Structural Properties of Recombinant Nonrepetitive and Repetitive Parts of Major Ampullate Spidroin 1 from *Euprosthenops australis*: Implications for Fiber Formation[†]

My Hedhammar,*[,‡] Anna Rising,[‡,§] Stefan Grip,[‡,§] Alejandra Saenz Martinez,[△] Kerstin Nordling,[‡] Cristina Casals,[△] Margareta Stark,[‡,§] and Jan Johansson[‡]

Department of Anatomy, Physiology and Biochemistry, Biomedical Centre, Swedish University of Agricultural Sciences, SE-751 23 Uppsala, Sweden, Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, SE-750 07 Uppsala, Sweden, and Departamento de Bioquímica y Biología Molecular I and CIBER (Centro Investigación Biomédica en Red) Enfermedades Respiratorias, Universidad Complutense de Madrid, 28040 Madrid, Spain

Received December 13, 2007; Revised Manuscript Received January 22, 2008

ABSTRACT: Spider dragline silk proteins, spidroins, have a tripartite composition; a nonrepetitive N-terminal domain, a central repetitive region built up from many iterated poly-Ala and Gly rich blocks, and a C-terminal nonrepetitive domain. It is generally believed that the repetitive region forms intermolecular contacts in the silk fibers, while precise functions of the terminal domains have not been established. Herein, thermal, pH, and salt effects on the structure and aggregation and/or polymerization of recombinant N- and C-terminal domains, a repetitive segment containing four poly-Ala and Gly rich coblocks, and combinations thereof were studied. The N- and C-terminal domains have mainly α-helical structure, and interestingly, both form homodimers. Dimerization of the end domains allows spidroin multimerization independent of the repetitive part. Reduction of an intersubunit disulfide in the C-terminal domain lowers the thermal stability but does not affect dimerization. The repetitive region shows helical secondary structure but appears to lack stable folded structure. A protein composed of this repetitive region linked to the C-terminal domain has a mainly α-helical folded structure but shows an abrupt transition to β-sheet structures upon heating. At room temperature, this protein self-assembles into macroscopic fibers within minutes. The secondary structures of none of the domains are altered by pH or salt. However, high concentrations of phosphate affect the tertiary structure and accelerate the aggregation propensity of the repetitive region. Implications of these results for dragline spidroin behavior in solution and silk fiber formation are discussed.

Spider dragline silk is among the toughest fibers known to man (*1*) and has attracted much attention in the pursuit of materials with good mechanical properties. Additional attractive features of dragline silk include a low density, biodegradability (*2*), and biocompatibility (*3*). This remarkable fiber is constituted of at least two similar proteins, major ampullate spidroin (MaSp)[1] 1 and MaSp2 (*4, 5*). MaSp1 is more abundant than MaSp2 in the dragline fiber; the ratio is estimated to be 3:2 (*6–8*). These spidroins are large, comprising 3000–4000 amino acids (*9*), and consist of nonrepetitive N- and C-terminal domains and between these an extensive repetitive part. The repetitive region contains stretches of alanine interrupted by glycine rich repeats. The dragline silk of *Euprosthenops* is one of the strongest and least extensible of all ragline silks examined so far (*10, 11*). These qualities might be due to the high alanine content (*12*). The conserved N-terminal domain of 130 amino acid residues is present also in flagelliform and cylindriform spider silk proteins (*13, 14*). Considering its high degree of conservation, the N-terminal domain is likely to have a crucial role. The C-terminal nonrepetitive region, likewise conserved, contains 100 amino acid residues, forms disulfide-connected dimers (*15*), and is crucial for fiber formation of recombinant partial spidroins (*16, 17*).

Dragline spidroins are produced in the major ampullate gland, located in the abdomen of the spider. The gland is composed of three dominant parts: tail, sac, and duct. Most of the protein for fiber formation is produced by secretory cells lining the interior of the tail (*11, 18*). In the lumen of the sac, the spidroins are stored in liquid form, termed the dope. The duct is 5 times longer than necessary to connect

[†] This study was supported by a European Commission grant ("Spiderman" Contract G5RD-CT-2002-00738) and the Swedish Research Council.

* To whom correspondence should be addressed. E-mail: my.hedhammar@afb.slu.se. Telephone: +46-18-4714191. Fax: +46-18-550762.

[‡] Department of Anatomy, Physiology and Biochemistry, Biomedical Centre, Swedish University of Agricultural Sciences.

[§] Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences.

[△] Universidad Complutense de Madrid.

[1] Abbreviations: 4Rep, repetitive region consisting of four poly-Ala and Gly rich units; 4RepCT, four repetitive units followed by the nonrepetitive C-terminal domain; CD, circular dichroism; CT, C-terminal domain; DSC, differential scanning calorimetry; HisTrxHis, dually His_6-tagged thioredoxin followed by a thrombin cleavage site; IMAC, immobilized metal ion affinity chromatography; MaSp, major ampullate spidroin; NT, N-terminal domain; SEC, size exclusion chromatography.

10.1021/bi702432y CCC: $40.75
Published on Web 02/23/2008

the sac with the spigot (*19*), narrows progressively, and has three limbs folded into an S shape (*20*). Along the duct, the conformation of the spidroins in the dope changes, and in the distal part, the liquid dope is converted into a solid fiber (*21*). How the spider manages to control the polymerization remains an intriguing question. When the spidroins are subjected to slow elongational flow in the gland, increasing sheer forces are likely to be generated as the duct narrows (*20*), which probably causes the proteins to align (*11, 22*). The conversion might also be facilitated by a drop in pH, caused by proton pumps in the distal part of the duct (*23*), and/or by changes in ion concentration (*24*). When the silk dope travels down the duct, both Na^+ and Cl^- concentrations decrease while the K^+ and phosphate concentrations increase (*24*).

A better understanding of fiber formation could guide experimental research to the development of superior fibrous materials. Moreover, the structure of the soluble spidroins attracts attention since the spider manages to keep this highly concentrated and stress-sensitive protein stored within the gland in a liquid state. Due to the territorial nature of spiders, any large-scale production of spider dragline silk would most likely require recombinant expression. The full-length spidroins are probably too large and repetitive to be expressed efficiently using presently available systems. Hence, parts of the repetitive region, the N- and/or C-terminal domains or combinations of these, have to be used. A recent study shows that a comparatively small repetitive unit fused to the C-terminal domain is sufficient, and necessary, for fiber formation (*17*). In this work, this fiber-forming unit, the N- and C-terminal nonrepetitive domains, and a unit representative for the repetitive regions only are characterized in terms of secondary, tertiary, and quaternary structure and stability. Moreover, fiber formation and aggregation propensities are analyzed under various conditions.

EXPERIMENTAL PROCEDURES

DNA Templates. A cDNA clone encoding a repetitive part and the C-terminal domain of MaSp1 from *Euprosthenops australis*, denoted clone 2 (EMBL databank accession number AJ973155) (*17*), and a cDNA clone encoding the N-terminal domain of MaSp1 from *E. australis*, denoted clone 55 (EMBL databank accession number AM259067) (*13*), were used as templates.

