Supercontraction and Backbone Dynamics in Spider Silk: ^{13}C and ^{2}H NMR Studies

Zhitong Yang,[†,‡] Oskar Liivak,[†,‡] Andreas Seidel,[‡] George LaVerde,[‡] David B. Zax,[§] and Lynn W. Jelinski[⊥]

Contribution from the Department of Physics, Center for Advanced Technology in Biotechnology, and Department of Chemistry & Chemical Biology, Cornell Center for Materials Research, Cornell University, Ithaca, New York 14853, and Chemistry Department, 240 T. Boyd Hall, Louisiana State University, Baton Rouge, Louisiana 70803

Received May 17, 2000. Revised Manuscript Received July 7, 2000

Abstract: The high-performance mechanical properties of certain spider silks can be radically altered by the addition of water. For example, unconstrained silk fibers from the major ampullate gland of the golden orb-weaving spider, *Nephila clavipes*, contract to about half of their original length when immersed in water. In this paper we use solid-state ^{13}C and ^{2}H NMR to study *N. clavipes* silk fibers, so as to address the molecular origins of supercontraction in the wet silk. Using ^{13}C NMR, we study backbone dynamics and demonstrate that, when in contact with water, a substantial fraction of the glycine, glutamine, tyrosine, serine, and leucine residues in the protein backbone show dramatic increases in the rate of large-amplitude reorientation. ^{2}H NMR of silk samples that incorporate leucine deuterated at one terminal methyl group provides a probe for dynamics at specific side chains along the fiber. Only a subset of these leucine residues is strongly affected by water. We suggest that the highly conserved YGGLGS(N)QGAGR blocks found in the silk protein play a major role in the supercontraction process. Amino acid sequences are proposed to produce artificial spider silk with similar mechanical properties, but without the undesired phenomenon of supercontraction. A possible use of the "supercontracting sequence" is also suggested.

Introduction

The promise that tailor-made proteins might be processed into materials with properties that equal—or even surpass—those of their natural counterparts has rekindled interest in biomimetic and bioinspired polymers.[1] Spider silk is one of the most exciting of these biomolecular systems. Rightfully regarded as Nature's high-performance fiber, the major ampullate gland (dragline) silk of the golden orb-weaving spider *Nephila clavipes* reveals outstanding toughness. Combining strength and stiffness that compete with those of steel[2] with elasticity similar to that of rubber,[3] spider silk is well designed to meet the demanding biological functions of catching prey and supporting the weight of a suspended spider.

Proteins based on the consensus amino acid sequence repeats of *N. clavipes* dragline silk have recently been expressed in bacteria,[4,5] and efforts to spin fibers from such proteins are currently underway.[6−8] A genetically engineered copy of spider dragline silk might find use in applications as varied as protective clothing, lightweight parachutes, restraining seat belts, and medical sutures.[9] Nonetheless, supercontraction remains a major impediment to most of these applications. *N. clavipes* dragline silk fibers shrink to about 55% of their original length when in contact with water,[10] and supercontraction is accompanied by a large increase in elasticity and a corresponding decrease in stiffness of the spider silk—sufficient so that the wet silk acts like a rubber.[11,12]

In other fibers, whether native or synthetic, supercontraction is observed only at high temperatures or in harsh solutions.[13,14] For example, in the cocoon silk of the silkworm *Bombyx mori*, supercontraction is observed only when the silk is immersed in a powerful swelling agent, such as an aqueous solution saturated with sodium thiocyanate.[15,16] Thus, an understanding of how the differences in the primary sequences of these proteins impact on supercontraction is critical if one wishes to eliminate the phenomenon from synthetic fibers inspired by spider dragline silk. Figure 1 compares three repetitive amino acid sequences which have been reported for silks: the dominant

[†] Department of Physics, Cornell University.
[‡] Center for Advanced Technology in Biotechnology, Cornell University.
§Department of Chemistry & Chemical Biology, Cornell Center for Materials Research, Cornell University.
[⊥] Louisiana State University.

