Construction, Cloning, and Expression of Synthetic Genes Encoding Spider Dragline Silk

John T. Prince,[‡,§] Kevin P. McGrath,[*,‡] Carla M. DiGirolamo,[‡,§] and David L. Kaplan[‡]

Biotechnology Division, U.S. Army Natick Research, Development, & Engineering Center, Natick, Massachusetts 01760-5020, and Geo-Centers, Inc., Newton Centre, Massachusetts 02159

Received December 22, 1994; Revised Manuscript Received March 1, 1995[⊗]

ABSTRACT: Synthetic genes encoding recombinant spider silk proteins have been constructed, cloned, and expressed. Protein sequences were derived from *Nephila clavipes* dragline silk proteins and reverse-translated to the corresponding DNA sequences. Codon selection was chosen to maximize expression levels in *Escherichia coli*. DNA "monomer" sequences were multimerized to encode high molecular weight synthetic spider silks using a "head-to-tail" construction strategy. Multimers were cloned into a prokaryotic expression vector and the encoded silk proteins were expressed in *E. coli* upon induction with IPTG.[1] Four multimers, ranging in size from 14.7 to 41.3 kDa, were chosen for detailed analysis. These proteins were isolated by immobilized metal affinity chromatography and purified using reverse-phase HPLC. The composition and identity of the purified proteins were confirmed by amino acid composition analysis, N-terminal sequencing, laser desorption mass spectroscopy, and Western analysis using antibodies reactive to native spider dragline silk. Circular dichroism measurements indicate that the synthetic spider silks have substantial β-sheet structure.

Silks are secreted into sets of specialized glands by insects and spiders and then processed into a fiber outside the glands. The role of these fibers is usually to protect an insect during metamorphosis, or in the case of spiders, to act as a net to catch prey. Spider silks are particularly remarkable materials, combining strength and toughness to absorb energy from flying insects without breaking. The dragline silk from the gold orb weaver, *Nephila clavipes*, displays excellent impact resistance and a balance of stiffness, strength and extensibility (Gosline *et al.*, 1984, 1986; Cunniff *et al.*, 1994). In general, spider silks are smooth, lustrous materials that show interesting UV reflectance patterns (Craig & Bernard, 1989) and solution behavior (Kerkam *et al.*, 1991). Spider silk fibers are insoluble in dilute acids and dilute alkali, resistant to most proteolytic enzymes (Tillinghast & Kavanagh, 1977; Tillinghast, 1984; Townley & Tillinghast, 1988), hydrolyzed by concentated sulfuric acid (Livengood, 1990) and solubilized by 9 M lithium bromide without hydrolysis (Mello *et al.*, 1994).

Spider silks have been expressed with cultured glands (Candelas & Cintron, 1981; Candelas & Lopez, 1983), and limited attempts to express native spider silk genes have been reported (Kaplan *et al.*, 1992a). Previous attempts to clone and express *Bombyx mori* silkworm silk genes have met with limited success. Oshima and Suzuki (1977) attempted expression of *B. mori* fibroin clones in *Escherichia coli* but found deletions in the repetitive domains. Even silk proteins and their genes isolated from native populations of *B. mori* exhibit size variations, which were attributed to recombination events (Sprague *et al.*, 1979). Because of difficulties encountered in cloning native silk genes, synthetic genes encoding *B. mori* silk-like sequences have been cloned and expressed in *E. coli*. A six amino acid repeat [GAGAGS] controls crystallinity in films and fibers (Cappello *et al.*, 1990; Ferrari *et al.*, 1993) and repeats such as [$(AG)_3$PEG] (McGrath *et al.*, 1992) and [$(AG)_3$EG] (Krejchi *et al.*, 1994) have been used in efforts to control crystalline order in protein-based materials.

Native silk proteins from the silkworm *B. mori* and spiders such as *N. clavipes* contain highly repetitive crystalline domains periodically interrupted by less crystalline or amorphous regions of sequence (Kaplan *et al.*, 1992a; 1992b). The crystalline region from *B. mori* fibroin is a 59 amino acid repeat [GAGAGSGAAG[SGAGAG]$_8$Y], with a 3:2:1 ratio of glycine, alanine, and serine (Stydom *et al.*, 1977; Tsujimoto & Suzuki, 1979); it is estimated to occur approximately 70 times within the protein. Repeated motifs from spider dragline silk are less conserved, but a 13 amino acid repeat [YGGLGSQGAGRGG] has been identified both in spider silk cDNA (Xu & Lewis, 1990) and by proteolysis of the native protein (Mello *et al.*, 1994).

Partial cDNA sequences of gold orb weaver spider *N. clavipes* silk proteins have been reported (Xu & Lewis, 1990, Hinman & Lewis, 1992). Two distinct cDNAs have been identified, encoding proteins designated spidroin I and spidroin II. Although these cDNA sequences are incomplete, much of the encoded proteins is repetitive in nature. For spidroin I, the consensus repeat is [GGAGQGGYGGLGSQGAGRGGLGGQGAG], usually followed by a poly[Ala] sequence containing 4−7 alanines. For spidroin II the consensus repeat is [GPGGYGPGQQGPGGYAPGQQPSGPGS], again followed by short poly[Ala] stretches but with considerably less sequence conservation between repeats.

* To whom correspondence should be addressed.
‡ U.S. Army Natick Research, Development, & Engineering Center.
§ Geo-Centers, Inc.
⊗ Abstract published in *Advance ACS Abstracts,* May 1, 1995.
[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid, disodium salt, IPTG, isopropyl β-thiogalactoside; NTA, nitrilotriacetic acid; PAGE, polyacrylamide gel electrophoresis; Tris-HCl, tris(hydroxymethyl)aminomethane, hydrochloric acid salt; YT, yeast−tryptone broth; nt, nucleotide.