Construction of Expression Vectors. DNA sequences encoding the N-terminal (NT) or C-terminal (CT) domain of *E. australis* MaSp1 were amplified by PCR using *LA Taq* polymerase (TaKaRa Bio, Saint-Germain-en-Laye, France) and clone 55 or clone 2, respectively, as the template. The following primers were used: NTsense (5′-TAGAATTCACACACTACACCATGGACAAAC-3′) and NT*antisense (5′-TAAAGCTTACGCTCCTGCGGATGCTGCTGC-3′) or NTantisense (5′-TAAAGCTTAAGCACTGACATCATTCATACCTG-3′) or CTsense (5′-TTAGGATCCAAGGTTATGGTCAGAGTAGTGCTTCTGCT-3′) and CTantisense (5′-TATAAGCTTAGCCCATTACTTGAGCCATGGC-3′), respectively (restriction enzyme cleavage sites underlined). The resulting PCR products were restricted with the corresponding enzymes (*Eco*RI and *Hin*dIII or *Bam*HI and *Hin*dIII) and ligated into the pHisTrxHis vector, encoding a His_6 tag, thioredoxin, another His_6 tag, and finally a thrombin cleavage site (HisTrxHis). The sequences of the constructs were verified, and the vectors were denoted pHisTrxHisNT*, pHisTrxHisNT, and pHisTrxHisCT. The NT* construct includes a short Ser and Ala rich motif located just downstream of the nonrepetitive domain (Figure 1A,B). The CT construct includes a similar Ser and Ala rich motif and a part of the proceeding glycine rich repeat located upstream of the nonrepetitive domain (Figure 1A,B). The DNA sequence encoding four repetitive poly-Ala and Gly rich blocks (4Rep) was amplified by PCR using clone 2 as the template and the primers 4Repsense (5′-TTAGGATC-

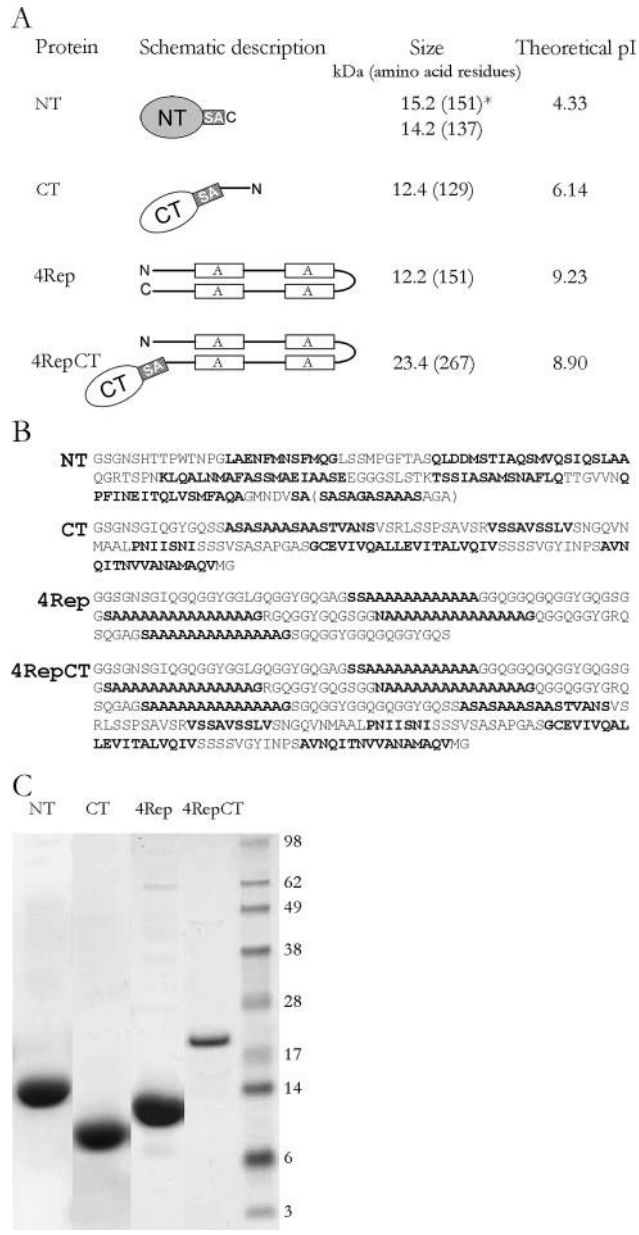

FIGURE 1: Description of the analyzed constructs. (A) Characteristics of recombinant *E. australis* MaSp1 proteins. White boxes represent polyalanine stretches, and black lines represent glycine rich regions. Gray boxes represent regions rich in serine and alanine. (B) Predictions of the secondary structures of the different parts. Segments predicted to be in α-helical conformations are shown in bold. Only segments with at least four consecutive residues in an α-helix conformation are shown. For NT, the segment in parentheses is present only in the NT* construct. (C) SDS−PAGE of purified NT, CT, 4Rep, and 4RepCT, as labeled above the lanes. Migration of low-molecular mass markers (kilodaltons) is shown at the right.

CAAGGTCAAGGTGGATATGGTGGACTA-3′) and 4Repantisense (5′-TATCTCGAGTTAACTCTGACCATAACCTCCTTGACCTTG-3′). The resulting PCR product was restricted with the corresponding enzymes (*Bam*HI and *Xho*I) and ligated into pHisTrxHis. The sequence of the construct was verified, and the vector was denoted pHisTrxHis4Rep. A DNA fragment encoding four repetitive poly-Ala and Gly rich blocks followed by the C-terminal domain (4RepCT) was cleaved from the pET32TrxHisS4RepCT vector (*17*) using the enzymes *Bam*HI and *Hin*dIII and ligated into pHisTrxHis, giving the pHisTrxHis4RepCT vector.

Expression of Recombinant MaSp1 Proteins. The different MaSp1 constructs were used to transform *Escherichia coli* BL21(DE3) cells (Merck Biosciences). The cells were grown at 30 °C in Luria-Bertani medium containing kanamycin to an OD_{600} of 0.8−1 and then induced with isopropyl β-D-thiogalactopyranoside and further incubated for up to 4 h at room temperature. Thereafter, cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0) supplemented with lysozyme and DNase I. After complete lysis, the 15000*g* supernatants were loaded onto a column packed with Ni-NTA Sepharose (GE Healthcare, Uppsala, Sweden). The column was washed extensively before bound proteins were eluted with 300 mM imidazole. Fractions containing the target proteins were pooled and dialyzed against 20 mM Tris-HCl (pH 8.0). MaSp1 proteins were released from the tags by proteolytic cleavage using a thrombin:fusion protein ratio of 1:1000 (w/w) at room temperature for 1–2 h. To remove the released HisTrxHis tag, the cleavage mixture was loaded onto a second Ni-NTA Sepharose column and the flow-through was collected. Protein samples were separated via SDS−PAGE and then stained with Coomassie Brilliant Blue R-250. The proteins were concentrated by ultrafiltration using a 5 kDa molecular mass cutoff cellulose filter (Millipore). The protein content was determined from the absorbance at 280 nm.

Size Exclusion Chromatography (SEC). The quaternary structures of NT and CT were analyzed using a Sephacryl S-100 HR (GE Healthcare) column (1.0 cm inside diameter, 37 cm bed height) equilibrated with 20 mM Tris (pH 7.4). The experiments were performed at room temperature using an ÄKTA basic system (GE Healthcare) at a flow rate of 0.5 mL/min. For analysis of reduced CT, it was incubated for 1 h in the presence of 2 mM DTT before being loaded onto a column equilibrated with 20 mM Tris and 0.3 mM DTT (pH 7.4). Complete reduction was confirmed by SDS-PAGE.

Circular Dichroism (CD) Spectroscopy. CD spectra from 250 to 190 nm and from 350 to 250 nm were recorded in 0.1 and 1.0 cm, respectively, path length quartz cuvettes using a J-810 spectropolarimeter (Jasco, Tokyo, Japan). The scan speed was 50 nm/min and the response time 2 s with an acquisition interval and bandwidth of 1 nm. The following buffers were used: 20 mM Tris-HCl (pH 8, 8.5, and 9), 20 mM sodium phosphate (pH 6, 6.3, 6.9, 7, 7.2, and 8), 20 mM sodium acetate (pH 4 and 5), and glycine-HCl (pH 2 and 3). Typically, protein concentrations of 20–30 μM were used. To study the effects of different salts on the conformation, potassium chloride, sodium chloride, potassium phosphate, or sodium phosphate was added to a final concentration of 100 or 300 mM (pH 6.3 and 7.2). Temperature scans from 20 to 90 °C were performed in 20 mM sodium phosphate buffer (pH 7.4) using a cuvette equipped with a Teflon lid. CD analysis of 4Rep and 4RepCT was done directly in the purification buffer [20 mM Tris-HCl (pH 8)], since these proteins have a strong tendency to aggregate over time, which led to protein loss during buffer exchange. The spectra obtained in Tris buffer were similar to those of the proteins at lower concentrations (8–12 μM) in phosphate buffer, although the Tris buffer was expected to shift to pH 6 during the thermal scan. The rate of temperature elevation was 0.5 °C/min. After the 90 °C scan, the samples were cooled to 20 °C, equilibrated for 15 min, and scanned again, to investigate the reversibility of the temperature-induced conformational changes. For thermal unfolding curves, the residual molar ellipticities at 208 and 275 nm were used, because the changes in signal were found to be most pronounced at these wavelengths.

