Cloning, Expression, and Assembly of Sericin-like Protein*

Received for publication, July 18, 2003, and in revised form, September 2, 2003 Published, JBC Papers in Press, September 8, 2003 DOI 10.1074/jbc.M307792200

Jia Huang, Regina Valluzzi, Elisabetta Bini, Brian Vernaglia, and David L. Kaplan‡
From the Departments of Biomedical Engineering, Chemical and Biological Engineering, Bioengineering Center, Tufts University, Medford, Massachusetts 02155

Recombinant sericin proteins of different molecular masses (17.4, 31.9, and 46.5 kDa), based on the 38-amino acid repetitive motif of native sericin, were cloned, expressed, and purified. The recombinant sericin self-assembled during dialysis (starting concentration of 2.5 mg/ml) forming twisted fibers. Circular dichroism and Fourier transform infrared spectroscopy studies demonstrated protein conformational transitions occurred from random coil to β-sheets during the dialysis. Congo red-stained recombinant sericin fibrils exhibited apple-green birefringence, indicating long-range order in the red-stained recombinant sericin fibrils exhibited apple-the most frequent pattern of amyloidogenic proteins, acids showed that the hydrophobicity pattern resembles 323, 453–461) definition of polar and non-polar amino Gutfreund, Y., and Gregoret, L. M. (2002) explore questions related to model in terms of hydrophobicity pattern with which to sericins from the silk gland of Bombyx mori suggest random coil with some interactions. Analysis of recombinant sericin sequence using Mandel-Gutfreund’s (Mandel-Gutfreund, Y., and Gregoret, L. M. (2002) J. Mol. Biol. 323, 453–461) definition of polar and non-polar amino acids showed that the hydrophobicity pattern resembles the most frequent pattern of amyloidogenic proteins, polar amino acid aggregates (PPPPP). Many β-proteins and peptides are designed to study amyloidogenesis using a polar/non-polar alternating pattern (PNPNN). Sericin-like proteins or peptides provide an alternative model in terms of hydrophobicity pattern with which to explore questions related to β-sheet formation and amyloidogenesis. The glue-like property of sericin is attributed to the hydrogen bonding between serine residues of sericin with serine residues in the fibroin structural components of silk fiber.

Silk of Bombyx mori is composed of two kinds of proteins: the fibroin which is synthesized in the posterior silk gland and the sericins produced by the middle silk gland. The sericins bind two fibroin fibers together in the anterior region of the gland as the fibers are spun to form the cocoon. Sericins represent a family of proteins whose molecular mass ranges from 20 to 310 kDa (2, 3) and are characterized by an unusually high serine content, 38.1 mol % (4). Sericins are soluble in hot water, which makes it possible to degum the silk fibroin threads. Two genes encode sericins, Ser1 and Ser2 (5–8). The different molecular weight sericins are the products of different splicing events at the transcript level. Some conformation studies with sericins from the silk gland of B. mori or regenerated sericin from B. mori cocoons suggested random coil with some β structure (9, 10). However, the samples used in these studies were mixtures of the various native sericin proteins, therefore the structure of the individual proteins is unknown and the contribution of specific sequences toward secondary structure is not confirmed. Sericin model peptides such as poly(-serine) (11–16), poly(O-benzyl-l-serine) (12), and poly(3-acetyl-l-serine) (13, 14) were synthesized for structural feature characterization. The ability of poly(serine) to fold into a β conformation depended on molecular weight. Low molecular weight poly(serine) (molecular weight 650) was soluble and adopted a random coil conformation in aqueous solution, whereas high molecular weight poly(serine) (molecular weight 105,000) formed β-sheets and was insoluble in water (11). However, homopolypeptides like poly(serine) do not match native sericin in terms of amino acid composition, sequence, and molecular weight. Therefore, to better understand sericin structure, interactions between sericin and fibroin, and the biological relevance of sericin in fiber structure, high molecular weight pure sericin-like proteins are required.