0006-2960/95/0434-10879$09.00/0

(A) DNA sequence of synthetic adapter inserted into pUC18 to create pUC-LINK

```
   XbaI    BamHI         NheI      BstEII                  SpeI     BamHI    XbaI
5' CT AGA GGA TCC ATG GCT AGC GGT GAC CTG AAT AAC ACT AGT GGA TCC T-3'
3'      T CCT AGG TAC CGA TCG CCA CTG GAC TTA TTG TGA TCA CCT AGG AGA TC-5'
           G   S   M   A   S   G   R   L   N   N   T   S   G   S
```

(B) SPI

```
5'-CT AGC GGT AGA GGC GGG CTG GGT GGC CAG GGT GCA GGT GCG GCT GCG GCT
     3'-G CCA TCT CCG CCC GAC CCA CCG GTC CCA CGT CCA CGC CGA CGC CGA
       S   G   R   G   G   L   G   G   Q   G   A   G   A   A   A   A

   GCC GCG GCA GCG GCC GCA GGC GGT GCC GGC CAA GGT GGC TAT GGC GGC CTG
   CGG CGC CGT CGC CGG CGT CCG CCA CGG CCG GTT CCA CCG ATA CCG CCG GAC
    A   A   A   A   A   A   G   G   A   G   Q   G   G   Y   G   G   L

   GGT TCT CAG GGG A-3'
   CCA AGA GTC CCC TGA TC-5'
    G   S   Q   G   T
```

(C) [SPII]

```
5' CT AGC GGT CCG GGC GGT TAT GGT CCG GGT CAA CAA A-3'
3'      G CCA GGC CGG CCA ATA CCA GGC CCA GTT GTT TGA TC-5'
       S   G   P   G   G   Y   G   P   G   Q   Q   T
```

FIGURE 1: DNA sequences of synthetic spider silk monomers and pUC-LINK adapter.

These cDNA sequences share similarities in overall organization and in regions of sequence conservation. In both proteins, the consensus repeats are rich in glycine and glutamine, with poly[Ala] regions integrated into larger repeating units. In addition, the C-terminal portions of the two proteins are reported to be very similar (Hinman & Lewis, 1992, Beckwitt & Arcidiacono, 1994).

We report here the successful construction, cloning, and expression of synthetic spider silk genes. For these constructs, the repetitive units were patterned after sequences identified from native dragline silk proteins. The sequences have been reverse-translated to DNA sequences whose codon selection was designed to maximize expression levels in *E. coli*. These DNA fragments were used as "building blocks" for the construction of larger multimers of synthetic spider silk. The recombinant proteins produced from these constructs will permit a systematic study of the structure–function relationships in spider silks, and in protein-based materials in general.

MATERIALS AND METHODS

Materials. The procedures involved in the manipulation and modification of DNA were performed as described by Sambrook *et al.* (1989). In all ligations, the insert-to-vector ratios were between 2:1 and 5:1. *E. coli* strains NM522 and SG13009pREP4 were obtained from Stratagene (La Jolla, CA) and Qiagen (Chatsworth, CA), respectively. All enzymes, sequencing kits, and DNA molecular weight standards were obtained from Pharmacia LKB Biotechnology (Piscataway, NJ) or New England Biolabs (Beverly, MA). pQE-9 and Ni-NTA agarose resin were obtained from Qiagen (Chatsworth, CA). Equilibrated phenol solutions were obtained from Ameresco (Solon, OH). Bio101 RPM kits were obtained from American BioAnalytical (Natick, MA). Ultrafree-MC tubes were obtained from Millipore (Bedford, MA).

Construction of pUC-LINK. The sequence shown in Figure 1A was synthesized as two fragments which when annealed have *Xba*I-compatible ends. Oligonucleotides were synthesized using the cyanoethyl phosphoramidite chemistry of McBride and Carothers (1983) and purified by reverse-phase HPLC. Purified oligonucleotides (10 μg, ca. 400 pmol) were phosphorylated at their 5′ termini using 50 units of T4 polynucleotide kinase. Complementary oligonucleotides were mixed in equimolar ratios and heated to 80 °C, and then slowly cooled to anneal and form duplexes. The DNA duplexes were extracted with phenol/chloroform (1:1) and precipitated by sodium acetate and 2-propanol. Recovered DNA was resuspended in 100 μL of 10 mM Tris-HCl (pH 8.0) containing 1 mM EDTA. After phosphorylation and annealing of complementary linker strands, the duplex DNA was ligated into pUC18 that had been digested with *Xba*I and dephosphorylated using calf intestinal phosphatase.

Construction of Synthetic Spider Silk DNA Fragments. The [SPI] sequence shown in Figure 1B was synthesized as four fragments, two 65 nt in length and two 49 nt in length. The sequences contained 5′-*Nhe*I and 3′-*Spe*I termini. The [SPII] sequence shown in Figure 1C was synthesized as two fragments, each 42 nt in length. Purified oligonucleotides (10 μg, ca. 400 pmol) were phosphorylated at their 5′ termini using 50 units of T4 polynucleotide kinase. Complementary oligonucleotides were mixed in equimolar ratios and heated to 80 °C, and then slowly cooled to anneal and form duplexes. The DNA duplexes were extracted with phenol/chloroform (1:1) and precipitated by sodium acetate and 2-propanol. Recovered DNA was resuspended to 100 μL of 10 mM Tris-HCl (pH 8.0) containing 1 mM EDTA.

