2920–2928
11. Ashida, J., Ohgo, K., Komatsu, K., Kubota, A., and Asakura, T. (2003) Determination of the torsion angles of alanine and glycine residues of model peptide compounds of spider silk (AGG)10 using solid-state NMR methods. *J. Biomol. NMR*. **25**, 91–103
12. Nakazawa, Y., Nakai, T., Kameda, T., and Asakura, T. (1999) A ^{13}C NMR study on the structural change of silk fibroin from *Samia cynthia ricin*. *Chem. Phys. Lett.* **311**, 362–366
13. Nakazawa, Y. and Asakura, T. (2003) Structure determination of a peptide model of the repeated helical domain in *Samia cynthia ricini* silk fibroin before spinning by a combination of advanced solid-state NMR methods. *J. Amer. Chem. Soc.* **125**, 7230–7237
14. Nakazawa, Y., Bamba, M., Nishio, S., and Asakura, T. (2003) Tightly winding structure of sequential model peptide for repeated helical region in *Samia cynthia ricini* silk fibroin studied with solid-stated NMR. *Protein Sci.* **12**, 666–671
15. Asakura, T., Yao, J., Yamane, T., Umemra, K., and Ulrich, A.S. (2002) Heterogeneous structure of silk fibers from *Bombyx mori* resolved by ^{13}C solid-state NMR. *J. Amer. Chem. Soc.* **124**, 8794–8795
16. Asakura, T. and Yao, J. (2002) ^{13}C CP/MAS NMR study on structural heterogeneity in *Bombyx mori* silk fiber and their generation by stretching. *Protein Sci.* **11**, 2706–2713
17. Yao, J., Ohgo, K., Sugino, R., Kishore, R., and Asakura, T. (2004) Structural analysis of *Bombyx* mori fibroin peptides with formic acid treatment using high-resolution solid-state 13C NMR spectrocscopy. *Biomacromolecules* **5**, 1763–1769
18. Sambrook, J., Frisch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY
19. Prince, J.T, McGrath, K.P, DiGirolamo, C.M., and Kaplan, D.L. (1995) Construction, cloning, and expression of synthetic genes encoding spider dragline silk. *Biochemistry* **34**, 10879–10885
20. Fukushima, Y. (1998) Genetically engineered syntheses of tandem repetitive polypeptides consisting of glycine-rich sequence of spider dragline silk. *Biopolymers* **45**, 269–279
21. Seidel, A., Liivak, O., and Jelinske, L.W. (1998) Artificial spinning of spider silk. *Macromolecules* **31**, 7633–6736
22. Seidel, A., Liivak, O., Calve, S., Adaska, J., Ji, G., Yang, Z., Grubb, D., Zax, D.B., and Jelinski, L.W. (2000) Regenerated spider silk: Processing, properties, and structure. *Macromolecules* **33**, 775–780
23. Arcidiacono, S., Mello, C.M., Butler, M., Welsh, E., Soares, J.W., Allen, A., Ziegler, D., Laue, T., and Chase, S. (2002) Aqueous processing and fiber spinning of recombinant spider silk. *Macromolecules* **35**, 1262–1266