Sericin 1 encoded by Ser1 consists of 70 repeats of the 38-amino acid motif (5): SSTGGSSNTDSNSNGSTSTGGSGSTNGSSNSRDGSY. The molecular mass of sericin 1 proteins is 76–284 kDa. The 38-amino acid repeat motif exhibits a high content of hydrophilic amino acids: 47.7% serine, 13.5% threonine, and 11.6% tyrosine, very close to the average amino acid composition of sericin (4): 38.1% serine, 7.5% threonine, and 4.7% tyrosine. Hence, a sericin-like protein can be synthesized based on this 38-amino acid motif. Proteins derived from genetic engineering are characterized by control over composition, sequence, molecular weight, and stereochemical purity. Motifs of silkworm and spider silk proteins have been studied in a similar manner to understand structures in different environments, and spider silk proteins have been modified for the purpose of controlled assembly (17–24). The aim of this study was to biosynthesize sericin-like proteins based on the 38-amino acid motif of sericin 1 using synthetic gene construction methods. Biophysical characterization of the purified recombinant sericin was accomplished using scanning electron microscopy, circular dichroism (CD),1 Fourier transform infrared (FTIR) spectroscopy, and Congo red (CR) staining to gain insight into the structure and assembly of this protein.

EXPERIMENTAL PROCEDURES

Materials—Oligonucleotides coding for the sercin 38-mer repeat from sercin 1 were purchased from MWG-Biotech (High Point, NC). All restriction enzymes, calf intestinal alkaline phosphatase, T4 polynucleotide kinase, T4 DNA ligase, DNA ladders, and pUC19 were purchased from New England Biolab (NEB, Beverly, MA). Qiagen (Valencia, CA) plasmid miniprep kits were used to purify plasmid from transformed

1 The abbreviations used are: CD, circular dichroism; FTIR, Fourier transform infrared; CR, Congo red; Ni-NTA, nickel-nitritotratic acid; MALDI-TOF, matrix-assisted laser desorption/ionization-time of flight; LDS, lithium dodecyl sulfate; LDS-PAGE, lithium dodecyl sulfate-polyacrylamide gel electrophoresis.
cells. Qiagen MinElute extraction kits and MinElute gel extraction kits were used to purify DNA fragment from agarose gels. Qiagen MinElute reaction cleanup kits were used to purify ligated oligonucleotide and DNA fragment from agarose gels. Qiagen MinElute extraction kits and MinElute gel extraction kits were used to purify DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse M13 primers. The obtained clone was named pUC19-SRC1.

**Construction of Synthetic B. mori Sericin Motif DNA**—Two sense and two antisense oligonucleotides of about 70 base (Fig. 1) were synthesized on a 0.05 μmol scale. After phosphorylation the complementary oligonucleotides were annealed by heating at 95 °C for 5 min and cooling slowly to room temperature. The two sets of annealed oligonucleotides were ligated to form one oligonucleotide monomer building block that was 132 nucleotides long encoding the sericin 38 amino acid motif. The ligated oligonucleotide was purified by Qiagen MinElute reaction cleanup kit. The ligation reaction was transformed into 100 μL MAX efficiency Stbl2™ competent cells. The correct clones of pUC19-SRC2 were determined by DNA sequencing with forward and reverse M13 primers. DNA inserts with more SRC1 repeats were constructed in the same strategy and termed pUC19-SRC1 and SRC1 monomer was 1:5 for ligating SRC1 into pUC19-SRC1 to generate pUC19-SRC2. Five microliters of the ligation reaction was transformed into 100 μL MAX efficiency Stbl2™ competent cells. The correct clones of pUC19-SRC2 were determined by DNA sequencing with forward and reverse M13 primers. DNA inserts with more SRC1 repeats were constructed in the same strategy and termed pUC19-SRCn, where n indicates the number of monomeric repeats. The longest insert contained twenty tandem repeats of SRC1 monomer.