Evaluation of Far-UV CD Spectra and Secondary Structure Prediction. Analysis of CD spectra for estimating secondary structure fractions was done using CDpro and the method CDSSTR (*25, 26*) using reference data set number 4, composed of 43 soluble proteins. The values given for helix and sheet contents are the sum of regular and distorted structures, respectively.

The secondary structure predictions shown were obtained using the hierarchical neural network method (HNN) (*27*). The neural network methods SSpro 4.5 and SSpro8 at the SCRATCH server (*28, 29*) displayed similar results. All secondary structure prediction programs used are available at the Expasy proteomic tools website.

Differential Scanning Calorimetry (DSC). Calorimetric measurements were performed in a Microcal VP differential scanning calorimeter (Microcal Inc., Northampton, MA) at a heating rate of 0.5 °C/min. Proteins (30–40 μM) dissolved in degassed 20 mM phosphate buffer (pH 7.4) were loaded in the sample cell, and degassed buffer was loaded in the reference cell. Calorimetric scans were collected for each sample between 10 and 100 °C. The reversibility of the thermal transition was evaluated by several cycles of heating and cooling. Standard Microcal Origin was used for data acquisition and analysis. The excess heat capacity functions were obtained after subtraction of the buffer baseline.

Turbidimetry. Proteins (30 μM) were analyzed at room temperature at different pHs (6.3 and 7.2) and different concentrations (0, 100, and 300 mM) of potassium chloride, sodium chloride, potassium phosphate, or sodium phosphate. At each time point, the absorbance at 340 nm was measured after gentle vortexing. Correlation between turbidity and aggregation was confirmed by SDS−PAGE of soluble and nonsoluble protein after 16000*g* centrifugation and by visual inspection of the relative amounts of pelleted protein.

Fiber Formation. Proteins (30 μM) were allowed to self-assemble into fibers as previously described (*17*) at different pHs (6.3 and 7.2) and different concentrations (0, 100, and 300 mM) of potassium chloride, sodium chloride, potassium phosphate, or sodium phosphate.

RESULTS

Schematic descriptions, sizes, calculated pI values, and secondary structure predictions for the different constructs of *E. australis* MaSp1 produced recombinantly are shown in Figure 1. NT and CT represent the nonrepetitive N- and

Table 1: Recoveries of Purified Recombinant Spider Silk Proteins

part	purified fusion protein (mg/L of culture[a])	purified target protein (mg/L of culture[a])
NT	150	70
CT	35	10
4Rep	40	15
4RepCT	25	10

[a] The amount of purified protein recovered from a 1 L shake flask culture with an OD of ~1.

Table 2: Results from Deconvolution of the CD Spectra Collected for the Different MaSp1 Constructs[a]

unit	helix (%)	sheet (%)	turn (%)	unordered (%)
NT at 20 °C	**54**	12	11	23
NT at 90 °C	6	**35**	25	**34**
4Rep at 20 °C	**33**	12	22	**33**
4Rep at 90 °C	7	**37**	24	**33**
CT at 20 °C	**57**	8	12	23
CT at 90 °C	8	**34**	25	**32**
4RepCT at 20 °C	**45**	12	21	22
4RepCT at 90 °C	4	**41**	23	32

[a] The value(s) of the dominant secondary structure(s) is in boldface.

Table 3: Melting Temperatures (T_m) for the Different MaSp1 Constructs Measured by Different Techniques

part	T_m measured by far-UV CD (°C)	T_m measured by near-UV CD (°C)	T_m measured by DSC (°C)
NT	60	54	58
CT	68 (reduced), 72 (nonreduced)	ND[a]	66 (nonreduced)
4RepCT	47	ND[a]	ND[a]

[a] Not determined.

C-terminal domains, respectively, and 4Rep is a repetitive region consisting of four poly-Ala and Gly rich units. 4RepCT is a combination of 4Rep and CT.

Expression and Purification of Recombinant MaSp1 Proteins. The MaSp1 parts were expressed as HisTrxHis fusion proteins and purified under nondenaturing conditions using immobilized metal ion affinity chromatography (IMAC). Proteolytic cleavage and isolation from the released HisTrx-His tag by passage through a second IMAC column yielded >95% pure proteins (Figure 1C). The recoveries of the different proteins are summarized in Table 1.

Structural Analysis and Temperature-Induced Conformational Changes. The different MaSp1 parts were investigated by near- and far-UV CD spectroscopy, DSC, and SEC. Results from deconvolution of the far-UV CD spectra are summarized in Table 2. Melting temperatures are summarized in Table 3.

(1) N-Terminal Domain. An initial construct representing the N-terminal domain (NT*, Figure 1) was found to undergo truncation over time. Mass spectrometry showed the main cleavage site to be SA↓SASA in the C-terminal part of the protein (data not shown; see Figure 1). Therefore, a second construct (NT), lacking the cleaved part, was generated. NT* and NT exhibited identical CD spectra at 20 °C, indicative of a helix-dominated secondary structure (Figure 2A). By deconvolution, the α-helical content was estimated to be 54% (Table 2). Upon being heated, NT* adopted mainly β-sheet structure at 90 °C, and this conformational change was irreversible (Figure 2A). NT, however, showed a transition to less ordered structure between 50 and 70 °C (Figure 2C and Table 3), compatible with cooperative unfolding, and showed mainly unordered structure at 90 °C (Figure 2A). After cooling to 20 °C, NT gave a spectrum very similar to the one before heating, but with a slightly lower amplitude (Figure 2A).

Near-UV CD spectra and thermal unfolding of NT (Figure 2B,C) showed that the tertiary structure is cooperatively unfolded with a T_m that is somewhat lower than that for melting of the secondary structure (Table 3). This unfolding of tertiary structure is not completely reversible, as the near-UV spectrum after cooling differs in shape.

DSC of NT gave a T_m of 58 °C and revealed that the unfolding is not reversible after heating to 100 °C (Figure 2D and Table 3).

SEC showed that NT migrates as a dimer (Figure 2E).

(2) C-Terminal Domain. CD spectra of CT indicated a predominantly α-helical structure and gave 57% helix content from deconvolution (Figure 3A and Table 2). Above 50 °C, the molar ellipticity started to decrease but CT displayed a typical α-helical pattern up to 80 °C. The spectrum collected at 90 °C indicated a β-sheet structure, which was kept upon cooling to 20 °C. After disulfide reduction, CT was more sensitive to thermal unfolding, as the T_m was lowered, compared to that under nonreducing conditions (Figure 3B and Table 3). SEC showed that CT exists as a dimer under both reducing and nonreducing conditions (Figure 2E).

DSC of nonreduced CT gave a T_m of 66 °C and showed that the unfolding is not reversible, as repeated scans gave no signal (Figure 3C and Table 3). DSC showed a broad heat transition, indicating a lack of cooperative unfolding. This is in line with the gradual melting of nonreduced CT seen by CD (Figure 3B).

Near-UV CD measurements of CT gave very weak signals, as expected from the low content of aromatic residues (see Figure 1).

(3) Repetitive Region. The CD spectrum of 4Rep displayed a rather low residual molar ellipticity (Figure 4A). A positive band at 190 nm and two minima at 205 and 219 nm gave a signature that is blue-shifted a few nanometers from the canonical α-helix spectrum. Above 20 °C, the CD signal gradually decreased in intensity without a change in shape (Figure 4C). Spectra above 35 °C showed a single minimum at 211 nm, implying conversion into a β-sheet-like state. The conformational change was irreversible and resulted in precipitated protein.

(4) Fiber-Forming Unit. CD analysis of 4RepCT gave two negative peaks at 206 and 220 nm (Figure 4B). This spectrum is almost identical, compared to a combination of the individual spectra of 4Rep and CT (Figure 4D). Upon being heated, 4RepCT was stable up to 40 °C and then abruptly underwent a structural conversion leading to a β-sheet state, with a T_m of 47 °C (Figure 4C). The weak CD signal after heating indicates that most of the protein has gone out of solution, which was corroborated by the visual appearance of small fibers in the cuvette.

Effects of pH and Salt on Conformation, Aggregation, and Fiber Formation. To investigate whether secondary structures of the different parts of *E. australis* MaSp1 are sensitive to pH changes, they were analyzed by CD between pH 4 and 9, using steps of 1 pH unit. No structural changes could be detected for any of the proteins (data not shown).