(1) Byrom, D. *Biomaterials: Novel Materials from Biological Sources*; Stockton: New York, 1991; pp 1−365.
(2) Gosline, J. M.; DeMont, M. E.; Denny, M. W. *Endeavour* **1986**, *10*, 37−43.
(3) Termonia, Y. *Macromolecules* **1994**, *27*, 7378−7381.
(4) Fahnestock, S. R.; Irwin, S. L. *Appl. Microbiol. Biotechnol.* **1997**, *47*, 23−32.
(5) Arcidiacono, S.; Mello, C.; Kaplan, D.; Cheley, S.; Bayley, H. *Appl. Microbiol. Biotechnol.* **1998**, *49*, 31−38.
(6) Trabbic, K. A.; Yager, P. *Macromolecules* **1998**, *31*, 462−471.
(7) Liivak, O.; Blye, A.; Shah, N.; Jelinski, L. W. *Macromolecules* **1998**, *31*, 2947−2951.
(8) Seidel, A.; Liivak, O.; Jelinski, L. W. *Macromolecules* **1998**, *31*, 6733−6736.
(9) Colgin, M. A.; Lewis, R. *Chem. Ind.* **1995**, *24*, 1009−1012.
(10) Work, R. W. *Text. Res. J.* **1977**, *47*, 650−662.
(11) Work, R. W. *J. Exp. Biol.* **1985**, *118*, 379−404.
(12) Gosline, J. M.; Denny, M. W.; DeMont, M. E. *Nature* **1984**, *309*, 551−552.
(13) Astbury, W. T.; Woods, H. J. *Philos. Trans. R. Soc. Ser. A* **1933**, *232*, 333−394.
(14) Speakman, J. B. *J. Soc. Chem. Ind.* **1931**, *50T*, 1−7.
(15) Asakura, T.; Demura, M.; Watanabe, Y.; Sato, K. *J. Polym. Sci. Polym. Phys.* **1992**, *30*, 693−699.
(16) Mellon, E. F.; Korn, A. H.; Hoover, S. R. *J. Am. Chem. Soc.* **1949**, *71*, 2761−2764.

10.1021/ja0017099 CCC: $19.00
Published on Web 08/30/2000

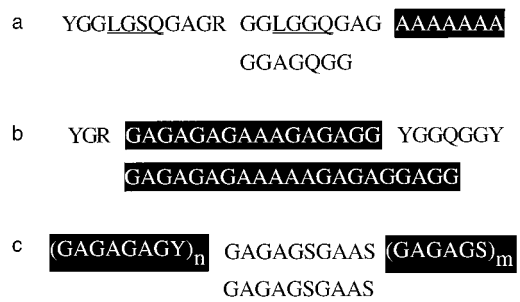

Figure 1. Repetitive amino acid sequences of the dominant proteins in (a) major and (b) minor ampullate gland silk of *N. clavipes*, and (c) the cocoon silk of the silkworm *B. mori*. LGXQ sequences are underlined; sequences thought to be dominant. β-sheets are shown with black background. Single-letter code abbreviation for amino acids is used; deletions and exchanges are neglected.

proteins in the major (MaSp1)[17] and minor (MiSp1)[18] ampullate gland silks of *N. clavipes*, and the cocoon silk of *B. mori*.[19] In each, the most abundant amino acids are glycine (Gly) and alanine (Ala). In two major respects, however, the supercontracting dragline silk differs from the other fibers. First, in MaSp1 the crystalline regions which adopt the β-sheet conformation are composed of poly(Ala) runs,[20,21] while in non-supercontracting *B. mori* cocoon silk the crystalline β-sheet regions are composed of poly(Gly-Ala) runs. The latter are also reasonably abundant in MiSp1. Second, significant numbers of leucine (Leu) residues are found only in the supercontracting MaSp1, and not in the other two non-supercontracting fibers. Furthermore, all leucine appears in conserved sequences LGXQ, accompanied by glycine, glutamine (Gln, Q), and an additional residue (X = S (serine, Ser), N (asparagine, Asn), or G).

