SUPPORTING INFORMATION

Conformation of spider silk proteins in situ in the intact major ampullate gland and in solution

Thierry Lefèvre, Jérémie Leclerc, Jean-François Rioux-Dubé, Thierry Buffeteau, Marie-Claude Paquin, Marie-Eve Rousseau, Isabelle Cloutier, Michèle Auger, Stéphane M. Gagné, Simon Boudreault, Conrad Cloutier, Michel Pézolet

Department of Chemistry, Department of Biology, Department of Biochemistry and Microbiology, Université Laval, Québec (Québec) G1K 7P4 Canada

Institut des Sciences Moléculaire (UMR 5255 CNRS), Université Bordeaux 1, Talence 33405 cedex France

Table of content:

- Experimental Section
  - Materials
  - Methods (Raman, VCD, and NMR spectroscopies)

- Figure S1. Typical repetitive amino acid sequences of the dragline silk proteins of the N.clavipes spider.

- References
**Experimental Section**

**Materials**

*N. clavipes* and *A. diadematus* spiders were collected in Florida and the province of Québec, respectively. They were raised in the laboratory in $20 \times 50 \times 60$ cm cages at $58 \pm 5\%$ relative humidity and $24 \pm 2^\circ$C. The spiders were fed four times a week with crickets of about one centimeter in length and with 3 drops of 1 or 10% glucose solution per week.

The peptide XAO (acetyl-XXAAAAAAAAAOO-amide) was synthesized using a solid-state method by the Service de synthèse de peptides, Université Laval (Québec, Canada), with a purity of 90% as determined by HPLC. It was dissolved directly in pure H$_2$O or D$_2$O (99.9% D).

The recombinant proteins MaSpI and MaSpII were provided by Nexia Biotechnologies Inc. (Montreal, Québec, Canada). The protein powders were dissolved in 6M guanidine hydrochloride (GdnHCl) at a concentration of 5 mg/ml and heated at 60°C for 60 min. To remove GdnHCl, the proteins were eluted through PD-10 Sephadex™ G-25 desalting columns (Amersham Biosciences, Baie d’Urfé, Québec, Canada) equilibrated with 50 ml of the desired buffer. The solution was centrifuged using an Amicon® Ultra-4 (Millipore, Mississauga, Ontario) filter unit (10,000 MWCO) at 6000 g to obtain a final concentration of 50 mg/ml. If needed, the solution was diluted with the same buffers to the desired concentration. MaSpI was dissolved in H$_2$O (pH 7) whereas MaSpII was dissolved in a phosphate buffer pH 11. The final protein concentration was verified by UV absorption spectroscopy. For the MaSpI-MaSpII mixture, the solutions were simply mixed at the appropriate molar ratio. The resulting pH was around 9. For VCD and some NMR measurements, proteins were solubilized in D$_2$O solutions, so that each step of the whole dissolution procedure was carried out in a nitrogen atmosphere. To adjust the samples to the desired pH, a MI-410 microelectrode
(Microelectrodes Inc., Bedford, New Hampshire) was used with 0.05M NaOH and HCl solutions.

**Methods**

1. **Raman spectromicroscopy**

For Raman measurements of the MA gland lumen, spiders were dissected, and rinsed with a phosphate buffer saline (PBS). The MA glands were carefully exposed paying a special attention not to untie them from the rest of the spider’s body. The glands were entirely covered in the phosphate buffer during the entire experiment to maintain as much as possible the physiological conditions. The laser beam was focused 50-100 µm below the surface of the gland epithelium using a long working distance 50× objective (0.9 NA-Olympus) to probe only the MA gland lumen. The thickness of the MA gland wall measured by optical microscopy of transverse microtomed cuts ranged between 20 and 50µm, depending on the spider species. The so-called A- and B-zones of the ampulla of the gland have been probed and no difference has been noted. Five glands from *N. clavipes* and four from *A. diadematus* have been investigated, recording 3 to 10 spectra of the spinning dope for each gland, with a high reproducibility. The validity of the results obtained inside the MA glands was confirmed by recording directly spectra of silk dope ribbons that were expelled from the ampulla of the gland after making a small incision in the membrane (data not shown).

Spectra were recorded at 20 ± 0.5°C using a LabRam 800HR Raman spectrometer (Horiba Jobin Yvon, Villeneuve d’Ascq, France) coupled to an Olympus BX 30 fixed stage microscope. The 514.5 nm line of an Ar⁺ laser (Coherent, INNOVA 70C Series Ion Laser, Santa Clara, CA) was used for *A. diadematus* whereas the 632.8 nm line of a He-Ne laser (Melles Griot, Carlsbad, CA) was used for *N. clavipes*. The laser beam intensity at the sample was approximately 5 mW (green line) and 7 mW (red line). The confocal hole and the
entrance slit of the monochromator were fixed at 400 and 100 µm, respectively. Data were collected by a one-inch open electrode Peltier-cooled CCD detector (1024 × 256 pixels). The acquisition time of a single spectrum was typically around one min. No sign of sample deterioration was observed under these experimental conditions.