**Modification of Expression Plasmid pET21a(+)—Two sense and anti-sense oligonucleotides of 36 base (Fig. 2) were synthesized on a 0.01 μmol scale for the construction of pET21a(+)-adaptor. Oligonucleotides were phosphorylated and purified, then ligated into pET21a(+) expression plasmid digested with BamHI and XhoI using the same method as described for gene construction. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Only pET21a(+) plasmid that could be digested with AgeI was chosen for DNA sequencing with T7 forward and reverse primers. Six histidines were introduced into pET21a(+) at the C terminus for the purpose of purification of the target protein by specific interaction with Ni-NTA resin.

**Construction of Multimeric B. mori Sericin Motif DNA**—The recombinant pUC19-SRC1 was digested with XhoI and SalI to release the 114 bp SRC monomer. pUC19-SRC1 was digested and dephosphorylated before comparing its concentration with that of the purified SRC monomer DNA fragment by 1.2% gel electrophoresis. The molar ratio of SRCn repeats was controlled in the range of 1:3–5. Five microliters of the ligation reaction was transformed into 100 μL MAX efficiency Stbl2™ competent cells. The correct clones of recombinant extension plasmid pET21a(+)–SRCn were determined by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.

**Protein Expression**—Recombinant expression plasmid pET21a(+)–SRCn was prepared from Stbl2™ cells was transformed into *Escherichia coli* strain DH5α. Hybridization oligonucleotides were synthesized on a 0.05 mol scale for the construction of pET21a(H11001)–SRCn. Oligonucleotides were phosphorylated and then ligated into pUC19. Five microliters of the ligation reaction was transformed into 100 μL Library efficiency DH5α™ competent cells. Recombinant plasmid was identified by DNA sequencing with forward and reverse T7 primers. The amino acid sequence of the sericin-like protein cloned in the pET21a(+)–SRCn plasmid is shown in Fig. 3.
was negatively charged at high pH and the charge repulsion slowed down the protein aggregation.

**LDS-PAGE and Western Blots**—LDS-PAGE was performed using NuPAGE 4–12% Bis-Tris gels (Invitrogen), and the protein bands were visualized with colloidal blue staining (Invitrogen). Western blotting to nitrocellulose was performed using XCell SureLock™ Mini-Cell and XCell II™ blot module with a NuPAGE transfer buffer, pH 7.2. The T7-tag fusion proteins were detected on blots using a T7-tag alkaline phosphatase conjugate. Standards containing the T7-s peptide tag were used with molecular masses of 15, 25, 35, 75, and 150 kDa and were detected by S-protein alkaline phosphatase conjugate. Target bands were visualized with chemiluminescent DCP-Star™ substrate. Membranes were developed with TBST (10 mM Tris-HCL, pH 8.0, 150 mM NaCl, 0.1% Tween 20) and 1% gelatin and then incubated with S-protein alkaline phosphatase conjugate according to the manufacturer's recommendations (Novagen).

**Amino Acid Composition Analysis**—Amino acid analysis was carried out at the Yale University W. M. Keck Biotechnology Resource Laboratory on a Beckman Model 7300 ion-exchange instrument following a 16 h hydrolysis at 115 °C in 100 μl of 6 N HCl and 0.2% phenol that also contained 2 nmol norleucine. The latter serves as an internal standard to correct for losses that may occur during sample transfers and drying. After hydrolysis, the HCl was dried in a Speedvac and the resulting amino acids dissolved in 100 μl Beckman sample buffer that contained 2 nmol homoserine with the latter acting as a second internal standard to independently monitor transfer of the sample onto the analyser.

**MALDI-TOF**—A Bruker Proflex™ mass spectrometer (Bruker, Billerica, MA) was used for molecular weight determinations. Recombinant protein concentration was 1 μg/μl, and the matrix a-cyano-4-hydroxycinnamic acid concentration was 10 g/ml in 50:50 water/acetonitrile solution. The protein and matrix were mixed at 1:1 ratio. Protein calibration standards including trypsinogen, protein A, and bovine albumin were used to calibrate mass spectrometer and the covered mass range is from 20 to 70 kDa.