Construction of Synthetic Silk Multimers. Multimerization of synthetic [SPI] was accomplished by one of two methods. In the first method, dephosphorylated pUC-LINK (150 ng, digested with both *Nhe*I and *Spe*I and purified from 1%

Table 1: List of Genetic Constructs

construct designation	combination strategy	expression system	expressed in *E. coli*?	purified/characterized?
$[SPI]_1$	none	pQE-9	no	ND[b]
$[SPI]_2$	1 into 1	pQE-9	analytical	ND[b]
$[SPI]_3$	1 × 3	pQE-9	analytical	yes/Western
$[SPI]_4$	1 × 4	pQE-9	1 L	fully analyzed
$[SPI]_5$	3 into 2	pQE-9	analytical	yes/Western
$[SPI]_6$	3 × 2	pQE-9	analytical	yes/Western
$[SPI]_7$	5 into 2	pQE-9	30 L	fully analyzed
$[SPI]_8$	4 into 4	pQE-9	analytical	no/Western
$[SPI]_9$	3 into 6	pQE-9	no	ND[b]
$[SPI]_{10}$	3 into 7	pQE-9	1 L	fully analyzed
$[SPI]_{11}$	4 into 7	pQE-9	analytical	yes/Western
$[SPI]_{12}$	10 into 2	pQE-9	analytical	no/Western
$[SPI]_{13}$	3 into 10	pQE-9	1 L	fully analyzed
$[SPI]_{14}$	4 into 10	pQE-9	analytical	no/Western[a]
$[SPI]_{17}$	7 into 10	pQE-9	analytical	ND[b]
$[SPI]_{18}$	8 into 10	pQE-9	no	ND[b]
$[[SPI]_3-[SPII]_1]_4$	(3 into 1) × 4	pQE-9/pET-21a	analytical	yes/Western
$[[SPI]_3-[SPII]_1]_4$	(3 into 1) × 4	pQE-9/pET-21a	analytical	yes/Western
$[[SPI]_5-[SPII]_1]_2$	(5 into 1) × 2	pQE-9/pET-21a	analytical	yes/Western

[a] Abnormal migration seen in Western—migrated at lower apparent molecular weight (ca. 12-mer). [b] Not determined.

agarose) was combined with ca. 100 ng of the [SPI] DNA duplex in ligation buffer. To this solution was added 1 μL of T4 DNA ligase; the reaction was mixed and incubated at 4 °C. This reaction was used to transform *E. coli* NM522 cells made competent made the method of Hanahan (1983). Transformed cells were identified by plating and incubation on medium containing 400 μg/mL ampicillin. Transformants containing the correctly constructed DNA fragments were identified by restriction digest analysis by *Bam*HI and by double digestion with *Nhe*I and *Spe*I. The $[SPI]_1$ synthetic insert was verified by double-stranded plasmid sequencing. NM522 containing the recombinant plasmids was grown in 5 mL of 4× YT containing 400 μg/mL ampicillin, and the plasmids were isolated by the method of Birnboim and Doly (1979). Synthetic inserts were liberated from the recombinant plasmids by digestion with *Nhe*I and *Spe*I and purified from 1% agarose. These inserts were self-ligated to form linear multimers of the target DNA. These were then redigested with *Nhe*I and *Spe*I, which cuts all inverted repeats. The digested multimers were recovered by 2-propanol precipitation and religated back into dephosphorylated pUC-LINK. Because of the diverse population of multimers, no insert to vector ratios are reported.

The second method of constructing higher-order multimers was to digest plasmids containing the monomer or multimers with *Nhe*I, generating a linearized vector-containing insert. A second insert-containing plasmid was digested with *Nhe*I and *Spe*I, liberating the multimerized DNA. Linearized plasmids and inserts were purified (using an Ultrafree-MC cartridge and the method recommended by the manufacturer) and ligated in different combinations to provide higher order multimers. Multimers constructed by this method are listed in Table 1, along with the strategy used in each construction. Multimerized DNA fragments encoding the synthetic spider silks were removed from the recombinant pUC-LINK plasmids by digestion with *Bam*HI, gel purified from 1% agarose, and inserted into *Bam*HI-digested and dephosphorylated pQE-9. The ligated mixture was used to transform the expression host strain SG13009pREP4. Identification of recombinant expression plasmids was accomplished by restriction analysis using the appropriate enzymes.

Protein Expression and Purification. For analytical protein expressions, cultures were grown to ca. $OD_{600} = 0.5-1.0$ in 4× YT medium containing 400 μg/mL ampicillin. Expression of the synthetic spider silks was induced by the addition of IPTG to a final concentration of 1 mM. After 1−2 h, 1 mL of cells was harvested and lysed in 0.2 mL of loading buffer. Expressions were analyzed by PAGE using 10% or 10−20% polyacrylamide gels. For large-scale expressions, cultures were grown to ca. $OD_{600} = 1.5-2.0$ in 1 L of 4× YT medium containing 400 μg/mL ampicillin, and protein expression was induced by the addition of IPTG to a final concentration of 1 mM. After 1−4 h the cells were harvested by centrifugation. Proteins were purified by affinity chromatography on a Ni-NTA column using the protocol recommended by the manufacturer. Purified proteins were exhaustively dialyzed against distilled deionized water and then lyophilized to provide a solid material of ca. 70% purity. Further purifications were carried out by reverse-phase HPLC using a C_{18} Vydac column with a linear gradient of water/acetonitrile containing 0.1% trifluoroacetic acid.