It has been hypothesized that in these materials, differences in the structures found in the crystalline regions are associated with the observed differences in their tensile strength.[22] When spider dragline silk is wetted, little change is observed in the structural properties of the densely packed, crystalline poly(Ala) runs—as would be expected for a hydrophobic region of the fiber.[21] Similarly, ^{13}C NMR studies of the dragline silk from spider *Nephila madagascariensis* show that the Ala-rich regions are also found in β-sheets. When dry, the Gly-rich regions prefer to form 3_1-helices; upon wetting, two distinct Gly environments are found.[23] There is, however, relatively little direct evidence for the changes in structure and dynamics found when the amorphous region of spider dragline silk is exposed to water. The amorphous region, and especially the repetitive sequences which contain Leu, are the focus of this study.

In this paper we report on ^{13}C and ^{2}H NMR studies of dynamics in spider dragline silk, and on the residue-specific changes in those spectra found to accompany supercontraction. High-resolution ^{13}C CP/MAS was used to probe dynamics at the fiber backbone and can rapidly survey the entire protein sequence without the need to resort to isotope labeling. To focus more specifically on the dynamic properties of the Gly residues in the silk fibers, static ^{13}C spectra were observed in a silk sample labeled by adding 1-^{13}C-Gly to the spider's feedstock. These experiments allow us to differentiate between fast-moving and immobilized Gly. We further discuss static ^{2}H NMR spectra observed in a silk sample labeled by adding Leu-d_3 to the spider's feedstock. As the ^{2}H label is found in only one methyl group at the end of the Leu side chain, these experiments are sensitive specifically to side-chain dynamics. Experiments on Leu-labeled silk were performed over a broad range of temperatures, both wet and dry, and the results suggest that it is the highly conserved YGGLGS(N)QGAGR sequences, found only in the spider dragline silk (Figure 1), that are responsible for supercontraction in this material. Finally, we propose modifications to the protein primary sequence that should produce fibers with more stable mechanical properties.

Experimental Section

1. Silk Sample Preparation. All silk samples were collected from three groups of adult female spiders *N. clavipes* from central Florida, using the forcible extraction method outlined by Work and Emerson.[24] Three times a week, each spider was silked for 40 min. A dissecting microscope was used to ensure that only silk from the major ampullate gland was collected. All spiders were fed a solution of Dulbecco's modified Eagle medium (DMEM) (Sigma Chemical Co., St. Louis, MO). The solution was hand-fed to the spiders 10 min before and immediately after each silking session. For group I, no additional supplements were supplied, and the silk samples extracted were assigned to the "natural abundance" category. Isotopically labeled amino acids (1-^{13}C-Gly and Leu-d_3) were added to DMEM and fed to group II and group III, respectively, to produce the two differently labeled silk samples. Dragline silk ^{13}C-labeled with 1-^{13}C-Gly was extracted from 20 spiders fed with a 4% w/v solution of 1-^{13}C-Gly (Cambridge Isotope Labs, Andover, MA) in DMEM. After five silking episodes, 104.1 mg of ^{13}C-labeled silk was harvested. Dragline silk ^{2}H-labeled at Leu was extracted from 20 spiders fed with a 2.5% w/v solution of L-Leu-d_3 (methyl-d_3) (C/D/N Isotopes, Pointe-Claire, Quebec, Canada)-enriched DMEM. After four silking episodes, 71.36 mg of deuterated silk was harvested. NMR experiments demonstrated that ^{2}H incorporation reached its peak 1 week after the Leu-d_3 was first introduced. The labeling levels achieved in these samples were monitored by electrospray ionization (ESI) mass spectroscopy.

2. Solid-State NMR. All solid-state ^{13}C NMR measurements were performed on a home-built spectrometer operating at 90.56 MHz. ^{13}C experiments were carried out in a triply tuned magic angle spinning (MAS) probe (Doty Scientific) used in double-resonance mode, and samples were consistently spun at 5 kHz. Chemical shifts were referenced to adamantane as an external secondary standard whose downfield peak was offset by 38.56 ppm from TMS. Cross-polarization (CP) experiments were performed with a Hartmann−Hahn matching field of 60 kHz and a mixing time of 2 ms. During the detection of the evolving ^{13}C magnetization, the ^{1}H decoupling field was increased to 90 kHz. Peak assignments were made using literature values of the chemical shifts.[25,26]