**Circular Dichroism Spectroscopy**—Purified SRC samples (starting concentration of 2.5 mg/ml) were dialyzed against 1 mM phosphate buffer, pH 8, using Slide-A-Lyzer dialysis cassettes (molecular weight cut-off = 3,500). Samples were removed at various time points and centrifuged at 14,000 × g for 10 min. The supernatant was analyzed by CD at 20 °C in a JASCO J-710 spectropolarimeter (Easton, MD) under a nitrogen atmosphere. Denatured purification buffer (pH 4.5) of the same volume as the SRC sample was also dialyzed against 1 mM phosphate buffer, pH 7.4, and was removed at various time points and analyzed by CD as background after filtering with a 0.22 μm filter. A quartz cell with a path length of 0.1 cm was used and the spectra were recorded between 260 and 190 nm at 1 nm intervals, with a response time 8 s, scan speed 10 nm/min, bandwidth 1 nm, and accumulation of 3 scans. The protein concentration was analyzed by bichinonic assay against a bovine serum albumin standard, and the experimental data were expressed as mean residue molar ellipticities [θ].

**Fourier Transform Infrared Spectroscopy**—FTIR studies were performed using a Bruker Equinox 55 FTIR spectrometer with an FTIR microscope attachment. Pellets were removed from SRC samples taken at various time points as described under "Circular Dichroism Spectroscopy." The supernatant was dried on ZnS crystals and examined with the FTIR microscope in the transmission mode. Hydrophilic ZnS crystal surface can preserve the solution-state conformation of protein (23). Both second derivatization and Fourier self-deconvolution were performed to determine the peaks in amide I band and ensure the fidelity of each method. The Gaussian × Lorentzian function was applied to curve fit data peaks using Peak Fit v4 (Jandel). The contribution of each component band to the amide I band was evaluated by dividing the total area of amide I band with area of each component band. For comparison, FTIR spectra of regenerated sericin was also taken after the drop of sericin solution dried. Regenerated silk sericin was prepared after treating cocoons of B. mori silkworm with boiling water for 60 min and filtering using 0.2 μm polyvinylidene difluoride filter (Millipore).

**Scanning Electron Microscopy**—Ten μl of a fibril sample formed after dialysis against distilled water, pH 6.3, for 10 h was taken from dialysis cassettes and loaded onto a silicon chip. After adsorption to the surface for 5 min, the sample was rinsed with deionized distilled water and air-dried. Fibril morphology was characterized with a LEO 982 Field Emission Scanning Electron Microscope (LEO Electron Microscopy, Inc., Thornwood, NY) at 1.0 kV.

**Congo Red Staining and Polarized Light Microscopy**—One hundred μl SRC protein fibril suspensions formed by dialysis against distilled water, pH 6.4, and 500 μl CR solution (100 μM in TBS solution, pH 7.4) were incubated at room temperature overnight. After centrifuge at 14,000 rpm for 5 min, the pellet was resuspended with 1 ml TBS solution, pH 7.4, and centrifuged again to pellet the sample. Unbound Congo red was removed by repeating the process twice. One-hundred μl TBS, pH 7.4, was added to the fibril pellet and 20 μl sample was pipetted onto a glass slide. The slide was then dried for a few hours. Birefringence was determined with a Nikon Eclipse E600 Polarized microscope.

**RESULTS**

**Construction of Recombinant Expression Plasmid pET21α(+)**

**SRCn**—Fig. 4 shows the results of electrophoresis analysis of SRC inserts of various length in the modified pET21α(+) ex-
pression plasmid and SRC monomeric insert in pUC 19. pET21a(+)SRCn was digested with NheI (restriction site in pET21a(+) plasmid) and AgeI (restriction site introduced with the adaptor) instead of BamHI because the inserts of SRCn can self-ligate before inserting into pET21a(+) to form pET21a(+)-(SRCn)n. Double digestion with NheI and AgeI separates multimerized SRCn from single SRCn constructs.