Western Assays. Western assays were performed using the Immuno-Blot Assay from Bio-Rad (Hercules, CA), using colorimetric detection with alkaline phosphatase. The proteins were electrophoretically transferred from the polyacrylamide gel to Immobilon-P^{SQ} paper (Millipore, Bedford, MA) using a semidry blotting apparatus (Integrated Separation Systems, Natick, MA). The primary antibody was a goat anti-spider silk antibody raised against washed spider silk and provided by Monsanto (St. Louis, MO). The secondary antibody was a swine anti-goat IgG(H+L) antibody conjugated to alkaline phosphatase (Boehringer Mannheim, Indianapolis, IN).

Characterization of Purified Proteins. Amino acid analyses were carried out on a Waters Pico-Tag system. Proteins were hydrolyzed in constant-boiling HCl containing 0.5% phenol, neutralized to remove residual acid, and derivatized using phenyl isothiocyanate (PITC). The PITC-amino acids were then separated and quantified using reverse-phase HPLC. N-Terminal sequencing was carried out on an Applied Biosystems Model 477A peptide sequencer. Laser

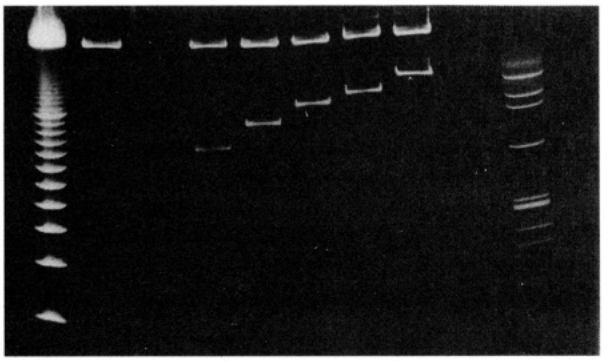

FIGURE 2: pUC-LINK[SPI] plasmids, *Bam*HI-digested and separated in a 4−16% polyacrylamide gel (Novex) by electrophoresis in a Tris−borate−EDTA (pH 8.0) buffer. Lanes: (1) 100-bp DNA ladder, (2) pUC-LINK[SPI]$_2$, (3) pUC-LINK[SPI]$_4$, (4) pUC-LINK-[SPI]$_6$, (5) pUC-LINK[SPI]$_8$, (6) pUC-LINK[SPI]$_{10}$, (7) pUC-LINK-[SPI]$_{12}$, (8) pUC-LINK[SPI]$_{14}$, (9) pUC-LINK[SPI]$_{17}$, and (10) X174 *Hae*III digest.

desorption mass spectroscopy was performed on a Finnegan LaserMat Mass Spectrometer using α-hydroxycinnamic acid as the desorption matrix.

Circular Dichroism. Circular dichroism spectra were recorded in a 10-mm path length cell at concentrations between 0.5 and 1 μM on an Aviv 60DS spectrophotometer. All spectra were recorded at 25 ± 0.1 °C. Protein solution concentrations were calculated by UV absorbance at 280 and 205 nm by the method of Scopes (1974). The instrument was routinely calibrated against *d*-camphorsulfonic acid. Secondary structure analysis was performed using algorithms provided by the manufacturer (Aviv Associates, Lakewood, NJ).

RESULTS

Construction of Monomeric and Multimeric [SPI] and [SPI]−[SPII] Hybrid DNAs. Plasmids containing correctly constructed tandem repeats of the synthetic DNA sequences were identified by restriction digest analysis using *Bam*H or by double digestion with *Nhe*I and *Spe*I. Multimers were formed by self-condensation of *Nhe*I−*Spe*I fragments or by insertion of *Nhe*I−*Spe*I fragments into *Nhe*I-digested vectors; these enzymes generate identical cohesive ends, and when the ligation joins a *Nhe*I and *Spe*I site together in a "head-to-tail" fashion (as in the self-condensation strategy), any internal restriction sites are destroyed. Subsequent digestion by *Nhe*I and *Spe*I will excise any inverted repeats within the multimerized fragments, ensuring correct maintenance of the protein reading frame. These multimers can also be inserted into either *Nhe*I- or *Spe*I-digested recombinant vectors to build, in a tightly controlled fashion, larger genes of any desired degree of multimerization. We have exploited these strategies for the controlled construction of multimers ranging in size from 1 to 18 repeats of the [SPI] sequence while retaining unique *Nhe*I and *Spe*I sites in the constructs. Table 1 lists the methods used in the construction of [SPI] multimers and [SPI]−[SPII] hybrids. Polyacrylamide gel electrophoresis of several of these constructs is shown in Figure 2.

Protein Expression. Our expression efforts utilized the commercially available expression vector pQE-9, which places the synthetic genes under the control of the bacteriophage T5 promoter and places a unique [His]$_6$ sequence at the N-terminus of the recombinant protein for purification by immobilized metal affinity chromatography. pQE9-[SPI] multimers were transformed into *E. coli* SG13009pREP4 and the encoded proteins were expressed upon induction by IPTG. Crude extracts were analyzed by polyacrylamide gel electrophoresis and Western analysis using an anti-spider silk antibody. Staining of the gels with Coomassie Brilliant Blue showed clear expression of [SPI]$_3$ (11.7 kDa), [SPI]$_4$ (14.7 kDa), and [SPI]$_5$ (17.6 kDa); [SPI]$_6$ (20.6 kDa) was faintly stained, and higher order multimers, which were readily detected by Western analysis, did not appear as visible bands (data not shown). A general trend observed was that as the size of the expressed protein increased the level of expression (measured as percent of total cell protein) decreased. This has been observed in the expression of other silk-like recombinant proteins (McGrath, 1991; McGrath *et al.*, 1992; J. Cappello, personal communication). The antibody used in the Western assays was not quantitative in its reaction with silk-like proteins but could reliably by used to detect synthetic spider silk proteins both in whole cell lysates and in purified form. In conjunction with the physical characterization of the purified materials (*vide infra*), the data support the expression of SPI multimers using the pQE9 expression system.