Two different spectrometers were used for ^{2}H NMR experiments, operating at either 55.26 or 58.77 MHz. Except where otherwise noted, ^{2}H spectra were acquired using the quadrupole echo pulse sequence,[27,28] where the separation τ_{echo} between 90° pulses of length 1.6 μs was 35 μs. The evolving signal was sampled with one complex point per microsecond, and scans were repeated at half-second intervals. In some experiments, τ_{echo} was varied between 20 and 100 μs so as to investigate the time scale of the dynamics in the silk. ^{2}H relaxation times (T_1) of the silk were measured by the inversion recovery method, where the

(17) Xu, M.; Lewis, R. V. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 7120−7124.
(18) Colgin, M. A.; Lewis, R. V. *Protein Sci.* **1998**, *7*, 667−672.
(19) Mita, K.; Ichimura, S.; James, T. C. *J. Mol. Evol.* **1994**, *38*, 583−592.
(20) Simmons, A.; Ray, E.; Jelinski, L. W. *Macromolecules* **1994**, *27*, 5235−5237.
(21) Simmons, A. H.; Michal, C. A.; Jelinski, L. W. *Science* **1996**, *271*, 84−87.
(22) Hayashi, C. Y.; Shipley, N. H.; Lewis, R. V. *Int. J. Biol. Macromol.* **1999**, *24*, 271−275.
(23) Kümmerlen, J.; van Beek, J. D.; Vollrath, F.; Meier, B. H. *Macromolecules* **1996**, *29*, 2920−2928.
(24) Work, R. W.; Emerson, P. D. *J. Arachnol.* **1982**, *10*, 1−10.
(25) Howarth, O. W.; Lilley, D. M. *Prog. NMR Spectrosc.* **1978**, *12*, 1−40.
(26) Saito, H. *Magn. Reson. Chem.* **1986**, 24, 835−852.
(27) Powles, J. G.; Strange, J. H. *Proc. Phys. Soc.* **1963**, *82*, 6−15.
(28) Davis, J. H.; Jeffrey, K. R.; Bloom, M.; Valic, M. I.; Higgs, T. P. *Chem. Phys. Lett.* **1976**, *42*, 390−394.

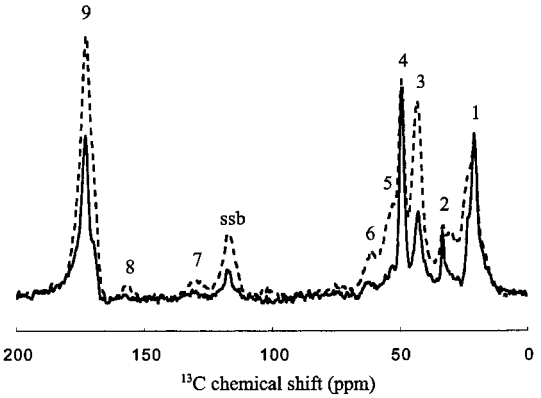

Figure 2. ^{13}C CP/MAS spectra of *N. clavipes* dragline silk in the dry and the water wetted state. Dashed line (- - -) is the spectrum of the dry silk; solid line (—) is the spectrum of the silk immersed in water. Peak assignments include the following: (1) alanine C_β; (2) glutamine C_γ; (3) glycine C_α; (4) alanine C_α; (5) glutamine C_α, leucine C_α, serine C_α, tyrosine C_α; (6) serine C_β; (7) tyrosine C_δ; (8) tyrosine C_ξ; (9) all carbonyl sites. Note the large signal losses in all of the signals in the spectrum of the wet silk, except for the signals assigned to the C_α and C_β carbons of alanine.

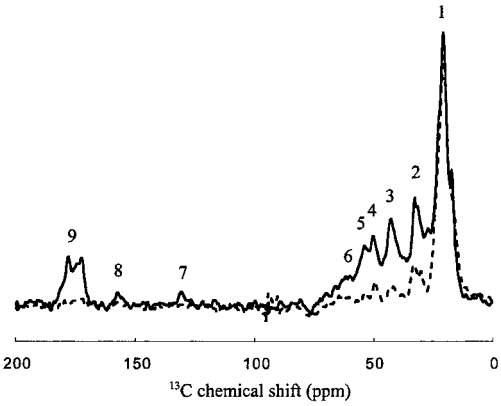

Figure 3. ^{13}C MAS Hahn spin−echo spectra of the dry and water-wetted spider silk using a delay time of 1 s. Dashed line (- - -) is the spectrum of the dry silk; solid line (—) is the spectrum of the wet silk. Peak assignments are as in Figure 2.

inverted magnetization was detected by a quadrupole echo sequence.