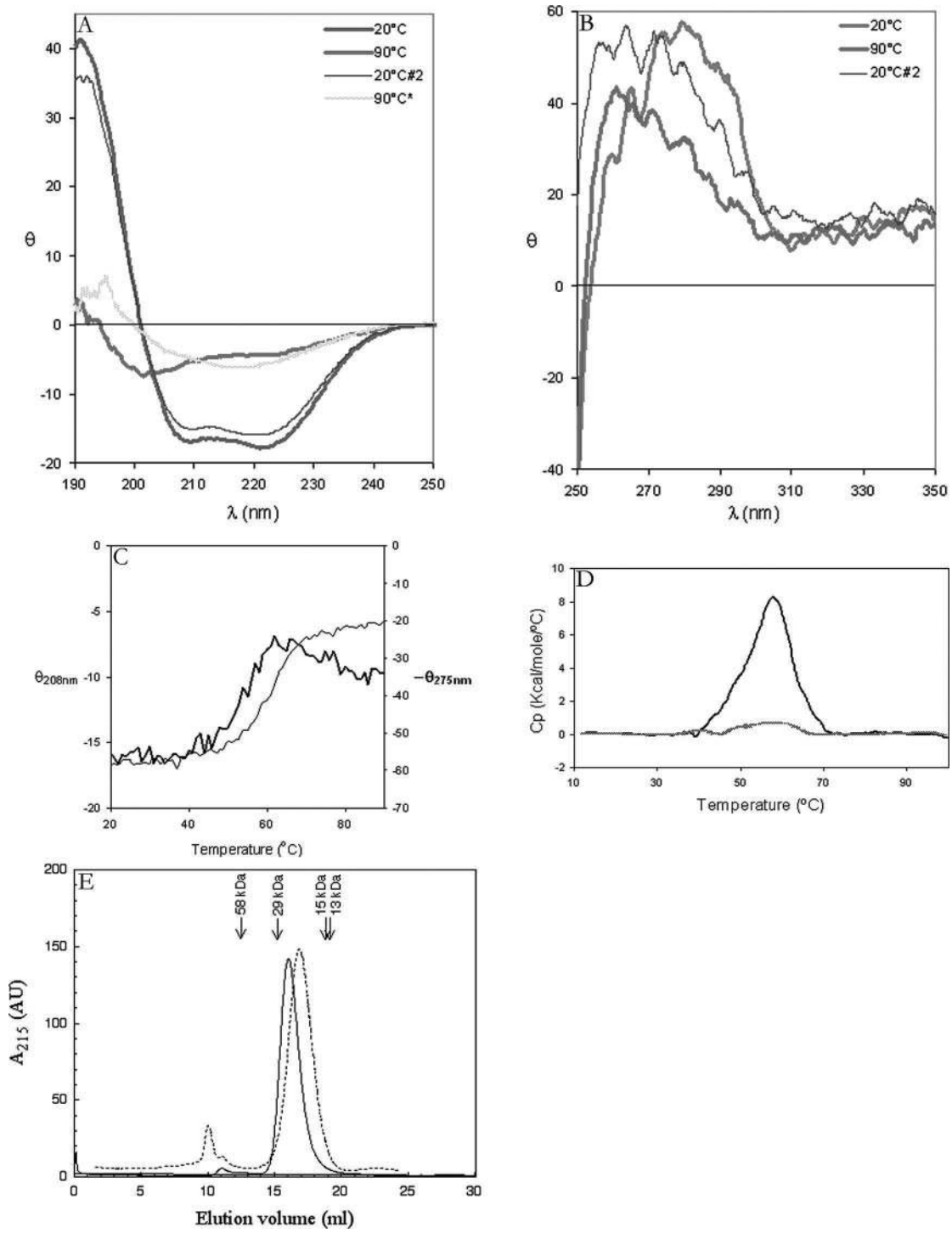

FIGURE 2: Thermal unfolding of NT. Far-UV (A) and near-UV (B) CD spectra of NT at 20 and 90 °C. The #2 spectra were collected after cooling to 20 °C. In panel A, a spectrum of NT* at 90 °C is included (90 °C*). (C) Residual molar ellipticities of NT at 208 nm (thin line) and 275 nm (thick line) vs temperature. The residual molar ellipticity (θ) is expressed in kilodegrees square centimeter per decimole. (D) Calorimetric scans of NT between 10 and 100 °C. The initial scan is shown in black and a repeated scan after cooling in gray. (E) Size exclusion chromatography of NT (solid line) and CT (dotted line). The arrows mark elution volumes of proteins with indicated masses.

Likewise, far-UV CD spectra for all MaSp1 parts remained unchanged at pH 6.3 and 7.2 in the presence of potassium chloride, sodium chloride, potassium phosphate, or sodium phosphate at a concentration of up to 300 mM (data not shown). The only observed effect was that 4Rep started to precipitate during prolonged incubation in 300 mM phosphate buffer.

To investigate the aggregation propensities, all proteins were incubated at low (100 mM) and high (300 mM) salt concentrations at pH 6.3 and 7.2 and the absorbance at 340 nm, indicative of turbidity due to aggregation, was measured over time. NT or CT did not aggregate to any measurable extent within 24 h under any of the conditions. In contrast, the aggregation rate of 4Rep and 4RepCT was accelerated in the presence of high concentrations of phosphate (as both sodium and potassium salts) (Figure 5A,B). SDS−PAGE analysis confirmed a decrease in the level of soluble protein and an increase in the level of nonsoluble protein (data not shown) that correlated well with the increase in turbidity.

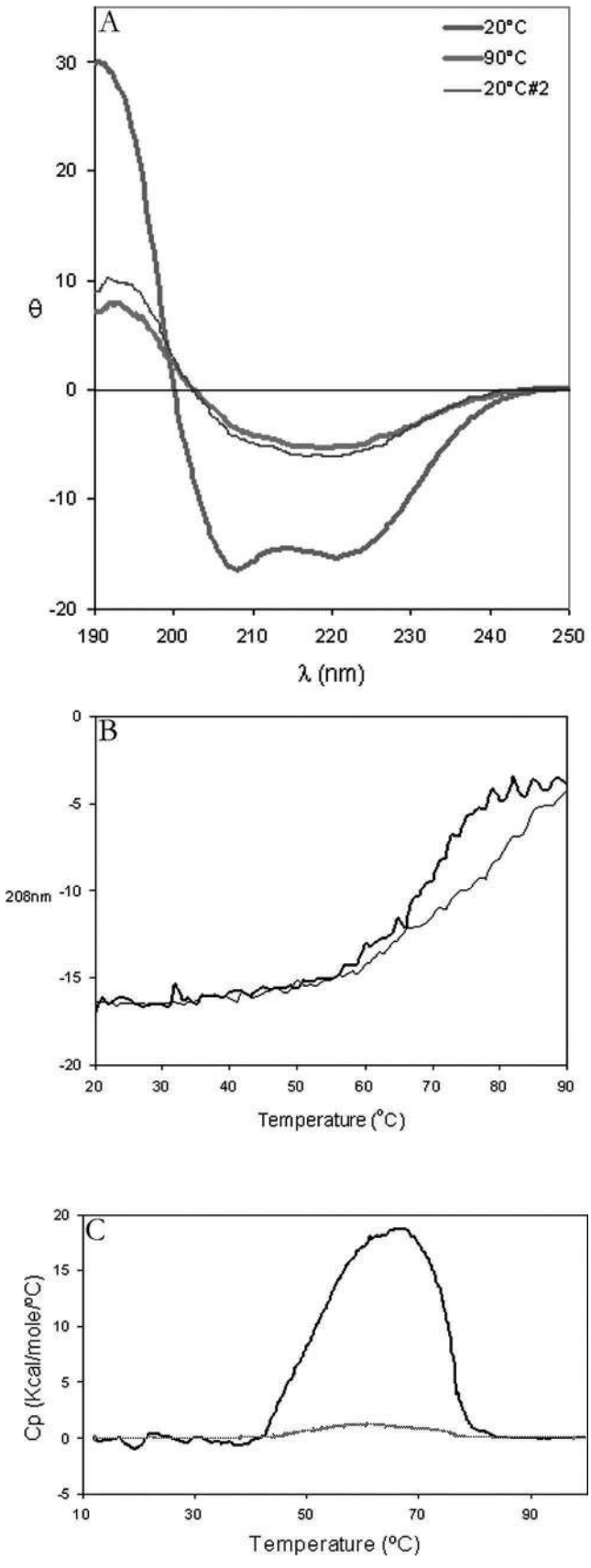

FIGURE 3: Secondary structure and thermal unfolding of CT. (A) Far-UV CD spectra of CT at 20 and 90 °C and after CT had cooled to 20 °C (#2). (B) Residual molar ellipticities at 208 vs temperature, under reducing (thick line) and nonreducing (thin line) conditions. The residual molar ellipticity (θ) is expressed in kilodegrees square centimeters per decimole. (C) Calorimetric scans of CT between 10 and 100 °C. The initial scan is shown in black and a repeated scan after cooling in gray.