IPTG-induced extracts of the [SPI]−[SPII] hybrids were subjected to PAGE and stained by Coomassie Brilliant Blue (data not shown). Bands corresponding to the molecular weights of the expected proteins were not seen; however, Western analysis confirmed the expression of silk-like proteins in the extracts and in fractions partially purified by immobilized metal affinity chromatography (data not shown). In all cases, the recombinant protein migrated in PAGE at the correct apparent molecular weight, in contrast with other silk-like proteins expressed in *E. coli* (McGrath, 1991; McGrath *et al.*, 1992; Creel *et al.*, 1992; Cantor *et al.*, 1994; J. Cappello, personal communication).

Protein Analysis. Proteins expressed from the pQE9 vector can be purified on a Ni-NTA column through the [His]$_6$ sequence positioned at the N-terminus. While expressions and purifications from almost all of the constructs listed in Table 1 were accomplished, we focused our characterization efforts on four distinct size variants of the [SPI] multimer sequences. Multimers containing 4, 7, 10, and 13 repeats of the 38 amino acid [SPI] sequence were chosen for analysis. Typical yields for [SPI]$_4$ (14.7 kDa), [SPI]$_7$ (23.5 kDa), [SPI]$_{10}$ (32.4 kDa), and [SPI]$_{13}$ (41.3 kDa) were 15, 7, 3, and 2 mg of lyophilized powder/L of culture when purified by Ni-NTA affinity chromatography (see Figure 3). These proteins were further purified by reverse-phase HPLC and subjected to amino acid composition analysis, N-terminal sequencing, and laser desorption mass spectroscopy. The results of the N-terminal sequencing on several of the constructs and on CNBr-cleaved [SPI]$_4$ confirm that the synthetic spider silk is produced from the pQE9 plasmid system. The laser desorption mass spectroscopy results (see Table 2) provide strong evidence that the synthetic spider silk multimers are correctly constructed and expressed. Amino acid compositions reflect a high Gly + Ala content, but the percentage of alanine in these analyses is somewhat lower than expected; as described in the Discussion section, we believe that the hydrophobic poly[Ala] stretches are resistant to hydrolysis and are not represented in the total compositional analysis.

Table 2: Physical Characteristics of Purified Spider Silks

sample	LDMS[a] exp (obs)	amino acid composition (in %)				
		Gly exp (obs)	Ala exp (obs)	Ser exp (obs)	Glx[b] exp (obs)	Leu exp (obs)
$[SPI]_4$	14 659 (14655)	37.9 (38.7)	28.8 (22.9)	6.2 (8.3)	6.8 (6.6)	5.1 (4.8)
$[SPI]_7$	23 567 (23558)	39.5 (41.0)	29.9 (26.6)	5.8 (6.3)	7.2 (7.5)	5.2 (5.2)
$[SPI]_{10}$	32 539 (32558)	40.3 (40.3)	30.1 (23.7)	5.9 (8.1)	7.4 (6.9)	5.3 (4.7)
$[SPI]_{13}$	41 479 (ND)[c]	40.7 (40.5)	30.3 (28.6)	5.9 (6.3)	7.7 (7.6)	5.4 (5.4)

[a] Laser desorption mass spectrometry; molecular masses reported in amu. [b] Glx = Glu + Gln. [c] Could not be determined experimentally.

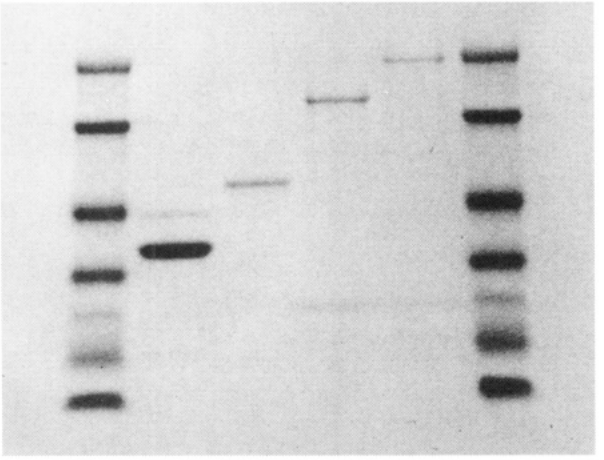

FIGURE 3: Coomassie blue staining of Ni-NTA- (Qiagen) purified extracts of IPTG-induced SG13009pREP4 containing pQE9[SPI] multimers. Lanes 1 and 6: Amersham rainbow marker standards (from the top) ovalbumin, 46 kDa; carbonic anhydrase, 30 kDa; trypsin inhibitor, 21.5 kDa; lysozyme, 14.3 kDa; aprotonin, 6.5 kDa; insulin B chain, 3.4 kDa; insulin A chain, 2.35 kDa. Lanes: (2) $[SPI]_4$, (3) $[SPI]_7$, (4) $[SPI]_{10}$, and (5) $[SPI]_{13}$.

