Primary Structure Elements of Spider Dragline Silks and Their Contribution to Protein Solubility†

Daniel Huemmerich,‡ Christopher W. Helsen,‡ Susanne Quedzuweit,‡ Jan Oschmann,§ Rainer Rudolph,§ and Thomas Scheibel*,‡

Department of Biotechnology, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany, and Department of Biotechnology, Martin-Luther-Universität Halle-Wittenberg, Kurt-Mothes-Strasse 3, 06120 Halle (Saale), Germany

Received May 19, 2004; Revised Manuscript Received July 16, 2004

ABSTRACT: Spider silk proteins have mainly been investigated with regard to their contribution to mechanical properties of the silk thread. However, little is known about the molecular mechanisms of silk assembly. As a first step toward characterizing this process, we aimed to identify primary structure elements of the garden spider's (*Araneus diadematus*) major dragline silk proteins ADF-3 and ADF-4 that determine protein solubility. In addition, we investigated the influence of conditions involved in mediating natural thread assembly on protein aggregation. Genes encoding spider silk-like proteins were generated using a cloning strategy, which is based on a combination of synthetic DNA modules and PCR-amplified authentic gene sequences. Comparing secondary structure, solubility, and aggregation properties of the synthesized proteins revealed that single primary structure elements have diverse influences on protein characteristics. Repetitive regions representing the largest part of dragline silk proteins determined the solubility of the synthetic proteins, which differed greatly between constructs derived from ADF-3 and ADF-4. Factors, such as acidification and increases in phosphate concentration, which promote silk assembly *in vivo* generally decreased silk protein solubility *in vitro*. Strikingly, this effect was pronounced in engineered proteins comprising the carboxyl-terminal nonrepetitive regions of ADF-3 or ADF-4, indicating that these regions might play an important role in initiating assembly of spider silk proteins.

Spider silks are protein polymers that display extraordinary physical properties (*1*). Among the different types of spider silks, draglines are most intensely studied. Dragline silks are utilized by orb weaving spiders to build frame and radii of their nets and as lifelines that are permanently dragged behind. For these purposes, high tensile strength and elasticity are required. The combination of such properties results in a toughness that is higher than that of most other known materials (*1*, *2*). Dragline silks are generally composed of two major proteins whose primary structures share a common repetitive architecture (*3*, *4*). Variations of a single repeat unit, which comprises up to 60 amino acids, are iterated several times to represent the largest part of a dragline silk sequence. These repeat units comprehend a limited set of distinct amino acid motifs. One motif found in all dragline silk repeat units is a block of typically six to nine alanine residues. In silk threads, several polyalanine motifs form crystalline β-sheet stacks leading to tensile strength (*5*, *6*). Glycine rich motifs such as GGX or GPGXX adopt flexible helical structures that connect crystalline regions and provide elasticity to the thread (*7*). Additionally, all investigated dragline silk proteins comprise regions at their carboxyl termini that display no obvious repetition pattern [nonrepetitive (NR)[1] regions]. Since so far no function and structural role could be assigned to these regions in the final thread, it is likely that they play a role in thread assembly.

Silk assembly *in vivo* is a remarkable process. Spider dragline silk proteins are stored at concentrations of up to 50% (w/v) (*8*) in the so-called major ampullate gland. Although a "dynamic loose helical structure" has been proposed for the proteins within the major ampullate gland (*8*), more recent data suggest a mainly random coil conformation for the proteins of the so-called A-zone, which represents the largest part of the gland (*9*, *10*). The highly concentrated protein solution forms the silk dope (spinning solution). Thread assembly is initiated during a passage of the dope through the spinning duct accompanied by extraction of water, sodium, and chloride (*11*, *12*). At the same time, the concentrations of the more lyotropic ions potassium and phosphate are increased and the pH drops from 6.9 to 6.3 (*11*−*13*). Assembly is finally triggered by mechanical stress, which is caused by pulling the thread out of the spider's abdomen (*14*).

† This work was supported by the Bundesministerium für Bildung und Forschung (D.H.), by the Fonds der Chemischen Industrie (D.H. and T.S.), and by the Deutsche Forschungsgemeinschaft (SFB563).

* To whom correspondence should be addressed. Phone: +49-89-28913179. Fax: +49-89-28913345. E-mail: thomas.scheibel@ch.tum.de.

‡ Technische Universität München.

§ Martin-Luther-Universität Halle-Wittenberg.

[1] Abbreviations: NR, nonrepetitive; *Ap*^{r}, ampicillin resistance gene; IPTG, isopropyl β-D-thiogalactoside; GdmCl, guanidinium chloride; GdmSCN, guanidinium thiocyanate; SDS, sodium dodecyl sulfate; PAGE, polyacrylamide gel electrophoresis; Tris, tris(hydroxymethyl)-aminomethane; CD, circular dichroism; rep-proteins, repetitive proteins; NR, nonrepetitive; cps, counts per second; MRW, mean residue weight; nd, not determined.

10.1021/bi048983q CCC: $27.50
Published on Web 09/29/2004

While some structural aspects of spider silk proteins have been unravelled, little is known about the contribution of individual silk proteins and their primary structure elements to the assembly process. Comparative studies of the two major dragline silk proteins of the garden spider *Araneus diadematus*, ADF-3 and ADF-4, revealed that, although their amino acid sequences are rather similar (*4*), they display remarkably different solubility and assembly characteristics. While ADF-3 is soluble even at high concentrations (*15*), ADF-4 is virtually insoluble and self-assembles into filamentous structures under specific conditions (unpublished results). To study molecular details of silk assembly, we systematically investigated the influence of repeat units and NR regions of ADF-3 and ADF-4 on solubility, assembly, and aggregation properties of the silk proteins.

EXPERIMENTAL PROCEDURES

Materials. Chemicals were obtained from Merck KGaA (Darmstadt, Germany) if not otherwise stated. Manipulation and modification of DNA were performed as described previously (*16*). Restriction enzymes were obtained from New England Biolabs (Beverly, MA), and ligase was obtained from Promega Biosciences Inc. (San Luis Obispo, CA). DNA purification was performed using kits from Qiagen (Hilden, Germany). Synthetic oligonucleotides were obtained from MWG Biotech AG (Ebersberg, Germany). All cloning steps were performed in *Escherichia coli* strain DH10B from Novagen (Madison, WI).

Construction of Cloning Vector pAZL. A cloning cassette with cohesive ends complementary to ones generated by *Bgl*II and *Hin*dIII was created by annealing two synthetic oligonucleotides, CC1 (GATCGAGGAGGATCCATGGGACGAATTCACGGCTAATGAAAGCTTACTGCAC) and CC2 (AGCTGTGCAGTAAGCTTTCATTAGCCGTGAATTCGTCCCATGGATCCTCCTC). Annealing was accomplished by decreasing the temperature of a 50 pmol/μL (each) oligonucleotide solution from 95 to 20 °C with an increment of 0.1 °C/s. Mismatched double strands were denatured at 70 °C followed by another temperature decrease to 20 °C. After the 20 °C−70 °C−20 °C cycle had been repeated 10 times, 10 additional cycles were performed with a denaturing temperature of 65 °C. The resulting cloning cassette was ligated with a pFastbac1 vector (Invitrogen, Carlsbad, CA) digested with *Bgl*II and *Hin*dIII. Both restriction enzyme recognition sequences were destroyed upon this cloning step. The resulting cloning vector was named pAZL.