Results

1. Rapid Dynamics Probed by ^{13}C NMR. Figure 2 shows the ^{13}C CP/MAS spectra of dry and water-wetted silk collected from the major ampullate gland of the spider *N. clavipes*. The total signal intensity observed in the spectrum of the latter (solid line) is significantly less than that observed in the spectrum of the former (dashed line). Molecular motion, which averages out the dipole−dipole interactions necessary for successful cross polarization, is responsible for the decrease in signal intensity.[29,30] Signal intensity loss, however, is not uniform across the spectrum but instead is concentrated at discrete chemical shift values. Because specific residues and conformations yield identifiable resonance frequencies, this decrease in signal intensity is a marker for increased residue-specific dynamics. In agreement with previous ^{2}H NMR results,[21] signals from Ala in the β-sheet (C_α at 49 ppm, C_β at 20 ppm, and indicated as peaks 4 and 1, respectively, in Figure 2)[25,26] show no change in intensity upon wetting and therefore undergo no change in dynamics. Signals arising from the amino acids Gln, Gly, Ser, Tyr (tyrosine, Y), and Ala in other conformations (with chemical shifts in the range 15−18 ppm) are considerably diminished; these residues are predominantly found in the Leu-rich region of the silk.[25,26] When water is added to the silk, the frequency of backbone dynamics in the amorphous region increases dramatically, while no change is observed in the extensive poly(Ala) runs.

To further verify this interpretation, and to quantify the dynamical rates associated with these motions, we exploited changes in carbon spin−lattice relaxation times (T_1) associated with wetting.[31] Short ^{13}C T_1 times (<1 s) typically indicate large-scale motion on a time scale comparable to the inverse of the Larmor frequency—in our case, a few nanoseconds. T_1-weighted spectra can be observed by using the natural polarization of the ^{13}C nuclear spins unenhanced by polarization transfer techniques. Figure 3 shows the MAS NMR spectra observed without cross polarization for the dry and wet silk. As probe ring-down made it difficult to observe the signal directly after a strong radio frequency pulse, these spectra were observed by applying a pair of pulses spaced by the rotor period (200 μs) and accumulating the rotor and spin−echo signal which arises one rotor period later. (A discussion of possible distortions due to short ^{13}C T_2 effects appears below.) The observed spectra were obtained with a recycle delay time $\tau_{rec} = 1$ s, which is much shorter than the typical T_1 for carbons in solids (10−100 s). As the signal recovery is proportional to $1 - \exp(-\tau_{rec}/T_1)$, any large signals observed in the wet silk with τ_{rec} shorter than 10 s must be associated with short T_1 relaxation times, and therefore with nanosecond time scale dynamics. These signals can be assigned to particular amino acid residues identified by chemical shift. For example, 40% of the Gly C_α carbon signal intensity is observed when $\tau_{rec} = 1$ s in the wetted silk. The corresponding sites in dry silk have been reported to have $T_1 \cong 9$ s, where only 10% of the maximum signal would be observed after 1 s.[20] While the data are insufficient to fully characterize the magnetization recovery curves (i.e., to identify whether it is a double, triple, or stretched exponential), they do serve to efficiently divide the total population of Gly residues into fast- (short T_1) and slow-moving (long T_1) components. This large decrease in T_1 allows us to estimate that the highly mobile Gly fraction constituted 35 ± 10% of the total, and that the dynamical correlation frequency of these fast relaxing Gly C_α carbons is $10^{8\pm1}$ Hz.