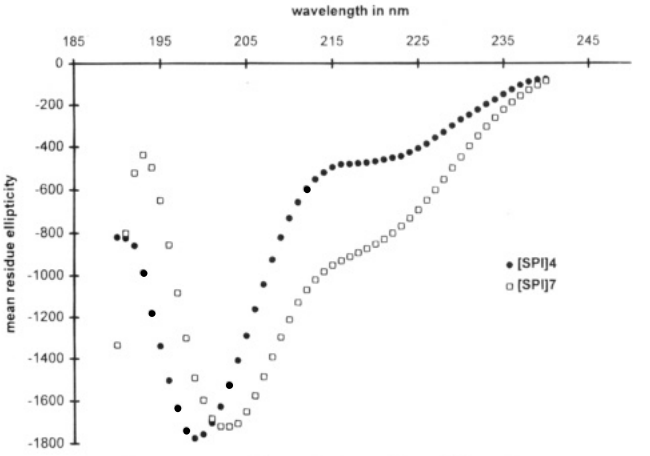

FIGURE 4: CD spectra of synthetic spider silks. Spectra were performed in distilled deionized water at protein concentrations of ~25 ng/mL.

Conformational Studies in Solution. Ni-NTA-purified $[SPI]_4$ and $[SPI]_7$ were used to perform a limited solubility study. Synthetic spider silk multimers were only slightly soluble in distilled deionized water but were at concentrations high enough to perform spectroscopic studies. Several attempts to increase the solubility of the multimers were tried; synthetic spider silk multimers did dissolve at higher concentrations in mixed aqueous−organic solvents, the most useful being water containing 0.1% trifluoroacetic acid or 0.1% HCl. Ni-NTA-purified $[SPI]_4$ and $[SPI]_7$ most likely contained some residual salts from the purification process that depressed protein solubilities.

Ni-NTA-purified $[SPI]_7$ was dissolved in water and its circular dichroism spectrum was measured from 240 to 190 nm. The result showed 66% β-sheet and 28% β-turn, in good agreement with the results of Raman spectroscopic analysis of the native silk in dragline fibers (Gillespie *et al.*, 1994). However, the low protein concentration (~25 ng/mL) precludes more definitive characterization. Ni-NTA-purified $[SPI]_4$ and $[SPI]_7$ were also dissolved in 0.1% trifluoroacetic acid and diluted to 70 and 83 ng/mL, respectively. The relative proportions of secondary structure were estimated at 55% β-sheet structure and 39% random coil for $[SPI]_4$ and 65% β-sheet structure and 27% random coil for $[SPI]_7$ (see Figure 4). In neither case was any appreciable amount of α-helix observed, as was suggested by earlier Fourier transform infrared analysis of dragline fibers (Hinman & Lewis, 1992).

DISCUSSION

Using *N. clavipes* dragline silks as a model, a series of synthetic spider silk proteins have been produced and characterized. Consensus protein repeats were defined by analysis of the published cDNA sequences and by our in-house efforts in peptide mapping of the *N. clavipes* major ampullate silk. These sequences were reverse-translated to their corresponding DNA sequences, with a codon selection designed to maximize expression levels in *E. coli*. The DNA sequences were multimerized using a "head-to-tail" construction strategy (McGrath, 1991; Ferrari *et al.*, 1993) in which two enzymes having identical cohesive ends are self-condensed to form large concatemers, which are subsequently digested by these same enzymes. Inverted repeats are still susceptible to enzymatic cleavage, while direct repeats are not. We used this approach to generate multimers of modest size (4−6 repeats) which were then inserted into plasmids containing other short multimers; in this fashion, we were able to generate all size variants from 1 to 14 repeats, as well as several constructs containing direct repeats as high as 2 kilobases in length.

The highly repetitive nature of these constructs raised concerns over the stability of the genes and the possibility of recombination. Previous in-house efforts at stable propagation of cDNAs from the major ampullate gland of *N. clavipes* in prokaryotic hosts revealed considerable recombination and loss of insert over time; this has also been reported from other laboratories (R. Lewis, personal communication). Other efforts at stabilizing silkworm cDNA sequences in *E. coli* (Oshima & Suzuki, 1977) have encountered similar results. No spider silk cDNAs larger than about 2.5 kilobases have been reported; this may represent an upper limit to the stability of such sequences in procaryotes. A number of published reports indicate that synthetic genes, which encode silk-like sequences virtually identical to those encoded by natural cDNAs, have both increased stability and increased expression levels (Ferrari *et al.*, 1993). In some cases, genes as large as 3.5 kilobases

of direct repeats have been successfully propagated and expressed in *E. coli*.

In general, there has been little evidence for instability in multimers of the [SPI] sequence. When *Bam*HI-digested pUC-[SPI] multimers are analyzed by gel electrophoresis there is a slight laddering effect for multimers containing 6 or more tandem repeats, with the highest rung being the strongest band. The distance between the rungs is approximately 100 bp, close in size to the 114 bp of the [SPI] monomer. The high copy number of pUC-based vectors, coupled with the high internal concentration of the [SPI] sequence in larger inserts, leads to a low level of recombinational events. We believe that this is most closely associated with the copy number of the plasmid vector, since the same inserts propagated in pBR322-based vectors show no such laddering. This evidence, in conjunction with the fact that the pUC-[SPI] multimers were grown in NM522 (a *rec*$^+$ strain), argues that these constructs are not susceptible to genetic rearrangements in *E. coli*.

Some of the constructs described here were designed to disrupt the regularity of the DNA sequences by inserting a smaller unit ([SPII], a sequence identified from spider silk cDNAs) into the repeated [SPI] sequences. This sequence, GPGGYGPGQQ, displays little homology to other silk sequences and may be associated with the amorphous or elastic regions in the silk fiber. The [SPII] sequence was not found in the cDNA encoding the [SPI] consensus repeat and has not yet been identified in peptide mapping of the dragline silk. However, a combination of these elements may help stabilize highly repetitive genes by the periodic interspersion of other sequences and allow for stable maintenance of genes encoding even higher molecular weight proteins. These hybrid materials will almost certainly have unique processing and mechanical behavior, and help to extend our understanding of structure−function relationships in protein solids.