Cloning of Silk Modules and NR Regions into the pAZL Vector. Three amino acid modules derived from dragline silk proteins ADF-3 and ADF-4 (Figure 1E) were back translated into a DNA sequence considering bacterial codon usage. Corresponding complementary DNA oligonucleotides A1 (TCCGTACGGCCCAGGTGCTAGCGCCGCAGCGGCAGCGGCTGGTGGCTACGGTCCGGGCTCTGGCCAGCAGGG) and A2 (CTGCTGGCCAGAGCCCGGACCGTAGCCACCAGCCGCTGCCGCTGCGGCGCTAGCACCTGGGCCGTACGGACC), Q1 (TCCGGGCCAGCAGGGCCCGGGTCAACAGGGTCCTGGCCAGCAAGGTCCGGGCCAGCAGGG) and Q2 (CTGCTGGCCCGGACCTTGCTGGCCAGGACCCTGTTGACCCGGGCCCTGCTGGCCCGGACC), and C1 (TTCTAGCGCGGCTGCAGCCGCGGCAGCTGCGTCCGGCCCGGGTGGCTACGGTCCGGAAAACCAGGGTCCATCTGGCCCGGGTGGCTACGGTCCTGGCGGTCCGGG) and C2 (CGGACCGCCAGGACCGTAGCCACCCGGGCCAGATGGACCCTGGTTTTCCGGACCGTAGCCACCCGGGCCGGACGCAGCTGCCGCGGCTGCAGCCGCGCTAGAACC) were synthesized and annealed as described above and ligated with the pAZL vector digested with *Bsg*I and *Bse*RI. NR regions of spider silk genes *adf-3* (gi|1263286) and *adf-4* (gi|1263288) (obtained from J. Gosline, Vancouver, BC) were amplified by PCR using the following primers: NR3f (GAAAAACCATGGGTGCGGCTTCTGCAGCTG TATCTG), NR3r (GAAAAGAAGCTTTCATTAGCCAGCAAGGGCTTGAGCTACAGATTG), NR4f (GAAAAACCATGGGAGCATATGGCCCATCTCCTTC), and NR4r (GAAAAGAAGCTTTCATTAGCCTGAAAGAGCTTGGCTAATCATTTG). PCR products and pAZL vector were ligated after digestion with *Nco*I and *Hin*dIII. Cloning of synthetic modules as well as PCR products resulted in the replacement of the cloning cassette's spacer, preserving the arrangement of its elements. For more efficient translation, the codon AGA (Arg), which is rarely translated in *E. coli*, was mutated to CGT (Arg) in NR3 and NR4 using PCR mutagenesis (*16*).

Construction of Synthetic Spider Silk Genes. Connecting of two gene fragments, e.g., single modules, module multimers, or NR, regions represented the basic step of the cloning strategy. For this purpose, the pAZL vector, containing the designated 5′-terminal gene fragment, was digested with *Bsa*I and *Bsg*I, while the vector comprising the 3′-terminal gene fragment was digested with *Bse*RI and *Bsa*I (Figure 1B). Ligation of the appropriate plasmid fragments yielded the connecting of the two gene fragments and led to the reconstitution of the pAZL vector's ampicillin resistance gene (Ap^r), which facilitated identification of correct constructs.

For gene construction, single modules were first connected to yield repeat units (Figure 1D). These were gradually multimerized and optionally linked with NR regions. Finally, synthetic gene constructs as well as NR regions were excised from the pAZL vector with *Bam*HI and *Hin*dIII and ligated with bacterial expression vector pET21a (Novagen) likewise digested, providing a T7 tag (MASMTGGQQMGR) coding sequence (*17*). The fidelity of all constructs was confirmed by DNA sequencing.

Gene Expression. All silk genes were expressed in *E. coli* strain BLR(DE3) (Novagen). Cells were grown at 37 °C in LB medium to an OD_{600} of 0.5. Before induction with 1 mM IPTG (isopropyl β-D-thiogalactoside), cells were shifted to 30 °C in the case of $(AQ)_{12}$, $(AQ)_{12}$NR3, $(QAQ)_8$, and $(QAQ)_8$NR3 and to 25 °C in the case of C_{16}, C_{16}NR4, NR3, and NR4, respectively. Alternatively, cells were grown in a fermenter to an OD_{600} of 40−50 using complex media (*18*) and the fed-batch technique (*19*). Again, before induction with 1 mM IPTG, cells were shifted to 25 or 30 °C. Cells expressing $(AQ)_{12}$, $(AQ)_{12}$NR3, $(QAQ)_8$, $(QAQ)_8$NR3, C_{16}, and C_{16}NR4 were harvested after induction for 3−4 h, while cells expressing NR3 and NR4 were harvested after 16 h.

Protein Purification. Cells were resuspended with 5 mL of buffer/g containing 20 mM *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid (HEPES) (pH 7.5), 100 mM NaCl, and 0.2 mg/mL lysozyme (Sigma-Aldrich, St. Louis, MO) and incubated at 4 °C for 30 min. Cells were lysed by