The difference between solid and dashed lines in Figure 3 also demonstrates that the C_α carbons of Gln, Leu, Ser, Tyr, and Arg (arginine, R)—found at shifts between 54 and 56 ppm[25,26]—have similarly shortened T_1 relaxation times. As the amino acid resonance frequencies are not individually resolved, no residue-specific conclusions are possible; nonetheless, we can estimate that as a group more than half of these C_α carbons relax with T_1's of less than 1 s. By similar logic, the C_α carbons in half of these residues also have correlation frequencies in the range of $10^{8\pm1}$ Hz.

2. Slower Dynamics Probed by ^{13}C NMR. The T_1-weighted ^{13}C measurements screen efficiently for rapid dynamics but are insensitive to slower dynamical modes. Where roughly half of the non-Ala residues (and nearly all of the Ala) show no evidence for T_1 changes upon wetting, and therefore move on time scales much slower than the nanosecond range, a different technique capable of probing slower modes is required. Dynamical averaging of the orientation-dependent (nonspinning)

(29) Earl, W. L.; VanderHart, D. L. *Macromolecules* **1979**, *12*, 762−767.
(30) Torchia, D. *J. Magn. Reson.* **1978**, 30, 613−616.
(31) Stejskal, E. O.; Memory, J. D. *High-Resolution NMR in the Solid State*; Oxford University Press: New York, 1996; pp 45−53.

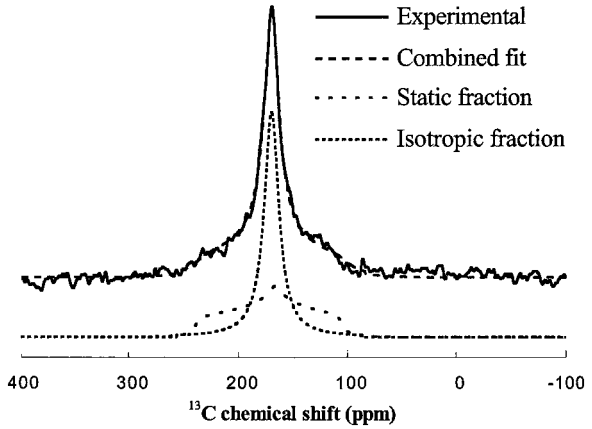

Figure 4. Static ^{13}C spin−echo spectrum of the carbonyl region of a wet silk sample derived from spiders fed with 1-^{13}C-glycine. The spectral fit assumes two glycine carbonyl populations, one nearly isotropically averaged and the other essentially static. There is an additional static contribution from ^{13}C incorporated into alanine. The best-fit components for both the isotropic and static fractions of the spectrum are also shown, where the isotropic portion is modeled as a Gaussian centered at 172 ppm with fwhm of 17 ppm.

chemical shift anisotropy (CSA) provides information on modes with time scales comparable to the inverse of the static powder pattern. For typical carbonyl groups, averaging of the ^{13}C CSA probes dynamics on a time scale of tens of microseconds.

A complication which arises in working with CSAs is that the static powder pattern representing the anisotropic distribution of magnetic environments found in ^{13}C NMR is comparable to the entire range of isotropic ^{13}C chemical shifts. Unless individual sites are selectively labeled, spectral overlap may be debilitating. To this end we prepared the 1-^{13}C-Gly silk sample described above. ESI mass spectroscopy revealed that in our silk sample the ^{13}C-carbonyl label had been incorporated into 37% of the Gly residues in the silk, while 13% of the Ala residues had also incorporated the carbonyl label. No significant incorporation was observed at other amino acid residues.[32]

Figure 4 shows the static, fully relaxed spectrum of the wet, carbonyl-labeled silk sample, collected using a spin−echo pulse sequence appropriate to spin $^1/_2$, where $\tau_{echo} = 200\ \mu s$ and on-resonance for the typical carbonyl isotropic shift. (A second spectrum observed with $\tau_{echo} = 40\ \mu s$ showed no significant intensity differences, indicating that there is little dynamics-induced dephasing on these time scales in the spin−echo spectrum.) The carbonyl ^{13}C tensors are spectroscopically far removed from any contributions from the C_α resonances. Furthermore, Gly and Ala comprise more than 70% of the amino acid residues in the silk protein, and only these residues have incorporated the ^{13}C label, so that the powder pattern of Figure 4 represents only the carbonyl sites from these two types of amino acid residues. Our goal is to characterize the distribution of carbon sites found in Figure 4 into broad (static) and narrowed (mobile) components (where the sites observed in the Bloch decay experiment in the wet silk described above contribute to the latter fraction).