These inserts were cloned into a bacterial expression vector and produced in sufficient quantity and purity to begin physical characterization. The recovery of these proteins from high cell density fermentations is uniformly low, averaging between 2 and 10 mg of purified product/L. The reason for such low production is not clear but may be due to the high G + C content of the synthetic genes. Efforts at maximizing yields through optimal codon usage may be offset by the formation of unusually stable regions of secondary structure in the mRNAs, as we have used the pQE-9 expression system to produce other synthetic proteins at levels 10−20-fold higher than seen in these systems.

Amino acid composition, N-terminal sequence analysis, laser desorption mass spectroscopy, and Western analysis using an anti-spider silk antibody confirm the identities of the expressed synthetic spider silk multimers. The only physical characterization data not fully supportive of [SPI] expression is the underrepresentation of alanine in the amino acid compositions of HPLC-purified materials. In the [SPI] repeat, all of the alanine residues are in a homopolymeric $[Ala]_{10}$ sequence. It is our hypothesis that this hydrophobic stretch of alanines is resistant to hydrolysis and therefore not fully represented in the total compositional analysis; similar results have been seen in compositional analyses of synthetic peptides containing poly[Ala] regions (T. Thannhauser, personal communication). All other amino acids are in the correct relative proportions, and the molecular masses of the proteins determined by laser desorption mass spectroscopy are correct to within the error limits of the experimental technique. Other evidence in support of this contention are the results from N-terminal sequence analysis and the positive response in Western analyses using an antibody reactive toward natural dragline silk proteins.

Circular dichroism measurements indicate that the synthetic spider silk multimer proteins have substantial β-sheet structure. These data are similar to circular dichroism measurements of *B. mori* silk fibroin, a protein with substantial β-sheet structure, when solubilized in water (Canetti *et al.*, 1989; S. Bedell, personal communication). However, both the Chou−Fasman and Robson−Garnier conformation prediction methods indicate that the synthetic spider silk multimers should have about 30% α-helical structure and very little β-sheet structure. While recent thermodynamic measurements of the β-sheet-forming propensities of amino acids in model protein systems (Kim & Berg, 1994; Minor & Kim, 1994; Smith *et al.*, 1994) indicate that the β-sheet-forming potential of amino acids predicted by these methods are generally correct, they are based in great measure on soluble, globular proteins; we have designed the synthetic spider silk proteins from well-characterized solid-state structures with high proportions of β-sheet structure. A variety of repeating poly(alanylglycine) polypeptides, including *B. mori* silk, form antiparallel β-sheet structures in the solid state, with repeating alanylglycine dyads incorporated into crystalline β-sheet regions. The general [Gly-Gly-X] motif present in both the native spider silk proteins and our recombinant proteins is not predicted to form a β-sheet (Lotz *et al.*, 1974); however, recent NMR studies on *N. clavipes* dragline silk containing labeled amino acids indicating that the poly[Ala] repeats present in the native spider dragline silk are housed within crystalline β-sheet domains in the spun fiber (Simmons *et al.*, 1994). The issue of whether or not the [Gly-Gly-X] motif is incorporated into a β-sheet has not been clearly resolved, but peptides containing these sequences have been shown to adopt a β-sheet conformation in solution upon heating (Hinman & Lewis, 1992).

The macroscopic mechanical properties of spider silk (Gosline *et al.*, 1984; Cunniff *et al.*, 1994) are derived from molecular-level interactions between protein chains both before and after spinning. Several studies have been carried out investigating these interactions at various stages of processing in silk glands and in the final fiber structure. Recently, Thiel *et al.* (1994) estimated a 50% volume fraction of crystalline material in dragline silk fibers, in good agreement with a Raman spectroscopic analysis of 56% β-sheet and 22% β-turn (Gillespie *et al.*, 1994). Early X-ray diffraction experiments indicated the presence of β-sheet crystals that run normal to the fiber axis in spider dragline silk, with a unit cell structure most closely resembling that of poly(alanine), and width approximately 40% crystallinity in the fiber (Warwicker, 1960). Recent X-ray measurements support the notion that poly(alanine) might be a better model for *N. clavipes* dragline silk than the silk from *B. mori*, despite the greater sequence similarity (Becker *et al.*, 1994).

No adequate picture has yet emerged that explains the unique functional properties observed with spider silk fibers. The approach described in this paper is the first step toward understanding the structure−function relationships in spider silk proteins. The influence of protein molecular weight and

composition on solution properties, fiber spinnability, and resulting mechanical properties can be examined with the methodologies described here and will permit a more systematic approach to understanding the effect of subtle changes in the repetitive domains on fiber properties.

ACKNOWLEDGMENT

We thank John Walker (Natick RD&E Center) for assistance with the circular dichroism measurements, Ted Thannhauser (Cornell University) for protein analyses, and Charlene Mello (Natick RD&E Center), Steve Arcidiacono (Natick RD&E Center), and Stanley Bedell (Geo-Centers, Inc.) for helpful discussions.

REFERENCES

Anderson, J. P., Cappello, J., & Martin, D. C. (1994) *Biopolymers 34*, 1049−1057.

Becker, M. A., *et al.* (1994) in *Silk Polymers, Materials Science & Biotechnology*, American Chemical Society Symposium Series 544, pp 185−196, American Chemical Society, Washington, DC.