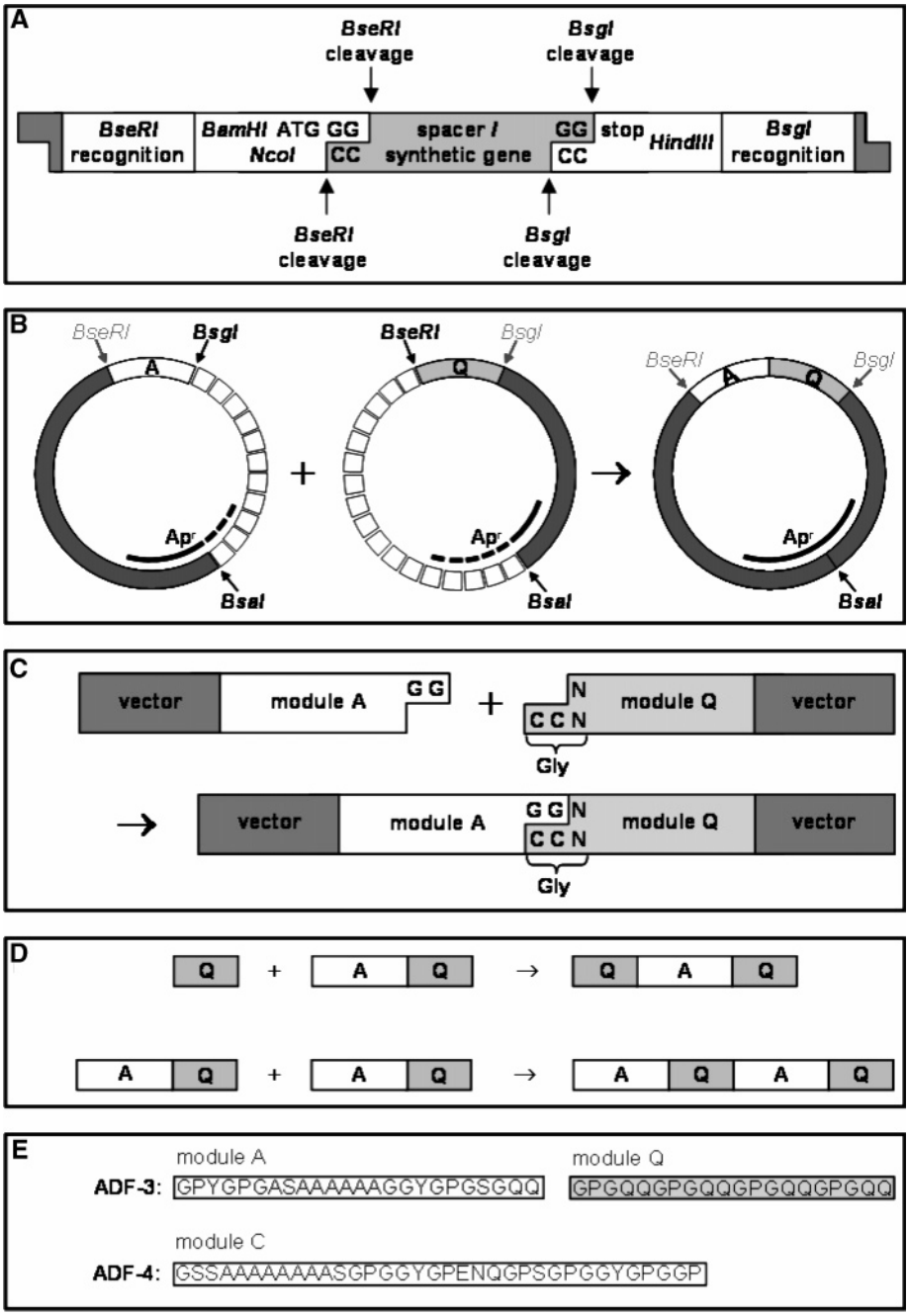

FIGURE 1: Cloning strategy for constructing synthetic spider silk genes. (A) The cloning cassette comprised restriction sites required for module multimerization (*Bsg*I and *Bse*RI) and for excising assembled genes (*Nco*I, *Bam*HI, and *Hin*dIII). During gene construction, the spacer region was replaced with modules and module multimers. (B) Site-directed connecting of two modules was accomplished by ligating two appropriate plasmid fragments. The vector's ampicillin resistance gene (Ap^r) was reconstituted. (C) Nucleotides required for linking two modules were confined within the first codon of each module. (D) Module multimers were connected like single modules, resulting in controlled assembly of synthetic genes. (E) Amino acid sequences of designed silk modules were derived from dragline silk proteins ADF-3 and ADF-4.

sonication using an HD/UW2200/KE76 ultrasonicator (Bandelin, Berlin, Germany), and genomic DNA was digested by incubating cell lysates with 0.1 mg/mL DNase I (Roche, Mannheim, Germany) in the presence of 3 mM $MgCl_2$ at 4 °C for 60 min. Insoluble cell fragments were sedimented at 50000*g* and 4 °C for 30 min. Soluble *E. coli* proteins of lysates containing $(AQ)_{12}$, $(AQ)_{12}$NR3, $(QAQ)_8$, $(QAQ)_8$NR3, C_{16}, and C_{16}NR4 were precipitated by heat denaturation at 80 °C for 20 min, while lysates containing NR3 and NR4 were heated to 70 °C for the same length of time. Precipitated proteins were removed by sedimentation at 50000*g* for 30 min. Silk proteins, which remained soluble during heat denaturation, were precipitated with 20% ammonium sulfate [$(AQ)_{12}$, $(AQ)_{12}$NR3, $(QAQ)_8$, $(QAQ)_8$NR3, C_{16}, and C_{16}NR4] or 30% ammonium sulfate (NR3 and NR4) at room temperature and harvested by centrifugation at 10000*g* for 10 min. Pellets of $(AQ)_{12}$, $(AQ)_{12}$NR3, $(QAQ)_8$, $(QAQ)_8$NR3, NR3, and NR4 were rinsed with a solution containing the same concentration of ammonium sulfate as used for precipitation and dissolved in 6 M guanidinium chloride (GdmCl). In contrast, C_{16} and C_{16}NR4 were washed with 8 M urea and dissolved in 6 M guanidinium thiocyanate (GdmSCN). All proteins were dialyzed against 10 mM NH_4HCO_3. Precipitates formed during dialysis were removed

Table 1: Selected Properties of Synthetic Silk Constructs and Authentic Spider Silk Proteins ADF-3 and ADF-4

	$(QAQ)_8$	$(AQ)_{12}$	C_{16}	NR3	NR4	$(QAQ)_8$NR3	$(AQ)_{12}$NR3	C_{16}NR4	ADF-3	ADF-4
molecular mass (kDa)[a]	47.5	48.1	47.7	13.3	11.9	59.3	59.8	58.1	56.1	34.9
extinction coefficient (276 nm) (M^{-1} cm^{-1})[b]	23200	34800	46400	4423	1523	27550	39150	47850	nd	nd
no. of charged amino acid residues[c] (positive/negative)	0/0	0/0	0/16	2/2	2/2	2/2	2/2	2/18	4/2	2/6
mean hydropathicity[d]	−1.252	−0.987	−0.464	0.401	0.438	−0.918	−0.710	−0.294	−0.628	−0.075
normalized mean hydropathicity, boundary hydropathicity[e]	0.361, 0.413	0.390, 0.413	0.448, 0.424	0.545, 0.413	0.548, 0.413	nd	nd	nd	0.399,[f] 0.415[f]	0.464,[f] 0.417[f]
midpoint temperature of thermal unfolding (°C)[g]	no	no	no	64	69	67	66	72	nd	nd
solubility (%, w/v)[h]	>30	>30	8	nd	nd	>30	>30	9	>28	<1

[a] The molecular mass of engineered proteins includes the T7 tag. [b] Extinction coefficients were calculated according to the method of Gill and Hippel (*20*). [c] Charged amino acid residues refer to silk gene sequences only; T7 tags comprise an additional arginine. [d] The mean hydropathicity was calculated as described previously (*37*). Hydrophobicity increases with hydropathicity values. [e] The mean hydropathicity was normalized to a range between 0 and 1. Boundary hydropathicity was calculated according to the method of Uversky et al. (*33*, *34*). If normalized hydropathicity values are less than the boundary value, proteins are predicted to be intrinsically unfolded. [f] Values of ADF-3 and ADF-4 refer to their repetitive sequences only. [g] Midpoint temperatures were determined by CD spectroscopy. [h] Values for ADF-3 and ADF-4 were taken from ref *15* and unpublished results.

by sedimentation at 50000*g* for 30 min, and the remaining soluble silk proteins were lyophilized. Prior to analysis, the lyophilized protein was dissolved in 6 M GdmSCN followed by dialysis against appropriate buffers. Aggregates were removed by sedimentation at 125000*g* for 30 min. Protein concentrations were determined photometrically in a 1 cm path length cuvette at 276 nm using calculated extinction coefficients (Table 1) (*20*). Identity of proteins was confirmed by sodium dodecyl sulfate−polyacrylamide gel electrophoresis [SDS−PAGE; 10% Tris-Glycine gels for >20 kDa proteins and 10−20% Tris-Tricine gels (Invitrogen) for <20 kDa proteins] followed by blotting onto polyvinylidene fluoride (PVDF) membranes (Millipore, Billerica, MA) and detection using a mouse anti-T7 monoclonal antibody (Novagen, 1:10000) as the primary antibody and an anti-mouse IgG peroxidase conjugate (Sigma-Aldrich, 1:5000) as the secondary antibody. Peroxidase activity was visualized using the ECL^{plus} Western blot detection kit from Amersham Biosciences (Piscataway, NJ).