Protein Expression and Purification—Sericin-like proteins were expressed in E. coli RY 3041 strain, a mutant of BL21(DE3). The C terminus of SlyD protein expressed by RY3041 was deleted so that SlyD protein cannot coelute with the target protein by binding to the Ni-NTA resin (25). The longest SRC insert contained 20 SRC1 tandem repeats (totally 1368 bp), but the largest SRC expressed, SRC12 (46.5 kDa) was much shorter than SRC20. The protein concentration of purified SRCn decreased with the length of the target protein, which may be because of the high content of serine placing excessive demand on Ser tRNA so that no intact protein can be expressed if the DNA encoding for the target protein is too long. Because the His-tag is designed at the C-terminal of the protein, no truncated target protein will be detected. The same issue has been observed previously in expression of spider silk genes (17, 22). Overall yield of purified SRCn was 15–36 mg/l depending on the molecular weight of SRCn.

The apparent molecular weights of the SRC proteins as estimated by LDS-PAGE were double the molecular weight of the SRC proteins (Fig. 5a). Reduced electrophoretic mobility has been observed previously for various proteins (27–32), and this can stem from unusual amino acid composition, intrinsic net charge or protein oligomerization. The pl values of SRC proteins are 5.98, 5.02, and 4.61 for SRC4, 8, and 12, respectively. Therefore, these acidic proteins bind the anionic surfactant lithium dodecyl sulfate at working pH 8.5 weakly due to electrostatic repulsion. The constant charge/mass ratio was lower than normal resulting in reduced electrophoretic mobility. Western blot analysis (Fig. 5b) also detected the SRC proteins based on the T7-tag.

Amino Acid Composition Analysis—Amino acid composition analysis confirmed the correct composition for the purified SRCn (Table I). SRC8 was chosen for further structural study because of the combination of high molecular weight and high expression yield compared with SRC4 and SRC12. The purity of the SRC8 protein sample for further structural study was 93% based on amino acid composition analysis. Analysis of SRC8 by MALDI-TOF showed a single peak of molecular mass of 32219 Da, close to the theoretical mass of 31947 Da.

Circular Dichroism—CD spectra of SRC8 dialyzed for 1 h showed a minimum at 196 nm, suggesting random coil structure (Fig. 6). However, there was no random coil maximum at 212 nm, the \( [\theta] \) value around 220 is below zero, suggesting there were other structures coexisting with random coil, such as \( \beta \)-sheet and \( \beta \)-turn (33, 34), which exhibit negative bands near 216 and between 220 and 230 nm, respectively. Therefore, the secondary structure of SRC8 dialyzed for 1 h appears to be the combination of random coil, \( \beta \)-turn, and \( \beta \)-sheet. The CD spectra changed as a function of time, indicating a gradual transition upon dialysis. After dialysis for 3.5 h, CD spectra showed the maximum at 196 nm, the minimum at 218 nm, diagnostic of \( \beta \)-sheet secondary structure.

Fourier Transform Infrared Spectroscopy—FTIR spectra of SRC8 dialyzed for 1 h is similar to regenerated native sericin, which has broad peaks around 1658 cm\(^{-1}\) (amide I), and 1538 cm\(^{-1}\) (amide II), indicating random coil and \( \beta \) turn conformations (35, 36) (Fig. 7). After dialysis for 3 h, the FTIR spectra
exhibit amide I and II vibrational modes at 1621, 1699 and 1522 cm\(^{-1}\), respectively, characteristic of \(\alpha\)-sheet conformation. The weak amide I component at 1697 cm\(^{-1}\) is the large splitting of the amide I mode caused by regular reversal of chain direction typical of the antiparallel \(\alpha\)-sheet architecture (37).

Quantification analysis using peak fit of FTIR spectra (Table II) revealed that after dialysis for 1 h, the SRC8 secondary structure was composed of 23% random coil, 41% \(\beta\)-turns and \(\beta\)-bends, 36% \(\alpha\)-sheet. \(\beta\)-turns and \(\beta\)-bends structure increased to 51% after dialyzed for 1.5 h with the significant loss of random coil structure to only 9% and gain of \(\alpha\)-sheet to 40%. No random coil structure was detected after dialysis for 3 h.