High concentrations of the corresponding chloride salts did not induce aggregation. Effects on the tertiary structure were analyzed using near-UV CD spectroscopy. After incubation in the presence of high concentrations of phosphate, the spectrum of 4Rep in particular was changed (Figure 5C,D).

4Rep aggregates after gentle wagging overnight. However, a few small fibers could also be detected (Figure 6A). The partitioning between aggregates and fibers was not visibly affected by pH or salt composition or concentration. Interestingly, CT formed fibers within hours, although they were smaller than those made of 4RepCT (Figure 6B,C). Under all tested conditions, formation of fibers of 4RepCT was visible after 10–20 min and the macroscopic appearance of the final fiber was the same (Figure 6C,D).

4Rep, CT, and 4RepCT were highly sensitive to sheer forces and tended to aggregate or form short fibers during handling, e.g., pipetting, even at low protein concentrations, typically 1 mg/mL. NT on the other hand was stable for weeks at room temperature. Via ultrafiltration, NT could be concentrated to at least 216 mg/mL. The protein solution was then yellow and viscous and could be stored at 4 °C for days without aggregating.

DISCUSSION

The conformation, stability, and behavior of dragline spidroins before and during fiber formation remain poorly understood. Several issues complicate the characterization of native soluble spidroins. In the major ampullate gland, the silk dope is highly concentrated (~30–50 wt %) (*30, 31*) and extremely sensitive to sheer stress, drying, and hydration (*31*). Solid state NMR data of lyophilized gland dope gave no evidence of β-sheet structures (*31*), and solution state NMR, Fourier transform infrared, and CD data suggested that the spidroins in the gland exist in a highly metastable state of dynamically averaged helical conformations (*31*). A more recent NMR spectroscopic study of native silk dope suggested a dynamically disordered structure, although other secondary structures having similar chemical shifts could not be excluded (*32*). CD analysis of proteins freshly extracted from the tail, proximal, and middle part of the silk gland indicated a structure poor in α-helix and β-sheet, while proteins extracted from the distal part of the gland suggested β-sheet structures (*33*). Together, these results indicate a polymorphic and flexible nature of the dope. This approach, employing recombinant expression of representative parts of *E. australis* MaSp1 and purification under nondenaturing conditions, allowed, for the first time, structural and stability studies of all individual domains of a spidroin in soluble form.

In particular, the nonrepetitive N-terminal domain has not been expressed or characterized previously, apart from an initial CD spectrum (*13*). Secondary structure prediction of NT yields five α-helices (Figure 1B). The CD data for NT reveal a predominantly α-helical structure, with an estimated helix content of 54% (Figure 2 and Table 2). The structure is largely maintained up to 45 °C, but at higher temperatures, the tertiary structure is melted, followed shortly by a loss of secondary structure. The secondary structure refolds after the sample is heated to 90 °C, although the tertiary structure seems not to re-form completely. However, if a short Ser and Ala rich segment is present C-terminally, as in NT*,

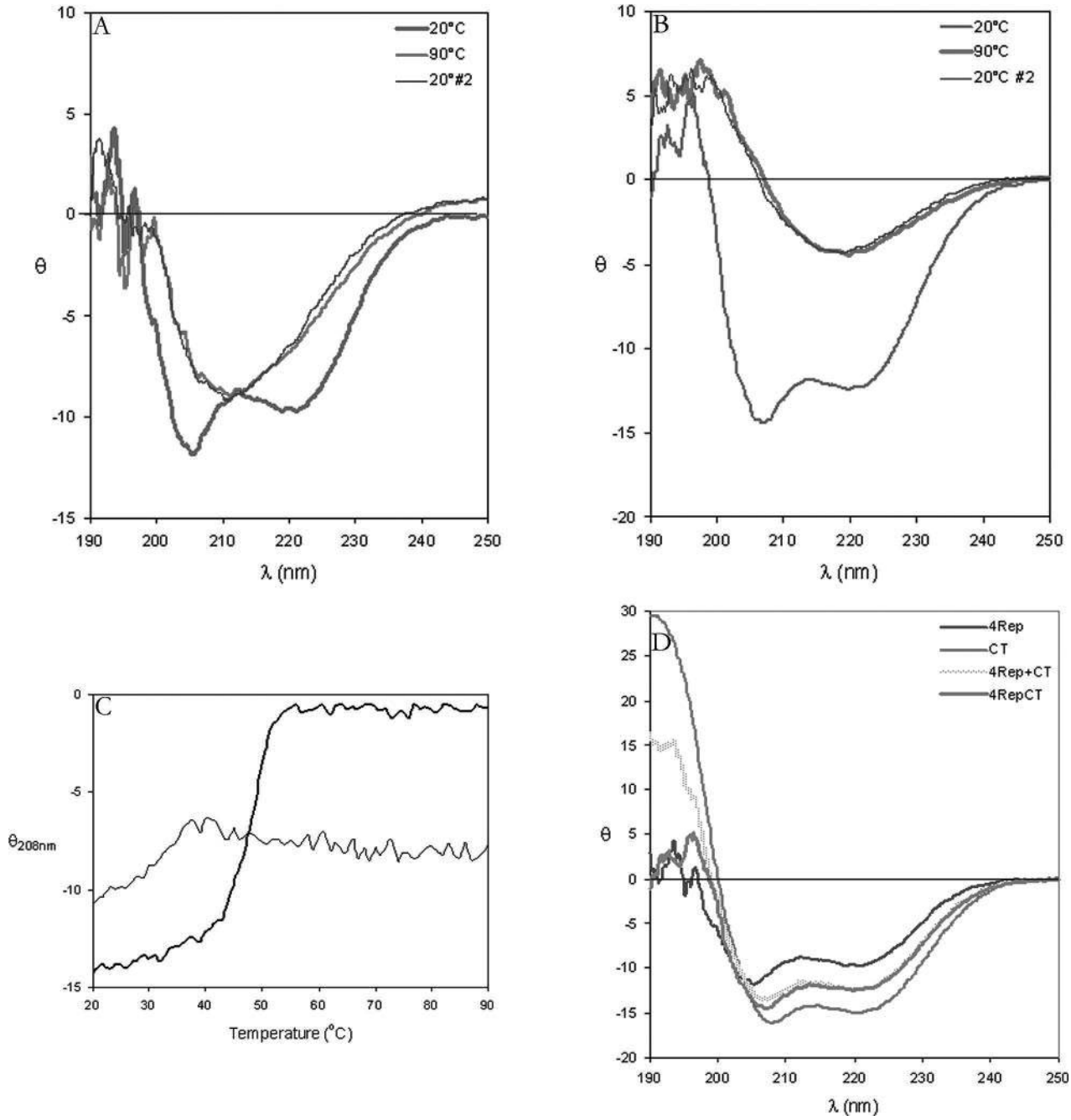

FIGURE 4: Secondary structure and thermal unfolding of 4Rep and 4RepCT. Far-UV CD spectra of 4Rep (A) and 4RepCT (B) at 20 and 90 °C and after CT had cooled to 20 °C (#2). (C) Residual molar ellipticities at 208 nm vs temperature for 4Rep (thin line) and 4RepCT (thick line). (D) Comparison between far-UV CD spectra of 4RepCT (green line), 4Rep (blue line), CT (red line), and the calculated combination of 4Rep and CT (dotted green line). The residual molar ellipticity (θ) is expressed in kilodegrees square centimeters per decimole.

the helical structure is destabilized above 50 °C and changes irreversibly to β-structure at 80 °C. This may reflect intermolecular interactions mediated by the Ser and Ala rich segment.