Fluorescence. Fluorescence spectra were recorded on a FluoroMax spectrofluorometer (Jobin Yvon Inc., Edison, NJ). Spectra were recorded using a protein concentration of 100 μg/mL in 10 mM tris(hydroxymethyl)aminomethane (Tris)-HCl (pH 8.0) at room temperature. The integration time was 1 s and the step size 0.5 nm, and bandwidths were 5 nm (excitation) and 5 nm (emission).

Secondary Structure Analysis. Far-UV circular dichroism (CD) spectra were obtained using a Jasco 715 spectropolarimeter equipped with a temperature control unit (Jasco International Co. Ltd., Tokyo, Japan). All spectra were recorded at a protein concentration of 150 μg/mL in 5 mM Tris-HCl (pH 8.0) in a 0.1 cm path length quartz cuvette at 20 °C. The scan speed was 20 nm/min, the step size 0.2 nm, the integration time set to 1 s, and the bandwidth 1 nm. Four scans were averaged and buffer-corrected. Thermal transitions were analyzed with a heating and cooling increment of 1 °C/min at 220 nm.

Solubility Assay. To determine the maximal concentration of soluble proteins, a 1 mg/mL [0.1% (w/v)] solution in 10 mM Tris-HCl (pH 8.0) was concentrated by ultrafiltration using a 10 000 Da molecular mass cutoff polyether sulfone membrane (Vivascience AG, Hannover, Germany). At distinct intervals, samples were taken from the solution until the protein started to precipitate. Samples were diluted in 10 mM Tris (pH 8.0) to determine protein concentrations photometrically.

Aggregation Assay. All samples were adjusted to 1 mg/mL in 10 mM Tris-HCl (pH 8.0). For testing ionic effects on silk protein aggregation, salts were added to final concentrations of 300 mM. The effect of acidification was investigated by adding HCl to a final concentration of 100 mM (pH 1). All samples were incubated at room temperature for 1 h. Protein precipitates were removed from all samples by sedimentation at 125000*g* for 25 min, and the amount of the remaining soluble protein was determined photometrically. Since the sum of soluble and aggregated protein had to equal the initial amount of soluble protein, the percentage of the aggregated protein could be calculated by subtracting the amount of soluble protein from the initially used amount of protein.

RESULTS

A Cloning Strategy for Designing Silk-like Proteins. Expression of authentic spider silk genes in bacterial hosts is inefficient (*21*) since some parts of the genes contain codons that are not efficiently translated in bacteria. In addition, gene manipulation and amplification by PCR are difficult due to the repetitive nature of silks. To investigate properties of spider silk proteins, cloning strategies have been employed using synthetic DNA modules with a codon usage adapted to the corresponding expression host. Synthetic genes were obtained which encoded proteins resembling the repetitive regions of spider silks (*22*−*25*). Importantly, none of these protein designs included the carboxyl-terminal NR regions that are found in all dragline silks.

We developed a seamless cloning strategy (*26*) that allowed controlled combination of different synthetic DNA modules as well as authentic gene fragments. Cloning vector pAZL was designed comprising a cloning cassette with a spacer acting as a placeholder for synthetic genes, and recognition sites for restriction enzymes *Bse*RI and *Bsg*I (Figure 1A). Since recognition and cleavage sites of these enzymes are eight (*Bse*RI) or 12 (*Bsg*I) nucleotides apart, translation start and stop codons as well as additional restriction sites required for the excision of assembled genes could be positioned close to the spacer.

In a first cloning step, the spacer region of pAZL was replaced with a synthesized DNA module (for module design, see below). Subsequently, two modules could be joined in a site-directed way (see Experimental Procedures and Figure 1B). The complementary 3′-single-strand extensions GG (sense) and CC (antisense) generated by cleavage with *Bsg*I and *Bse*RI were used for connecting two modules (Figure 1C). Thus, the DNA sequence required to link two modules was confined to a glycine codon (GGX). Glycine is naturally abundant in spider silk proteins (~30%); therefore, modules could be designed without the need to search for restriction endonuclease recognition sites which, after translation, match authentic amino acid sequences. Since the arrangement of the cloning cassette's elements remained unchanged upon cloning and multimerization, a variety of module combinations could be constructed (Figure 1D).

Design, Synthesis, and Purification of Synthetic Spider Silks. We chose dragline silk proteins ADF-3 and ADF-4 (3) from the garden spider *A. diadematus* as templates for the synthetic constructs. The partially identified primary structure of ADF-3 largely consists of repeat units, which all comprise a consensus sequence that includes a polyalanine motif. The length of individual repeat units is determined by varying the numbers of the GPGQQ motif. To mimic the repetitive sequence of ADF-3, we designed two modules. The sequence of one module, termed A, was derived from the polyalanine-containing consensus sequence (Figure 1E). The sequence of a second module termed Q contained four repeats of the GPGQQ motif. To study repeat units of different lengths, one or two Q modules were combined with one A module to obtain (AQ) or (QAQ). These repeat units were multimerized to generate synthetic genes encoding repetitive proteins (rep-proteins) $(AQ)_{12}$ and $(QAQ)_8$.

The repetitive part of ADF-4 is generally composed of a single conserved repeat unit displaying only slight variations. We combined these variations and designed one consensus module termed C (Figure 1E), which we multimerized to obtain rep-protein C_{16}. The number of module repeats in all synthetic genes was chosen to encode proteins with a similar molecular mass (~50 kDa).

ADF-3 and ADF-4 both display NR regions at their carboxyl termini, comprising 124 and 109 amino acids, respectively. Gene sequences encoding these regions were amplified by PCR, and codons problematic for bacterial expression were changed to more suitable codons by site-directed mutagenesis (see Experimental Procedures). Therefore, all of our synthetic genes could be combined with the appropriate authentic NR regions, yielding genes encoding rep-NR proteins $(AQ)_{12}$NR3, $(QAQ)_8$NR3, and C_{16}NR4. Additionally, NR3 and NR4 could be expressed individually.