**DISCUSSION**

Almost all the SRC8 protein self-assembled into fibrillar structure after dialysis longer than 10 h and was removed from supernatant by centrifugation. No significant CD signal could be detected for SRC8 samples dialyzed longer than 10 h. When dialyzed against distilled water, fibrillation of SRC8 was faster at pH 6.3 than at pH 11. The pI of SRC8 is about 6, and fibrillation progressed faster when SRC8 had no net charge. An excess of uncompensated charges could hinder \(\beta\)-sheet forma-
tion by high-charge repulsion (41). SRC8 was totally unfolded by 8 M urea in purification buffer, so it was random coil at the beginning of dialysis. With the removal of urea, hydrogen bonding initiated the formation of β-turns and β-sheets. Most of the amino acids in SRC protein are polar (serine, asparagines, threonine, tyrosine, aspartic acid, and arginine), which are capable of linking β-strands together into sheets by hydrogen bonding between polar side chains or between main chain amid and polar side chain (42–44). This kind of interaction is called “polar zipper” and is responsible for inherited neurodegenerative diseases (45, 46) and is frequently found in the transmembrane regions of membrane proteins in which it helps to form extensive hydrogen bond networks between transmembrane helices (47). One known preference of globular protein sequences that can form β-strands is that they have polar/non-polar amino acid alternating binary patterns (48–51). According to the definition of polar and non-polar amino acids proposed by Mandel-Gutfreund (1), polar amino acids (P) include Thr, Thr, Asn, Gln, Tyr, Cys, Lys, Arg, His, Asp, and Glu, and non-polar amino acids (N) include Gly, Ala, Val, Leu, Ile, Met, Pro, Phe, and Trp. The binary pattern of the 38 amino acid sericin motif in the SRC proteins is PPNPPPNNPPPPPP-PNNNNPNNPPNNPPPPNPNN. Sericin mainly consists of polar amino acid domains composed of 3–15 polar amino acids connected by 1–2 non-polar amino acids. The polar/non-polar binary pattern of SRC proteins is similar to that of amyloidogenic proteins in which the occurrence of PPPPP and NNNNN pattern are ranked first and fourth (1). Amyloidogenic proteins are well known to self-assemble into amyloid fibrils from a partially unfolded state, which are pathological aggregates of protein in Alzheimer’s and prion diseases (52). The diameter of the fibers formed by SRC8 was larger than that of amyloid fibrils, which are always 6–13 nm in diameter independent of the amino acid sequence. The reason for this difference is unclear, but these findings suggest that the sericin-like proteins provide a useful model to improve our understanding of mechanisms that influence fibril diameter in amyloid-related structures. Circular dichroism characterization of regenerated native sericin extracted from silkworm cocoons showed similar conformational transitions to that of recombinant sercin proteins (53). However, gels formed from native sericins can be transformed with heat back into a solution state characterized by random coil, whereas the SRC fibrils described in the present study are stable and the conformation is not reversible upon heating. This difference of reversibility of conformational transitions between regenerated sericin and recombinant sericin-like protein likely originates from the amino acid sequence. Regenerated native sericin contains domains of less repetitive amino acid sequence (7), such as ETGEEELFDWYSKIEKGKPVPHMKVIP, that can form more random hydrogen bonding patterns so that water molecules are retained in the polymer network after gelation. The SRC fibrils behave more like amyloids, which are composed of strongly intermolecular bonded, dehydrated, and densely packed β-sheets (54). Because of its high propensity to form β-sheets, the domains of sericin sequence that resemble this 38-amino acid motif should contribute to the β conformation content of native sericin structure previously studied by CD (9, 10).

The amino acids that exhibit highest overall potential to form turns are glycine, asparagines, aspartic acid, serine, and proline. All of them except proline are present in the 38-amino acid repetitive motif of sericin (55). The β-sheet structure was likely propagated from β-turns because the decrease in percentage of β-turns was accompanied by the increase of β-sheet content (56). The FTIR spectrum of SRC8 dialyzed for 3.5 h was close to that of the sericin model compounds poly(β-L-phenylalanine) and poly(β-serine) (12) and sericin 1 peptide (35) in the solid state, although their sequences are different. The FTIR spectra gave consistent results and further confirmed the conformational transition of SRC8 from random coil to β-sheet during fibrillation. Curve fitting of freshly prepared regenerated native sericin from the FTIR spectrum showed that it has a similar secondary structure to that of SRC8 dialyzed for 1 h (data not shown).