The recovery of soluble NT fused to HisTrxHis is remarkably high, and much higher than for the corresponding CT fusion protein (Table 1). NT is also stable for weeks at ambient temperatures and not affected by changes in pH or high salt concentration. Moreover, NT could be concentrated to extremely high levels (>216 mg/mL) without forming aggregates. The combined stability and solubility of the N-terminal region suggests a role as a solubility enhancing domain for spidroins. Whether such a function is important mainly during spidroin synthesis, secretion, storage, and/or fiber formation remains to be investigated. The C-terminal domain is present in the major ampullate gland extracts as well as in dragline silk fibers (*15*). However, it is notable that N-terminal sequence analysis of spider silk revealed sequences from only the repetitive region and not from the N-terminal domain (*7*) and that silk-like fibers can be formed in vitro from recombinant spidroins lacking the N-terminal domain (*17*).

CT is predominantly α-helical; secondary structure prediction yields five α-helices (53% of the protein), and CD data show 57% helix (Figures 1B and 3A and Table 2). The last four of the predicted α-helices are located within the

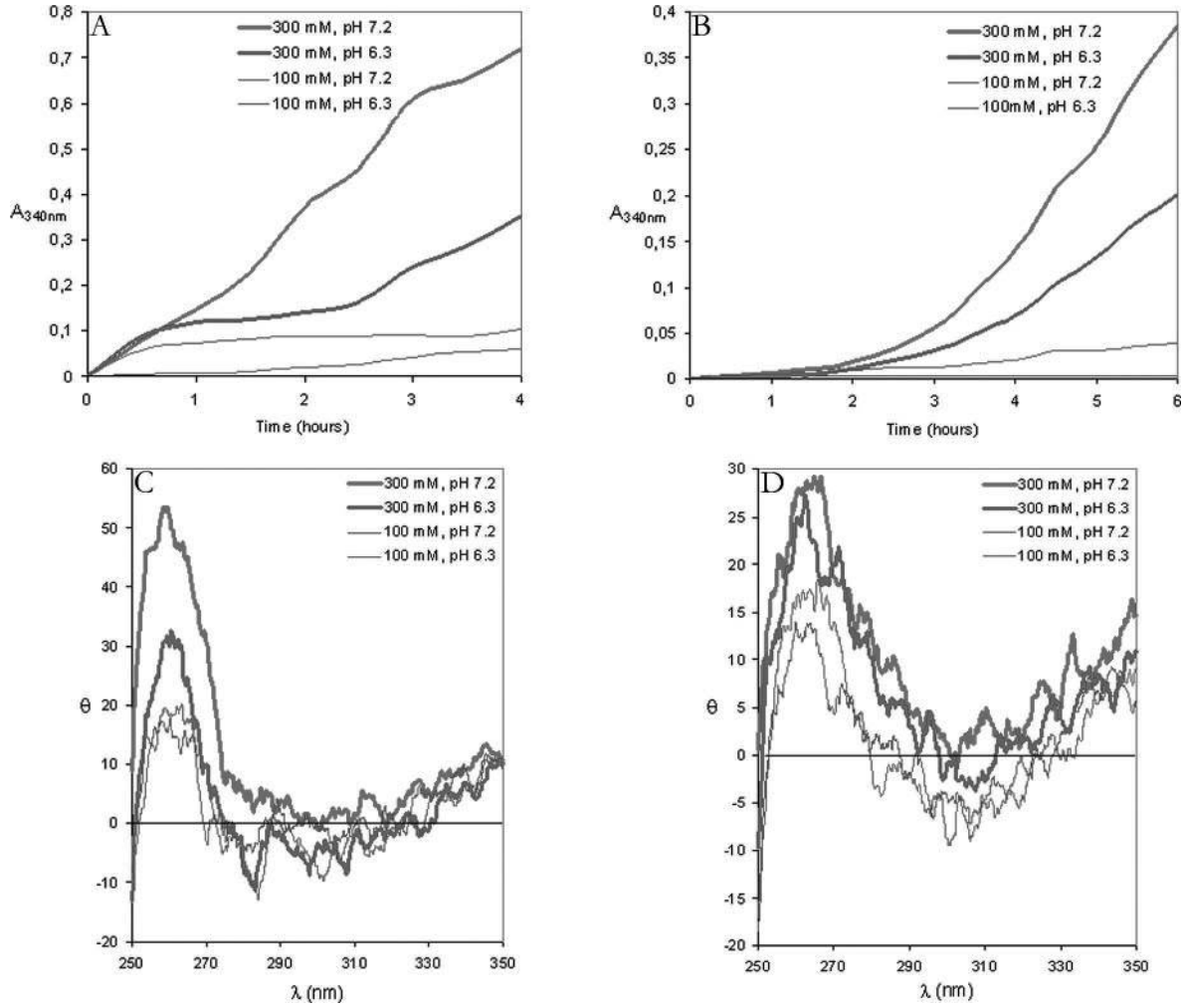

FIGURE 5: Effects of phosphate on 4Rep and 4RepCT. Turbidity of 4Rep (A) and 4RepCT (B) measured as the absorbance at 340 nm vs time at different pH values and phosphate concentrations. Near-UV CD spectra of 4Rep after incubation for 3 h (C) and 4RepCT after incubation for 5 h (D) at the indicated concentrations of phosphate buffers. The residual molar ellipticity (θ) is expressed in kilodegrees square centimeters per decimole.

conserved, nonrepetitive part of CT. The helicity now found for *E. australis* MaSp1 CT is consistent with studies of recombinant nonrepetitive C-terminal regions from *Araneus diadematus* dragline silk proteins ADF-3 and ADF-4 (*34*). The α-helical conformation of CT gradually melts above 50 °C and changes irreversibly to β-sheet structures above 80 °C (Figure 3). This is partly contradictory to previous studies, where the thermal transitions of recombinant proteins containing nonrepetitive C-terminal regions from ADF-3 and ADF-4, both with and without a repetitive part, were found to be completely reversible (*34*). The discrepancies could be due to differences in polypeptide sequence. However, this could also be due to differences in protein concentration and/or recombinant production strategy; e.g., a heat denaturation step was used for recovery of the ADF proteins, but not for CT or 4RepCT.

Native dragline silk spidroins form dimers under denaturing and nonreducing conditions (*15, 35*) which is most likely in part caused by formation of a disulfide bridge by a Cys that proceeds a conserved segment in the C-terminal domain. This segment has been proposed to form an amphipathic helix (*15, 36*), which is in line with the prediction shown in Figure 1B; the segment in question corresponds to the second last helix. We note that the Cys is located at the hydrophobic surface of the predicted amphipathic helix, which suggests that dimerization of the C-terminal domain is driven primarily by hydrophobic interactions and that formation of the disulfide then locks the dimer covalently. Consistent with this idea, dimerization was now observed for constructs containing CT, as judged by nonreducing SDS−PAGE (data not shown). Moreover, CT migrates as a dimer during SEC under both reducing and nonreducing conditions (Figure 2E). The differences in thermal unfolding of CT under reducing and nonreducing conditions show that disulfide formation has a stabilizing effect. However, this effect is comparatively small since no structural changes are seen at 20 °C, and also since reduced CT is thermally stable up to 50 °C (Figure 3). Disulfide-dependent dimerization may not be crucial for silk fiber formation, since, for example, most minor ampullate spidroins lack Cys in their C-terminal region (*15*). In line with this, 4RepCT allows fiber formation also under reducing conditions (data not shown). A Cys → Ser mutated recombinant ADF-4 protein, expressed using a baculovirus system, has also been shown to form interacellular nanofibers with similarities with the nonmutated variant (*36*). It is notable that not only CT but also NT migrates as a dimer during size exclusion chromatography (Figure 2E). How intermolecular interactions provided by the dual dimerization sites of the spidroins contribute to fiber formation needs to be further investigated. It is likely that the presence of domains