After bacterial synthesis, all silk proteins were purified by a heat step followed by an ammonium sulfate precipitation. The identity of the proteins was confirmed by immunoblotting, using antibodies directed against T7 peptide tag sequences, attached to the amino-terminal end of all silk proteins (Figure 2A). Although all rep-proteins and all rep-NR proteins had similar molecular masses (Table 1), they displayed different migration velocities when subjected to SDS−PAGE. This effect might be caused by aberrant binding of dodecyl sulfate to the proteins, leading to variation of the proteins' net charges. Besides full-length proteins, immunoblotting revealed traces of proteins with lower

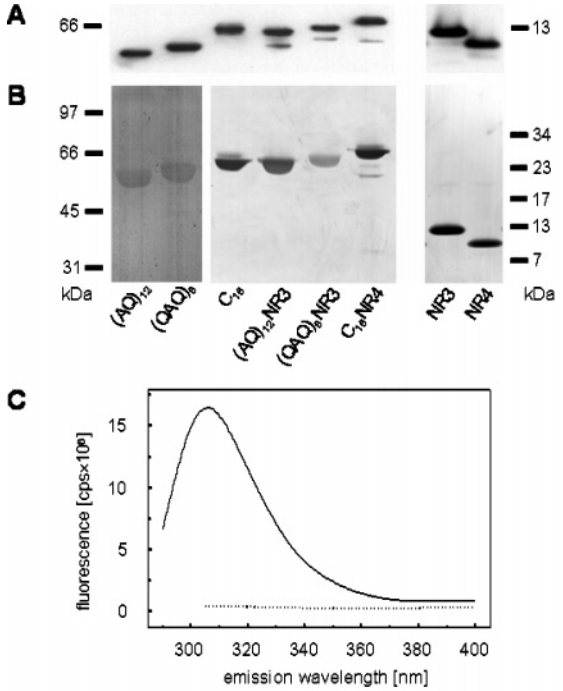

FIGURE 2: Analysis of spider silk proteins. (A) T7 tags of recombinant silk proteins were detected after Western blotting with an anti-T7 tag antibody. (B) Proteins were subjected to SDS−PAGE followed by silver staining. Due to weak staining of $(AQ)_{12}$ and $(QAQ)_8$, the contrast of the image was increased electronically. (C) Fluorescence emission spectra of purified C_{16}NR4 are shown with an excitation wavelength of 280 (—) or 295 nm (•••).

molecular masses within preparations of rep-NR proteins. Binding of the anti T7-tag antibody to these proteins identified them as silk proteins lacking part of their carboxyl-terminal end. After each purified protein was analyzed by SDS−PAGE and silver staining, no additional bacterial proteins were detected in all protein preparations (Figure 2B). Protein purity was further determined by measuring fluorescence emission. Incident 280 nm light leads to excitation and fluorescence emission of tyrosines and tryptophans, while 295 nm light exclusively excites the latter. Since none of the designed spider silk proteins comprised tryptophans, fluorescence emission upon excitation with 295 nm would have been indicative of contaminating *E. coli* proteins, which on average contain 1.5% tryptophan (*27*). Fluorescence measurements of all silk protein preparations revealed emission spectra akin to the spectrum of tyrosine, which occurs abundantly in the silk proteins. In contrast, no tryptophan fluorescence could be detected, indicating the high purity of the protein preparations (data of C_{16}NR4 shown in Figure 2C).

Bacterial production of synthetic silk proteins in Erlenmeyer flasks yielded similar protein amounts for all constructs. Yields of individual preparations ranged from 10 to 30 mg of purified protein per liter of culture medium. Fermentation of cells was employed to investigate the possibility of upregulating protein synthesis. Yields of $(QAQ)_8$NR3 and C_{16}NR4 thus could be increased to 140 and 360 mg/L, respectively.

Rep-NR Proteins Consist of a Poorly Structured Repetitive Region and a Highly Structured Nonrepetitive Domain. Secondary structure was investigated by CD spectroscopy. Rep-proteins exhibited spectra typical for intrinsically un-

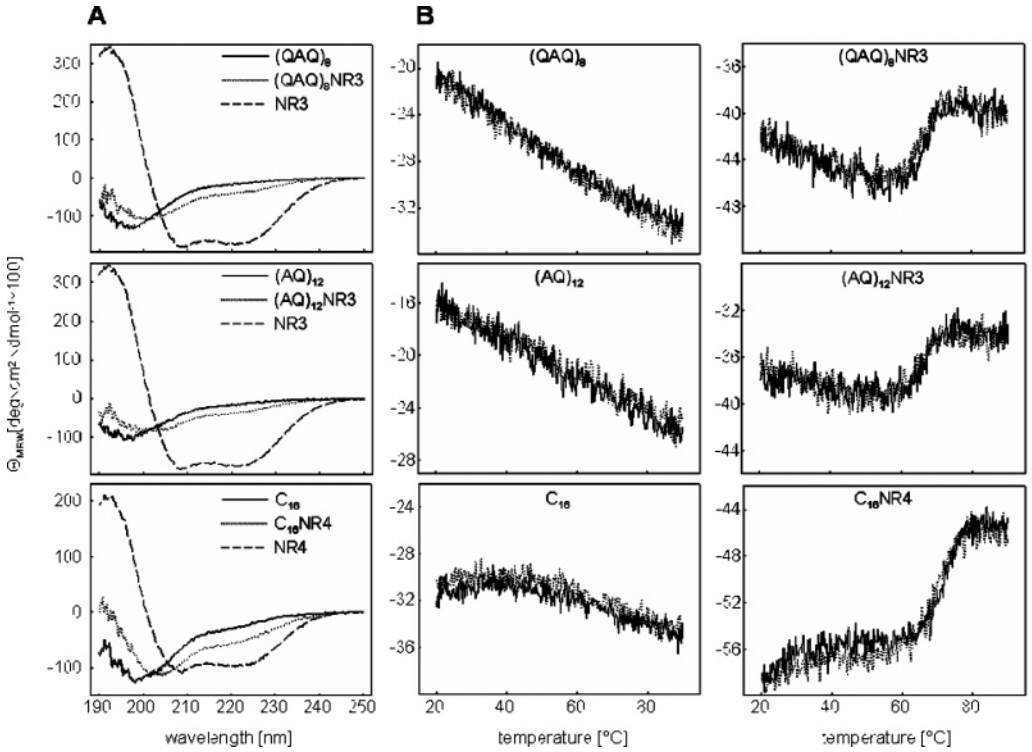

FIGURE 3: Secondary structure and temperature transitions of spider silk proteins. (A) CD spectra of rep-proteins (—), rep-NR proteins (•••), and NR proteins (– – –) were recorded at 20 °C. (B) Mean residue weight (MRW) ellipticities of soluble synthetic spider silk proteins were measured at 220 nm while heating to 90 °C (—), followed by cooling to 20 °C (•••).

structured proteins. In contrast, NR proteins exhibited spectra indicative of secondary structure comprising α-helices (Figure 3). Spectra of rep-NR proteins roughly corresponded to a combination of the rep and NR spectra weighted according to their share in the rep-NR proteins (Figure 3 and data not shown). Therefore, it is likely that the rep-NR proteins are composed of a repetitive region displaying mainly random coil structure and a carboxyl-terminal non-repetitive folded protein domain.