Beckwitt, R., & Arcidiacono, S. (1994) *J. Biol. Chem. 269*, 6661−6663.

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513.

Candelas, G. C., & Cintron, J. (1981) *J. Exp. Zool. 212*, 1−6.

Candelas, G. C., & Lopez, F. (1983) *Comp. Biochem. Physiol. 74*, 637−641.

Canetti, M., Seves, A., Secundo, F., & Vecchio, G. (1989) *Biopolymers 28*, 1613−1624.

Cappello, J., Crissman, J., Dorman, M., Mikolajczak, M., Textor, G., Marquet M., & Ferrari, F. (1990) *Biotechnol. Prog. 6*, 198−202.

Craig, C. L. & Bernard, G. D. (1989) *Ecology 71*, 616−623.

Cunniff, P. M., Fossey, S. A., Auerbach, M. A., Song, J. W., Kaplan, D. L., Adams, W. W., Eby, R. K., Mahoney, D., & Vezie, D. L. (1994) *Polym. Adv. Technol. 5*, 401−410.

Ferrari, F. A., Richardson, C., Chambers, J., Causey, S. C., Pollock, T. J., Cappello, J., & Crissman, J. W. (1993) U.S. Patent 5,-243,038.

Fossey, S. A. & Kaplan, D. (1994) in *Silk Polymers, Materials Science & Biotechnology*, American Chemical Society Symposium Series 544, pp 270−282, American Chemical Society Washington, DC.

Fossey, S. A. Nemethy, G., Gibson, K. D., & Scheraga, H. (1991) *Biopolymers 31*, 1529.

Geddes, A. J., Parker, K. D., Atkins, E. D. T., & Beighton, E. (1968) *J. Mol. Biol. 32*, 343.

Gillespie, D. B., Viney, C., & Yager, P. (1994) in *Silk Polymers, Materials Science & Biotechnology*, American Chemical Society Symposium Series 544, pp 155−168, American Chemical Society, Washington, DC.

Gosline, J. M., Denny, M. W., & DeMont, M. E. (1984) *Nature 309*, 551−552.

Gosline, J. M., DeMont, M. E., & Denny, M. W. (1986) *Endeavour 10*, 37−43.

Hanahan, D. (1983) *J. Mol. Biol. 166*, 317.

Hinman, M. B., & Lewis, R. V. (1992) *J. Biol. Chem. 267*, 19320−19324.

Kaplan, D. L., Fossey, S., Mello, C. M., Arcidiacono, S., Senecal, K., Muller, W., Stockwell, S., Beckwitt, R., Viney, C., & Kerkam, K. (1992a) *Materials Res. Soc. Bull. 10*, 41−47.

Kaplan, D. L., Fossey, S., Viney, C., & Muller, W. (1992b) *Materials Res. Soc. Symp. Proc. 255*, 19−29.

Kerkam, K., Viney, C., Kaplan, D., & Lombardi, S. (1991) *Nature 349*, 596−598.

Kim, C. A., & Berg, J. M. (1993) *Nature 362*, 267−270.

Krejchi, M. T., Atkins, E. D. T., Waddon, A. J., Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1994) *Science 265*, 1427−1432.

Livengood, C. D. (1990) in *Polymers-Fibers and Textiles: A Compendium* (Kroschowitz, J. I., Ed.) Encyclopedia Reprint Series, pp 789−797, John Wiley and Sons, New York.

Lotz, B., & Keith, H. D. (1971) *J. Mol. Biol. 61*, 201.

Lotz, B., Brack, A., & Spach, G. (1974) *J. Mol. Biol. 87*, 193−203.

McBride, L. J., & M. H. Caruthers (1983) *Tetrahedron Lett. 24*, 245.

McGrath, K. P. (1991) Ph.D. Dissertation, University of Massachusetts at Amherst.

McGrath, K. P., Fournier, M. J., Mason, T. L., & Tirrell, D. A. (1992) *J. Am. Chem. Soc. 114*, 727−733.

Mello, C. M., Senecal, K., Yeung, B., Vouros, P., & Kaplan, D. L. (1994) in *Silk Polymers, Materials Science & Biotechnology*, American Chemical Society Symposium Series 544, pp 67−79, American Chemical Society, Washington, DC.

Minor, D. L., & Kim, P. S. (1994) *Nature 367*, 660−663.

Oshima, Y., & Suzuki, Y. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5363−5367.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Scopes, R. K. (1974) *Anal. Biochem. 59*, 277−282.

Simmons, A., Ray, E., & Jelinski, L. W. (1994) *Macromolecules 27*, 5235−5237.

Smith, C. K., Withka, J. M., & Regan, L. (1994) *Biochemistry 33*, 5510−5517.

Sprague, K. U., Roth, M. B., Manning, R. F., & Gage, L. P. (1979) *Cell 17*, 407−413.

Stydom, D. J., Haylett, T., & Stead, R. H. (1977) *Biochem. Biophys. Res. Commun. 79*, 932.

Thiel, B., Kunkel, D., Goess, K., & Viney, C. (1994) *Biopolymers 34*, 1089−1097.

Tillinghast, E. K. (1984) *Insect Biochem. 14*, 115−120.

Tillinghast, E. K., & Kavanagh, E. J. (1977) *J. Exp. Zool. 202*, 213−222.

Townley, M. A., & Tillinghast, E. K. (1988) *J. Arachnol. 16*, 303−319.

Tsujimoto, Y., & Suzuki, Y. (1979) *Cell 16*, 425−436.

Warwicker, J. O. (1960) *J. Mol. Biol. 2*, 350.

Xu, M., & Lewis, R. V. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 7120−7124.

BI9429593