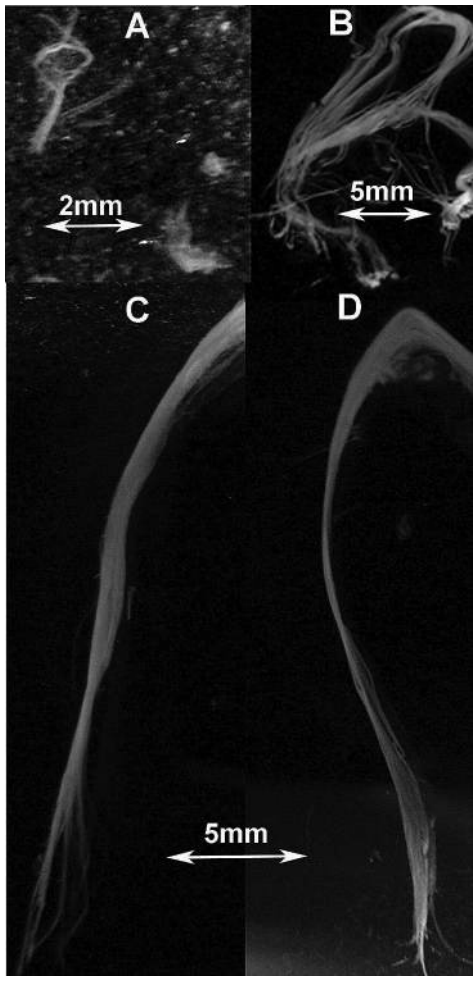

FIGURE 6: Fibers formed by different MaSp1 parts. Small fibers made of 4Rep (A) and CT (B) in 20 mM BisTris (pH 7.2). Fibers made from 4RepCT in 20 mM BisTris (pH 7.2) (C) and in the same buffer supplemented with 500 mM potassium phosphate (D).

prone to dimerize in each terminus promotes end domain-connected multimers.

Interactions between the N- and C-terminal domain have been suggested to be important for silk assembly (*37*). However, CD spectra of a 1:1 molar mixture of NT and CT were indistinguishable from that of the calculated combination of the two individual spectra, suggesting that no NT−CT interaction that leads to conformational change takes place (data not shown). Moreover, NT loaded onto a column containing bound HisTrxHis-4RepCT did not retard the reaction, further suggesting that NT and CT do not interact strongly.

The CD signature of 4Rep is blue-shifted a few nanometers from the canonical α-helix spectrum and shows a 222 nm: 208 nm ratio of 0.8. This could imply atypical helical structure, possibly 3_{10}-helix (*38*), or contributions from random coil or 3_1 structures (*39*). The rather high Tyr content in the protein may also contribute to this pattern (*40*). Previous investigations of a recombinantly expressed repetitive region of *A. diadematus* dragline silk proteins ADF-3 and ADF-4 showed spectra typical for unstructured proteins (*34*). These proteins contain shorter alanine stretches (six to eight residues) which might not be able to form as stable helices as the 12–15-residue poly-Ala segments of 4Rep. CD spectra of 4Rep display rather low ellipticities (Figure 4A) which implies a marginally stable fold. Therefore, the content of secondary structure elements obtained by deconvolution (Table 2) may be an overestimation. When the sample is heated, the CD signal gradually decreases in intensity without an accompanying change in shape. This is likewise compatible with melting of a structure that lacks a stable fold. However, already upon being heated above 35 °C, 4Rep undergoes conversion to a β-sheet-like structure. Even at room temperature, 4Rep is unstable and aggregates.

4RepCT, like 4Rep, displays a low molar ellipticity (Figure 4B). The CD spectrum of 4RepCT is almost identical in shape compared to a calculated combination of the spectra of 4Rep and CT, suggesting that the two parts do not interact in 4RepCT. As opposed to 4Rep, 4RepCT is stable to 40 °C, before it abruptly undergoes a structural conversion leading to a β-sheet state. This heat-induced conversion is accompanied by generation of short macroscopic fibers. These data show that CT covalently linked to 4Rep favors ordered polymerization over unordered aggregation of the repetitive region. We hypothesize that this is a function of the C-terminal domain also in native spidroins. CT is obviously essential for efficient fiber formation, as 4Rep without CT forms only tiny fibers, and CT can even form short fibers on its own (Figure 6). However, fiber formation for CT may require the presence of the Ser and Ala rich segment (Figure 1). Likewise, through comparison of NT and NT*, as well as of CT and 4RepCT, it seems likely that the repetitive region is prone to inducing a structural change to β-sheet structure.

Recently, we showed that, under physiological conditions, 4RepCT spontaneously forms macroscopic fibers that resemble native spider silk (*17*). This is the first recombinant spidroin reported to retain the property to spontaneously form fibers in vitro in a physiological buffer and under ambient conditions. Moreover, this is the first recombinant spidroin that is isolated under nondenaturing conditions. Taken together, this implies that the solution structure of the spidroin is important for fiber formation. The CD signature of the soluble state of 4RepCT shows similarities with spectra of freshly prepared dope (*33*). However, other studies, where the gland dope is blotted and redissolved in water before analysis, indicate a predominantly disordered structure (*41–43*). From our experience with 4RepCT, it appears likely that blotting of silk dope results in formation of spidroin polymers which are not soluble in water. 4RepCT is highly polymerogenic, and within minutes, macroscopic fibers are spontaneously formed. A CD spectrum of spontaneously formed 4RepCT fibers indicates a substantial fraction of β-sheet structure (*17*). This confirms that the fibers are a result of ordered polymerization.

Several environmental factors, e.g., acidification and changes in ion concentration, have been suggested to promote silk assembly (*24, 30*). During its transit through the duct, water is removed from the liquid silk, and an exchange of Na^+ for K^+ accompanies this process (*44*). The pH of the dope is lowered during this travel (*23*), and an increase in phosphate concentration in the final thread compared to the dope has been reported (*24, 45*). These alterations have been suggested to induce structural changes in the spidroin. Consistent with this theory, addition of potassium chloride to a final concentration of 300 mM to a 0.9 wt % silk dope solution induced aggregation and nanofibril formation (*30*).

Moreover, the optimum sheer rate for sheer thickening of the silk dope was found to occur at pH 6.3 (*30*). Herein, no influence on the secondary structure of any of the different parts of MaSp1 was found under these conditions. It is not surprising that the secondary structure of 4Rep is unchanged between pH 4 and 8, since it lacks residues with pK_a values in this interval (Figure 1). The C-terminal domain contains two highly conserved glutamic acid residues, one of them succeeding the Cys that forms an interchain disulfide bridge. However, the secondary structure of CT is unchanged even at pH 2. The N-terminal domain is the most charged of the spidroin domains; it contains three Asp, five Glu, and two Lys residues, one Arg, and one His, quite evenly distributed. However, the secondary structure of NT is likewise unaffected at pH 2–8. It is still plausible that lowering the pH or altered ion composition in the gland or duct weakens repulsive charge interactions and thereby facilitates polymerization.

The secondary structures of the *E. australis* MaSp1 parts also remain unchanged in the presence of potassium chloride, sodium chloride, potassium phosphate, or sodium phosphate up to a concentration of 300 mM. However, after prolonged incubation in 300 mM phosphate, 4Rep and 4RepCT aggregate (Figure 5A,B). The aggregation probably does not require secondary structure alternations, since no changes in the far-UV CD spectra were observed. Near-UV CD, however, shows that aggregation is associated with tertiary structure changes (Figure 5). The increased level of aggregation could be due to salt-induced hydrophobic interactions. HPO_4^{2-} is a stronger chaotroph than $H_2PO_4^-$ and Cl^-, which might explain why phosphate ions are more potent than Cl^- in inducing aggregation, and why aggregation is more pronounced at pH 7.2 than at pH 6.3 (Figure 5). Fiber formation does not seem to be affected by phosphate ions to the same degree as aggregation (Figure 6). This may in part be explained by the fact that for fiber formation to occur, the spidroins are subjected to sheer forces, while aggregation was studied under nonsheer conditions. Moreover, fiber formation for 4RepCT starts within minutes, while it takes hours before aggregated 4RepCT is detected (Figure 5).

Covalent attachment to the C-terminal domain seems to limit the aggregation propensity of the repetitive domain, as the aggregation rate for 4RepCT is slower than that for 4Rep (Figure 5). This supports our hypothesis that the C-terminal domain prevents amorphous aggregation by regulating polymerization. This observation contradicts earlier studies, in which covalent attachment to the C-terminal domain of *A. diadematus* dragline spidroins was found to increase the level of aggregation of repetitive regions (*34*). The latter proteins were purified under denaturing conditions, which may affect aggregation properties. The different results may also be explained by differences in the sequence and length of the proteins that were studied.