Silk Proteins Refold after Thermal and Chemical Denaturation. When thermally induced structural changes were investigated via CD spectroscopy, no cooperative temperature transitions were observed for rep-proteins between 20 and 90 °C, an effect which has also been observed for other intrinsically unfolded proteins (*28*, *29*) (Figure 3). Since NR proteins displayed secondary structure content, thermal unfolding should be detectable at elevated temperatures. Accordingly, cooperative thermal transitions were observed. Midpoints of temperature transitions were 64 °C (NR3) and 69 °C (NR4) (Table 1). Rep-NR proteins also displayed cooperative unfolding with transition temperatures 2−3 °C above the values of the corresponding NR domains [67 °C for $(QAQ)_8$NR3, 66 °C for $(AQ)_{12}$NR3, and 72 °C for C_{16}NR4 (Figure 3B and Table 1)]. This indicates that thermal unfolding of rep-NR proteins is determined mainly by the NR domains, which seem to be slightly stabilized when connected with a repetitive sequence. Furthermore, all thermal transitions were completely reversible. The reversibility of the structural changes upon heating explained the high recovery of soluble silk proteins after the heat step employed during protein purification. Tris was used to buffer all solutions investigated by CD spectroscopy, because of good spectral properties and little capacity to promote silk protein aggregation. Because of the strong temperature dependence of Tris-buffered solutions, the pH of the samples was expected to shift from pH 8 to 6 upon heating from 20 to 90 °C (*16*). However, temperature transitions of silk proteins in phosphate buffer at pH 8, displaying a temperature-independent p*K* value, revealed equal midpoints of thermal unfolding (data not shown), although they were not entirely reversible probably due to protein aggregation (see below). This indicated that thermal unfolding of silk proteins was not influenced by thermally induced changes in the pH in Tris-buffered solutions.

Chemical de- and renaturation of silk proteins was employed during the purification of silk proteins. To test the effect of this treatment on secondary structure, CD was measured of rep-NR proteins in Tris buffer, after dialysis against 6 M GdmCl and renaturation by dialysis against Tris buffer. The identical spectra of the initial and refolded proteins indicated that chemical denaturation is completely reversible (data not shown).

The Solubility of Silk Proteins Is Determined by Their Repetitive Sequences. To obtain the high protein concentrations found in the dope, silk proteins must be highly soluble. We tested the maximum concentrations at which rep-proteins and rep-NR proteins remained soluble, to identify primary structure elements that determine solubility. All proteins comprising modules A and Q could be concentrated by ultrafiltration to more than 30% (w/v) without forming visible aggregates, regardless of the presence of the NR domain. In contrast, proteins containing module C could only be concentrated to 8% (w/v) (C_{16}) and 9% (w/v) (C_{16}NR4) (Table 1). Both proteins transformed into a gel-like state upon further concentration (data not shown). Thus, solubility of the silk proteins was solely determined by their repetitive sequences and was not influenced by the NR domain.

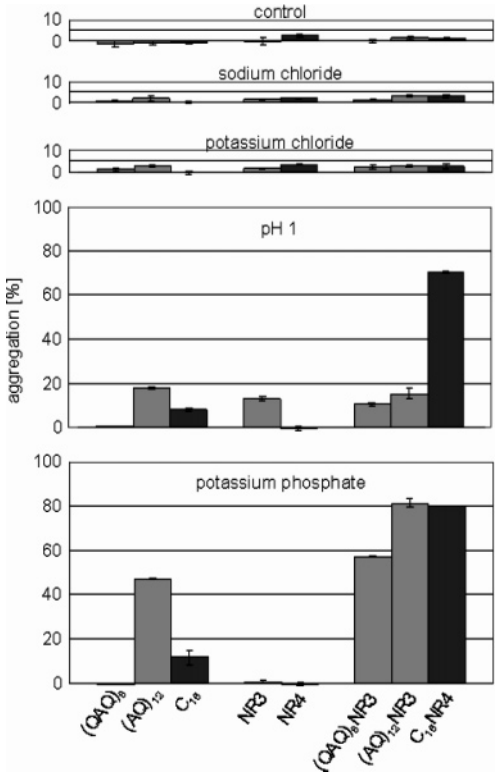

FIGURE 4: Aggregation of synthetic spider silk proteins. Aggregation of proteins was assessed after incubation for 1 h in buffer (control), in the presence of 300 mM NaCl, or 300 mM KCl, at pH 1 or in the presence of 300 mM potassium phosphate. Bars for proteins derived from ADF-3 are light gray and from ADF-4 are dark gray.

Potassium Does Not Promote Aggregation of Synthetic Silk Proteins, Independent of Their Primary Structure. pH and ions, such as potassium and phosphate, are involved in natural silk assembly. Here we wanted to investigate how these factors promote the assembly of synthetic silk proteins. Since we were unable to imitate the authentic assembly process, which includes orientation of the involved proteins by extensional flow, we performed an aggregation assay using protein solutions not displaying orientational order. None of the tested rep-, rep-NR, and NR proteins displayed significant aggregation (<5%) when incubated in buffer for 1 h, indicating that all proteins were intrinsically soluble under the testing conditions (Figure 4). To investigate whether addition of ions caused aggregation by increasing the ionic strength, proteins were incubated with 300 mM sodium chloride. However, no aggregation was observed. In contrast to sodium, potassium has previously been reported to specifically promote silk aggregation (*30*), yet 300 mM potassium chloride also showed no influence on the solubility of the synthetic silk proteins (Figure 4).

Addition of Phosphate and Acidification Initiate Aggregation of Rep-Proteins Depending on Their Primary Structure. Phosphate has been known to be added to the dope during the natural spinning process. Potassium phosphate (300 mM) caused no aggregation of $(QAQ)_8$ and weak precipitation of C_{16} (12%). In contrast, $(AQ)_{12}$ displayed a stronger tendency to aggregate (47%) upon phosphate treatment. Similar results were obtained using sodium phosphate, indicating that the effect is specifically caused by phosphate ions (data not shown).

The exact function of acidification during spider silk assembly has not yet been determined. However, it seems likely that negatively charged groups are protonated, reducing the net charge of spider silk proteins. Phosphoryl groups of phosphoamino acid residues, which have been detected in dragline silk (*31*), display pK_a values (*32*) near the range of the pH shift observed during the spinning process and thus could be involved in triggering silk assembly. Since the synthetic silk proteins did not contain such groups, we aimed to mimic charge reduction by protonating terminal and side chain carboxyl groups. $(QAQ)_8$ and $(AQ)_{12}$, displaying only the terminal carboxyl group, showed no (<5%) and weak (18%) aggregation at pH 1. Interestingly, protonation of C_{16}'s 16 glutamate residues also caused only weak aggregation (8%) (Figure 4).