2. **VCD spectroscopy**

Protein, polypeptide, and peptide components were dissolved in D$_2$O at a concentration of 50 mg/ml, except the peptide XAO (29 mg/ml). The VCD spectra were recorded with a ThermoNicolet Nexus 670 FTIR ThermoElectron spectrometer (Madison, WI) equipped with a home-made VCD optical bench. In this apparatus, the light beam was focused on the sample with a BaF$_2$ lens, passing through an optical filter (1850-800 cm$^{-1}$), a BaF$_2$ wire grid polarizer (Specac Ltd, Orpington, UK), and a ZnSe Type II/ZS50 photoelastic modulator (Hinds Instruments, Hillsboro, OR). The light was then focused with a ZnSe lens onto a 1×1 mm$^2$ HgCdTe MCTA* E6032 detector (ThermoElectron). The photoelastic modulator was adjusted for maximum efficiency at 1600 cm$^{-1}$. Samples were held in a CaF$_2$ BioCell™ with a fixed path length of 45 µm (BioTools Inc., Jupiter, FL). VCD spectra were recorded at room temperature at a resolution of 4 cm$^{-1}$, by coadding 72,000 scans. Calculations were performed via the standard ThermoElectron software, using Happ-Genzel apodization, de-Haseth phase-correction and a zero-filling factor of one. Calibration spectra were recorded using a birefringent plate (CdSe) and a second BaF$_2$ wire grid polarizer. Baseline corrections were performed by subtracting the raw spectra of D$_2$O. For the sake of comparison, the VCD spectra were normalized so that all samples have their amide I’ absorbance maximum equal to 1.
3. NMR spectroscopy

The Hα resonances of the MaSpI and MaSpII proteins were assigned using a combination of 2D DQF-COSY spectra and 2D $^{13}$C-HSQC spectra at natural abundance. 5 mg/ml samples were used in 90% H$_2$O/10% D$_2$O for MaSpI (pH 5) and in 90% H$_2$O/10% D$_2$O (phosphate buffer pH 11) for MaSpII. Other samples were directly dissolved in pure D$_2$O. The 2D $^{13}$C-HSQC spectra were acquired on a Varian Inova-600 MHz (Palo Alto, CA) NMR spectrometer and the 2D DQF-COSY were acquired on a Varian Inova-400 MHz NMR spectrometer. For all samples, the temperature was kept constant at 23°C. The 2D $^{13}$C-HSQC spectra were recorded with 512 complex data points, an acquisition time of 0.073 s, a recycle delay of 1 s and 104 transients. 256 increments were recorded in the F1 dimension. The spectra for the samples with 90% H$_2$O were recorded with presaturation of the water peak. The 2D DQF-COSY spectra were recorded with 128 transients, 1k complex data points, a recycle delay of 1 s and an acquisition time of 0.160 s. 512 increments were recorded in the F1 dimension. For both the 2D DQF-COSY and the $^{13}$C-HSQC spectra, linear prediction was used in the F1 dimension with a square sinebell function shifted by 90°. The chemical shifts were referenced relative to internal sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) set to 0.00 ppm.

Chemical shift correction factors. The chemical shifts (δ) of the Hα resonances of MaSpI and MaSpII are presented in Table 1. Some of the chemical shifts were corrected (δ$_R$) because of the sequence-dependent effects on chemical shifts. More specifically, it is necessary to use sequence-dependent correction factors (Δδ), especially in unordered proteins, to compare the chemical shifts of one residue in different environment (different neighbours). We have used the sequence-dependent correction factors given by Schwarzinger et al.\textsuperscript{2} and calculated with GGXGG model peptides, X representing the 20 natural amino acids. It is important to note that the method described by Schwarzinger was slightly modified to compare the experimental chemical shifts measured for the MaSpI and MaSpII proteins with those
obtained in GGXGG peptides. For example, in a sequence of five alanine residues, the resonance of the H\(^{\alpha}\) of the central alanine has to be shifted upfield by 0.08 ppm because of three correction factors \(\Delta\delta_{(A-2)}\) (0.02 ppm), \(\Delta\delta_{(A-1)}\) (0.03 ppm), and \(\Delta\delta_{(A+1)}\) (0.03 ppm) associated with the first, the second and the fourth alanine in the sequence, respectively. The correction factor for the fifth alanine \(\Delta\delta_{(A+2)}\) is zero in this case.

For MaSpII, some residues such as proline, serine and glutamine have different environments throughout the amino acid sequence. It is therefore necessary to use different correction factors for each different environment that a given type of residue experiences. For example, the proline residues experience 4 different neighbour environments (Figure S1), one with a correction factor \(\Delta\delta_{(1)}\) of 0.00 ppm (proline content of 34%), a second with \(\Delta\delta_{(2)}\) = 0.05 ppm (proline content \(x = 39\)%), a third with \(\Delta\delta_{(3)}\) = 0.02 ppm (proline content \(y = 13.5\)%), and a fourth with \(\Delta\delta_{(4)}\) = -0.02 ppm (proline content \(z = 13.5\)%). Calling \(w, x, y, \) and \(z\) the contents of residues of a given type in the neighbour environment associated to correction factors, \(\Delta\delta_{(1-4)}\), we have:

\[
\delta_x = \delta + (w \times \Delta\delta_{(1)}) + (x \times \Delta\delta_{(2)}) + (y \times \Delta\delta_{(3)}) + (z \times \Delta\delta_{(4)})
\]

For proline, this gives:

\[
\delta_x = 4.42 \text{ ppm} + 0.02 \text{ ppm} = 4.44 \text{ ppm}
\]
MaSpI: (AGAAAAAGGAGQGGYGGLGSQGAGRGGGLG)\textsubscript{n}

MaSpII: (SAAAAAAAGPGQQPGGYGPGQQQGGYGPQGQPSGPG)\textsubscript{n}

Figure S1. Typical repetitive amino acid sequences (one letter code) of the dragline silk proteins of the *N. clavipes* spider.

References