CONCLUSIONS

The soluble form of *E. australis* MaSp1 consists of a dimeric α-helical N-terminal domain, a flexible helical repetitive region, and a dimeric α-helical C-terminal domain that is stabilized by an interchain disulfide bridge. All of these parts are structurally tolerant to changes in pH and ion concentration. The N-terminal domain is highly soluble, which may be important for spidroin processing, and the C-terminal domain appears to dictate ordered polymerization of the repetitive region. Aggregation of the repetitive region is induced by phosphate ions, but whether this is important for fiber formation remains to be established.

ACKNOWLEDGMENT

We thank Dr. Sophia Schedin Weiss for providing thrombin.

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*, 3295–3303.
2. Viney, C., Huber, A. E., Dunaway, D. L., Kerkam, K., and Case, S. T. (1994) Optical Characterization of Silk Secretions and Fibers, in *Silk Polymers: Materials Science and Biotechnology* (Kaplan, D., Adams, W. W., Farmer, B., and Viney, C., Eds.) pp 120–136, American Chemical Society, Washington, DC.
3. Vollrath, F., Barth, P., Basedow, A., Engstrom, W., and List, H. (2002) Local tolerance to spider silks and protein polymers in vivo. *In Vivo 16*, 229–234.
4. Hinman, M. B., and Lewis, R. V. (1992) Isolation of a clone encoding a second dragline silk fibroin. *Nephila clavipes* dragline silk is a two-protein fiber. *J. Biol. Chem. 267*, 19320–19324.
5. Xu, M., and Lewis, R. V. (1990) Structure of a protein superfiber: Spider dragline silk. *Proc. Natl. Acad. Sci. U.S.A. 87*, 7120–7124.
6. 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.
7. Sponner, A., Schlott, B., Vollrath, F., Unger, E., Grosse, F., and Weisshart, K. (2005) Characterization of the protein components of *Nephila clavipes* dragline silk. *Biochemistry 44*, 4727–4736.
8. Sponner, A., Unger, E., Grosse, F., and Weisshart, K. (2005) Differential polymerization of the two main protein components of dragline silk during fibre spinning. *Nat. Mater. 4*, 772–775.
9. Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A., and Hayashi, C. Y. (2007) Blueprint for a high-performance biomaterial: Full-length spider dragline silk genes. *PLoS One 2*, e514.
10. Madsen, B., Shao, Z. Z., and Vollrath, F. (1999) Variability in the mechanical properties of spider silks on three levels: Interspecific, intraspecific and intraindividual. *Int. J. Biol. Macromol. 24*, 301–306.
11. Vollrath, F., and Knight, D. P. (2001) Liquid crystalline spinning of spider silk. *Nature 410*, 541–548.
12. Pouchkina-Stantcheva, N. N., and McQueen-Mason, S. J. (2004) Molecular studies of a novel dragline silk from a nursery web spider, *Euprosthenops* sp. (Pisauridae). *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol. 138*, 371–376.
13. Rising, A., Hjalm, G., Engstrom, W., and Johansson, J. (2006) N-Terminal nonrepetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins. *Biomacromolecules 7*, 3120–3124.
14. Motriuk-Smith, D., Smith, A., Hayashi, C. Y., and Lewis, R. V. (2005) Analysis of the conserved N-terminal domains in major ampullate spider silk proteins. *Biomacromolecules 6*, 3152–3159.
15. 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.
16. Ittah, S., Cohen, S., Garty, S., Cohn, D., and Gat, U. (2006) An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. *Biomacromolecules 7*, 1790–1795.
17. Stark, M., Grip, S., Rising, A., Hedhammar, M., Engstrom, W., Hjalm, G., and Johansson, J. (2007) Macroscopic fibers self-assembled from recombinant miniature spider silk proteins. *Biomacromolecules 8*, 1695–1701.
18. Frische, S., Maunsbach, A. B., and Vollrath, F. (1998) Elongate cavities and skin-core structure in *Nephila* spider silk observed by electron microscopy. *J. Microsc. 189*, 64–70.
19. Bell, A. L., and Peakall, D. B. (1969) Changes in fine structure during silk protein production in the ampullate gland of the spider *Araneus sericatus*. *J. Cell Biol. 42*, 284–295.

20. Knight, D. P., and Vollrath, F. (1999) Liquid crystals and flow elongation in a spider's silk production line. *Proc. R. Soc. London, Ser. B 266*, 519–523.
21. Work, R. W. (1977) Mechanisms of major ampullate silk fiber formation by orb-web-spinning spiders. *Trans. Am. Microsc. Soc. 96*, 170–189.
22. 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.
23. Vollrath, F., Knight, D. P., and Hu, X. W. (1998) Silk production in a spider involves acid bath treatment. *Proc. R. Soc. London, Ser. B 265*, 817–820.
24. Knight, D. P., and Vollrath, F. (2001) Changes in element composition along the spinning duct in a *Nephila* spider. *Naturwissenschaften 88*, 179–182.
25. Johnson, W. C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins 35*, 307–312.
26. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis. *Anal. Biochem. 287*, 243–251.
27. Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000) NPS@: Network protein sequence analysis. *Trends Biochem. Sci. 25*, 147–150.
28. Cheng, J., Randall, A. Z., Sweredoski, M. J., and Baldi, P. (2005) SCRATCH: A protein structure and structural feature rediction server. *Nucleic Acids Res. 33*, W72–W76.
29. Pollastri, G., Przybylski, D., Rost, B., and Baldi, P. (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins 47*, 228–235.
30. Chen, X., Knight, D. P., and Vollrath, F. (2002) Rheological characterization of *Nephila* spidroin solution. *Biomacromolecules 3*, 644–648.
31. 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.
32. 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.
33. Kenney, J. M., Knight, D., Wise, M. J., and Vollrath, F. (2002) Amyloidogenic nature of spider silk. *Eur. J. Biochem. 269*, 4159–4163.
34. Huemmerich, D., Helsen, C. W., Quedzuweit, S., Oschmann, J., Rudolph, R., and Scheibel, T. (2004) Primary structure elements of spider dragline silks and their contribution to protein solubility. *Biochemistry 43*, 13604–13612.
35. Mello, C. M., Senecal, K., Yeung, B., Vouros, P., and Kaplan, D. (1994) Initial characterization of *Nephila clavipes* dragline protein, in *Silk polymers. Materials science and biotechnology* (Kaplan, D., Wade, W. W., Farmer, B., and Viney, C., Eds.) pp 67–79, American Chemical Society, Washington, DC.
36. Ittah, S., Michaeli, A., Goldblum, A., and Gat, U. (2007) A model for the structure of the C-terminal domain of dragline spider silk and the role of its conserved cysteine. *Biomacromolecules 8*, 2768–2773.
37. Jin, H. J., and Kaplan, D. L. (2003) Mechanism of silk processing in insects and spiders. *Nature 424*, 1057–1061.
38. Toniolo, C., Polese, A., Formaggio, F., Crisma, M., and Kamphuis, J. (1996) Circular Dichroism Spectrum of a Peptide 310-Helix. *J. Am. Chem. Soc. 118*, 2744–2745.
39. Sreerama, N., and Woody, R. W. (1994) Poly. (pro)II helices in globular proteins: Identification and circular dichroic analysis. *Biochemistry 33*, 10022–10025.
40. Bhattacharjee, S., Toth, G., Lovas, S., and Hirst, J. D. (2003) Influence of tyrosine on the electronic circular dichroism of helical peptides. *J. Phys. Chem. B 107*, 8682–8688.
41. Dicko, C., Knight, D., Kenney, J. M., and Vollrath, F. (2004) Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. Temperature and concentration effects. *Biomacromolecules 5*, 2105–2115.
42. Dicko, C., Vollrath, F., and Kenney, J. M. (2004) Spider silk protein refolding is controlled by changing pH. *Biomacromolecules 5*, 704–710.
43. Dicko, C., Knight, D., Kenney, J. M., and Vollrath, F. (2005) Conformational polymorphism, stability and aggregation in spider dragline silks proteins. *Int. J. Biol. Macromol. 36*, 215–224.
44. 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.
45. 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.

BI702432Y