NR Domains Amplify the Response to Factors that Promote Aggregation. To investigate the influence of NR domains on aggregation, NR proteins and rep-NR proteins were incubated at low pH and with phosphate. Despite the likely protonation of the two negatively charged amino acid residues of each NR domain, only weak (NR3, 13%) or no (NR4, 0%) aggregation was observed at pH 1. Acidification of $(QAQ)_8$NR3 and $(AQ)_{12}$NR3 caused weak aggregation (10 and 15%, respectively), which was in the range displayed by the corresponding rep-proteins. Interestingly, although the NR4 domain alone did not precipitate, C_{16}NR4 showed strong aggregation at pH 1 (70%). Thus, the combination of the repetitive C_{16} and the NR4 domain, which did not significantly aggregate upon acidification by themselves, led to a protein highly sensitive to this aggregation-promoting factor. Similar results were obtained for the addition of phosphate. While neither NR3 nor NR4 showed aggregation in the presence of phosphate (1 or 0%, respectively), the addition of the NR domains to the repetitive regions caused an increased aggregation of the rep-NR proteins in comparison to rep-proteins [57% for $(QAQ)_8$NR3, 81% for $(AQ)_{12}$NR3, and 80% for C_{16}NR4].

DISCUSSION

Using a cloning strategy that allows seamless and controlled assembly of DNA modules, synthetic genes encoding spider silk-like proteins were constructed. The design of proteins yielded different combinations of repeat units and naturally occurring NR regions, for systematic testing of the properties of such single primary structure elements. Structural analysis by CD spectroscopy revealed that repetitive regions are mainly unstructured in their soluble state, displaying properties common to other intrinsically unfolded proteins (*28*, *29*). The same conformational state has been proposed for the largest part of the major ampullate content (*10*). In contrast, NR regions were found to represent independently folding protein domains. Because of the high degree of similarity between the NR domains of ADF-3 and ADF-4 (81% similar and 67% identical), it can be assumed that both exhibit similar structures and might fulfill related functions.

Thermal and chemical denaturation of rep-NR proteins is reversible. Further, CD data obtained from refolded rep-NR proteins and from natural silk dope (*9*) are remarkably similar. Therefore, it can be assumed that even after treatment with heat and chaotropic reagents during purification and

sample preparation all investigated spider silk components in aqueous solutions were in a conformational state comparable to that of natural silk proteins within the dope.

According to Uversky et al., intrinsically unfolded proteins can be predicted on the basis of their net charge and mean hydropathicity. The net charge of a protein is used to calculate a "boundary" hydropathicity. If the normalized mean hydropathicity of the protein is below the boundary value, the protein is predicted to be intrinsically unfolded (*33*, *34*). In accordance with our measured data, this is the case for the repetitive sequences $(QAQ)_8$ and $(AQ)_{12}$ (Table 1). Intrinsic unfolding of a protein indicates that interactions of the amino acid residues with the surrounding solvent are more favorable than with amino acids of the same or other polypeptide chains. Accordingly, $(QAQ)_8$ and $(AQ)_{12}$ are soluble even at high concentrations. In contrast, C_{16} displays a hydropathicity slightly above the boundary value. While still revealing some properties of intrinsically unfolded proteins, interactions between polypeptide chains are becoming more favorable at high concentrations, leading to aggregation of the protein and resulting in a lower solubility in comparison to those of $(QAQ)_8$ and $(AQ)_{12}$ (Table 1).

Mean hydropathicities of NR domains were clearly larger than their calculated boundary values, which is consistent with the data presented in this work, identifying them as folded protein domains. Solubility of folded proteins is determined only by amino acid residues exposed to the surface. Therefore, the mean hydrophobicity of NR domains cannot be used as a parameter to estimate their contribution to the solubility of spider silk proteins. However, since NR sequences constitute only a small fraction of spider silk proteins, they likely have little influence on the proteins' solubilities. Accordingly, rep-NR proteins displayed the same solubility properties as the corresponding rep-proteins, indicating that solubility is determined by the composition of the repetitive sequences only. The solubility and calculated hydropathicity of $(QAQ)_8$ and $(AQ)_{12}$ correlate well with the values of authentic ADF-3 (Table 1). C_{16} and ADF-4 both display lower solubility, although C_{16} does not share the high intrinsic insolubility of ADF-4. This difference can be explained by a higher hydropathicity and a lower net charge for ADF-4 (1.0%) than for C_{16} (2.9%) (Table 1).

Reducing negative charges by protonation of the silk proteins' carboxyl groups was expected to mainly affect proteins comprising the C module. Accordingly, proteins composed of modules A and Q, which do not contain aspartates or glutamates, did show only weak aggregation upon protonation. C_{16}, even after neutralization of its 16 negative charges, remained mostly soluble. Strikingly, fusing the NR4 domain, which did not show any response to acidification by itself, with the weakly aggregating C_{16} resulted in a protein highly sensitive to protonation. Thus, charge reduction of the repetitive region and the presence of the NR domain are required for initiating aggregation. The pH used in this experiment was chosen beyond conditions found during the natural spinning process to demonstrate the effect of protonation. Although NR4 did not show any measurable aggregation tendency, there is a possibility that some of its properties are altered at low pH in comparison to physiological conditions.

Phosphate, like other lyotropic ions, is known to increase the surface tension of water, promoting hydrophobic interactions (*35*). In the case of spider silk proteins, it is likely that the addition of phosphate initiates interactions between the hydrophobic polyalanine motifs, causing the aggregation of the proteins. Accordingly, aggregation of $(AQ)_{12}$ was pronounced in comparison to that of $(QAQ)_8$ which contains one-third fewer polyalanine motifs. C_{16} displaying the longest and greatest number of polyalanine motifs, however, did not show the strongest aggregation upon phosphate treatment. A possible explanation for this unexpected result might be the repulsion of negatively charged glutamate side chains and phosphate ions, leading to a weakening of the phosphate's lyotropic effect. Even though both NR domains did not respond to the addition of phosphate, their addition to the rep-proteins strongly increased phosphate sensitivity.

In the case of acidification as well as after addition of phosphate, it appeared that the NR domains amplified a primary response of the repetitive regions to aggregation-promoting factors. As the NR domains did not respond to these factors themselves, this enhancement of sensitivity might be caused, for instance, by changes in the oligomeric status of the silk proteins. NR domains have been found to form disulfide-bridged dimers (*36*). Further oligomerization might lead to increased local concentrations of polypeptide sequences required for initiating aggregation which is assisted by solvent conditions that favor the formation of intermolecular interactions.

Our protein engineering approach, which combines synthetic repetitive sequences with authentic NR regions, reveals that proteins closely resembling authentic silk proteins can be produced in high yields. The bacterial expression system as well as the simple and cheap purification process, which can easily be scaled up, provides the basis for cost-efficient industrial-scale production of spider silk-like proteins. On the basis of our studies, the molecular mechanisms of spider silk assembly can be further investigated, which will provide the knowledge required for artificially spinning silk threads from recombinant proteins and for gaining new materials for biotechnological and medical applications.