Congo red is a symmetrical sulfonated azodye that preferentially stains amyloid fibrils by intercalation between the anti-parallel β-pleated sheets (57–60). Congo red-stained amyloid fibrils show an apple-green birefringence under polarized light (61–63), indicating the bound CR molecules are aligned with respect to each other due to the organized cross β-sheet structure of amyloid fibrils. Accordingly, the apple-green birefringence of CR-stained SRC fibrils suggests that the long-range arrangement of the β-sheets necessary for birefringence exists in SRC8 fibrils. Congo red does not bind well to poly(β-serine) (40), which suggest β-sheets formed by poly(β-serine) may form amorphous aggregates. The scanning electron microscopy and CR staining results suggest that fibers of biosynthetic sericin are hierarchical assembled and include protofibrils. Biosynthetic sericin fibrils exhibit birefringence, a diagnostic characteristic of amloid fibrils. However, the diameters of the two fibrils discussed above remains an open question.

Silk fibroin of B. mori contains 46% glycine, 29% alanine, 12% serine, 4.8% tyrosine, and 2.5% valine (64). About 75% of all serine residues occur within the sequence of the repetitive region (65), which forms crystalline domains. It was also found that ~75% of the serine residues in fibroin interact with carboxyl groups on adjacent chains to form intermolecular hydrogen-bonding networks in the fibroin fiber (66). The remaining serine residues in B. mori fibroin and serine residues in sericin likely contribute to intermolecular hydrogen bonding related to sericin coating the two fibroin fibers. Therefore, biosynthetic sericin is a useful model protein to study sericin and fibroin interactions during the course of silk fiber extrusion through silkworm spinneret.

In conclusion, during dialysis biosynthetic sericin self-assemble into helically twisted fibrillar structures. CD and FTIR results suggested that the conformation of the sericin transformed from random coil to β-sheet, and the final fibril structure predominantly comprised of long-range ordered antiparallel β-sheets that gave rise to apple-green birefringence after Congo red binding. Self-assembly is a key for peptide and protein assembly into fibrils (67), sheets (68) or membranes (69), tubes (70), or three-dimensional matrices. Most self-assembled β-peptides and proteins exhibit alternating polar/non-polar binary patterns and have been well studied (48–51). Biosynthetic sericin sequences are composed of a high content of polar uncharged amino acids that polar/non-polar hydrophobicity patterns similar to amyloidogenic proteins that can form ordered β-sheets. The study of biosynthetic sericins can provide improved understanding of protein self-assembly and the general mechanisms of fibril formation.

REFERENCES
1. Mandel-Gutfreund, Y., and Gregoret, L. M. (2002) J. Mol. Biol. 323, 453–461
2. Kaplan, D., Adams, W. W., Farmer, B., and Viney, C. (1994) Silk Polymers: Materials Science and Biotechnology, pp. 10–11, American Chemical Society, Washington, D.C.
3. Gama, T. I., Nokuchi, T., and Lauffer, H. (1977) Insect Biochemistry 7, 285–295
4. Lucas, P., and Rudall, K. M. (1968) Comprehensive Biochemistry, Vol. 26B, 475–588, Elsevier Science Publishers, Amsterdam
5. Okamoto, H., Inokuchi, K., and Suzuki, Y. (1982) J. Biol. Chem. 257, 15192–15199
6. Michaille, J. J., Garel, A., and Prudhomme, J. C. (1980) Gene 86, 177–184
7. Garel, A., Deleage, G., and Prudhomme, J. C. (1997) Insect Biochem. Mol. Biol. 27, 469–477
8. Couble, P., Michaille, J. J., Garel, A., Couble, M. L., and Prudhomme, J. C.