ACKNOWLEDGMENT

We thank Anja Mialki for technical assistance and John Gosline and Paul Guerette for kindly providing clones of *adf-3* and *adf-4*.

REFERENCES

1. Gosline, J. M., Guerette, P. A., Ortlepp, C. S., and Savage, K. N. (1999) The mechanical design of spider silks: from fibroin sequence to mechanical function, *J. Exp. Biol. 202* (Part 23), 3295−3303.
2. Vollrath, F., and Knight, D. P. (2001) Liquid crystalline spinning of spider silk, *Nature 410*, 541−548.
3. Guerette, P. A., Ginzinger, D. G., Weber, B. H., and Gosline, J. M. (1996) Silk properties determined by gland-specific expression of a spider fibroin gene family, *Science 272*, 112−115.
4. Gatesy, J., Hayashi, C., Motriuk, D., Woods, J., and Lewis, R. (2001) Extreme diversity, conservation, and convergence of spider silk fibroin sequences, *Science 291*, 2603−2605.
5. Simmons, A. H., 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.
6. 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.

7. van Beek, J. D., Hess, S., Vollrath, F., and Meier, B. H. (2002) The molecular structure of spider dragline silk: folding and orientation of the protein backbone, *Proc. Natl. Acad. Sci. U.S.A. 99*, 10266−10271.
8. Hijirida, D. H., Do, K. G., Michal, C., Wong, S., Zax, D., and Jelinski, L. W. (1996) ^{13}C NMR of *Nephila clavipes* major ampullate silk gland, *Biophys. J. 71*, 3442−3447.
9. Kenney, J. M., Knight, D., Wise, M. J., and Vollrath, F. (2002) Amyloidogenic nature of spider silk, *Eur. J. Biochem. 269*, 4159−4163.
10. Hronska, M., van Beek, J. D., Williamson, P. T., Vollrath, F., and Meier, B. H. (2004) NMR characterization of native liquid spider dragline silk from *Nephila edulis*, *Biomacromolecules 5*, 834−839.
11. Knight, D. P., and Vollrath, F. (2001) Changes in element composition along the spinning duct in a *Nephila* spider, *Naturwissenschaften 88*, 179−182.
12. Vollrath, F., Knight, D., and Hu, X. W. (1998) Silk production in a spider involves acid bath treatment, *Proc. R. Soc. London, Ser. B 265*, 817−820.
13. Tillinghast, E. K., Chase, S. F., and Townley, M. A. (1984) Water extraction by the major ampullate duct during silk formation in the spider, *Argiope aurantia* Lucas, *J. Insect Physiol. 30*, 591−596.
14. Knight, D. P., Knight, M. M., and Vollrath, F. (2000) Beta transition and stress-induced phase separation in the spinning of spider dragline silk, *Int. J. Biol. Macromol. 27*, 205−210.
15. Lazaris, A., Arcidiacono, S., Huang, Y., Zhou, J. F., Duguay, F., Chretien, N., Welsh, E. A., Soares, J. W., and Karatzas, C. N. (2002) Spider silk fibers spun from soluble recombinant silk produced in mammalian cells, *Science 295*, 472−476.
16. Sambrook, J., and Russell, D. (2001) *Molecular Cloning*, Cold Spring Harbor Laboratory Press, Plainview, NY.
17. Kroll, D. J., Abdel-Malek Abdel-Hafiz, H., Marcell, T., Simpson, S., Chen, C. Y., Gutierrez-Hartmann, A., Lustbader, J. W., and Hoeffler, J. P. (1993) A multifunctional prokaryotic protein expression system: overproduction, affinity purification, and selective detection, *DNA Cell Biol. 12*, 441−453.
18. Reiling, H. E., Laurila, H., and Fiechter, A. (1985) Mass-Culture of *Escherichia coli* Medium Development for Low and High-Density Cultivation of *Escherichia coli*-B/R in Minimal and Complex Media, *J. Biotechnol. 2*, 191−206.
19. Yee, L., and Blanch, H. W. (1992) Recombinant protein expression in high cell density fed-batch cultures of *Escherichia coli*, *Bio/Technology 10*, 1550−1556.
20. Gill, S. C., and von Hippel, P. H. (1989) Calculation of Protein Extinction Coefficients from Amino-Acid Sequence Data, *Anal. Biochem. 182*, 319−326.
21. Arcidiacono, S., Mello, C., Kaplan, D., Cheley, S., and Bayley, H. (1998) Purification and characterization of recombinant spider silk expressed in *Escherichia coli*, *Appl. Microbiol. Biotechnol. 49*, 31−38.
22. 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.
23. Fahnestock, S. R., and Irwin, S. L. (1997) Synthetic spider dragline silk proteins and their production in *Escherichia coli*, *Appl. Microbiol. Biotechnol. 47*, 23−32.
24. Lewis, R. V., Hinman, M., Kothakota, S., and Fournier, M. J. (1996) Expression and purification of a spider silk protein: a new strategy for producing repetitive proteins, *Protein Expression Purif. 7*, 400−406.
25. Scheller, J., Guhrs, K. H., Grosse, F., and Conrad, U. (2001) Production of spider silk proteins in tobacco and potato, *Nat. Biotechnol. 19*, 573−577.
26. Padgett, K. A., and Sorge, J. A. (1996) Creating seamless junctions independent of restriction sites in PCR cloning, *Gene 168*, 31−35.
27. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12, *Science 277*, 1453−1474.
28. Kim, T. D., Ryu, H. J., Cho, H. I., Yang, C. H., and Kim, J. (2000) Thermal behavior of proteins: heat-resistant proteins and their heat-induced secondary structural changes, *Biochemistry 39*, 14839−14846.
29. Uversky, V. N., Lee, H. J., Li, J., Fink, A. L., and Lee, S. J. (2001) Stabilization of partially folded conformation during α-synuclein oligomerization in both purified and cytosolic preparations, *J. Biol. Chem. 276*, 43495−43498.
30. Chen, X., Knight, D. P., and Vollrath, F. (2002) Rheological characterization of nephila spidroin solution, *Biomacromolecules 3*, 644−648.
31. Michal, C. A., Simmons, A. H., Chew, B. G., Zax, D. B., and Jelinski, L. W. (1996) Presence of phosphorus in *Nephila clavipes* dragline silk, *Biophys. J. 70*, 489−493.
32. Hoffmann, R., Reichert, I., Wachs, W. O., Zeppezauer, M., and Kalbitzer, H. R. (1994) ^{1}H and ^{31}P NMR spectroscopy of phosphorylated model peptides, *Int. J. Pept. Protein Res. 44*, 193−198.
33. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins 41*, 415−427.
34. Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics, *Protein Sci. 11*, 739−756.
35. Arakawa, T., and Timasheff, S. N. (1985) Theory of protein solubility, *Methods Enzymol. 114*, 49−77.
36. Sponner, A., Unger, E., Grosse, F., and Weisshart, K. (2004) Conserved C-termini of Spidroins are secreted by the major ampullate glands and retained in the silk thread, *Biomacromolecules 5*, 840−845.
37. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol. 157*, 105−132.

BI048983Q