More than 90% of Ala residues in MaSp1 in the native silk are found in the crystalline β-sheets. Their contribution to the spectrum of Figure 4 can be modeled using tabulated values of the carbonyl tensor in poly(Ala)[33] and the assumption that no dynamical changes are induced when the silk is wetted. Similarly, Van Beek et al.[34] have determined the CSA tensor parameters for the carbonyl carbon of Gly in *N. madagascariensis* silk, which we assume describe equally well the same sites in our silk sample. This provides both the powder pattern describing immobile Gly residues and the isotropic shift value characteristic of mobile Gly sites in wet *N. clavipes* silk. Using these input parameters, and the relative intensities of the Ala and Gly contributions to the spectrum (determined from incorporation levels measured by mass spectroscopy), a non-linear least-squares fit[35] was used to determine the distribution of Gly residues into either immobile or rapidly moving (i.e., dynamics >20 kHz) sites. By this method we estimate that 40 ± 5% of Gly residues are found in the broad powder pattern, representing residues which undergo little dynamics when wet. The remainder (60 ± 5%) reorient at rates faster than 20 kHz, and therefore contribute to the motionally averaged portion of the powder pattern. Of this latter 60%, two-thirds were demonstrated above to reorient in the wet silk; this leaves about one-third of the averaged intensity, or one-fifth of the total Gly residues, mobile on a time scale more rapid than 20 kHz but slower than 100 MHz.

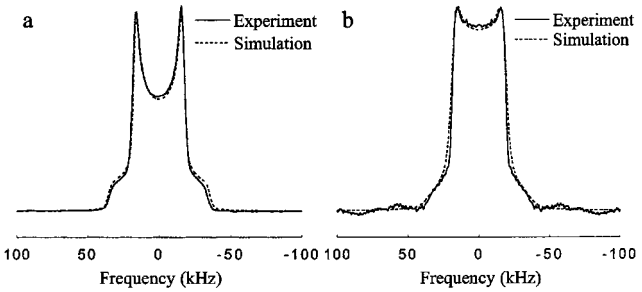

Figure 5. Ambient-temperature ^2H NMR spectra of (a) polycrystalline L-leucine-d_3 (methyl-d_3) powder (C/D/N, Pointe-Claire, Quebec) and (b) unoriented dry dragline silk collected from spiders fed with this compound. In addition to the experimental spectra, simulated line shapes and difference spectra are shown.

3. Side-Chain Dynamics in Dry Deuterated Silk. Figure 5 compares the room-temperature ^2H NMR spectra of polycrystalline L-Leu-d_3 powder to that of dry silk incorporating deuterated Leu. The spectrum of the former (Figure 5a) exhibits the typical Pake doublet powder pattern, and a fit which assumes that the electric field gradient (EFG) is axially symmetric ($\eta = 0$) is reasonable. The two prominent singularities split by approximately 40 kHz are diagnostic of a methyl group rotating rapidly about its C_3 symmetry axis, with little additional dynamical averaging.

ESI mass spectroscopy analysis of our deuterated silk suggests that the methyl-d_3 unit is incorporated in 41% of Leu residues. Amino acid analysis suggests that roughly 4.5% of the amino acid residues in MaSp1 correspond to Leu. Where many of the Leu environments are chemically distinct, one might imagine that the ^2H spectrum of the labeled silk would be horribly

(32) In natural abundance, 1.1% of each amino acid residue contains a ^{13}C in the carbonyl. After the spiders were fed as described, we found 37% of the glycine residues contain ^{13}C while 13% of alanine residues contain ^{13}C. As 45% of the amino acids in the major ampullate silk fibroin are glycine, and 28% are alanine, this results in a carbonyl region spectrum which reflects primarily the glycine carbonyl, with a small additional contribution from alanine.

(33) Duncan, T. M., *A Compilation of Chemical Shift Anisotropies*, 2nd ed.; Farragut Press: Chicago, 1994.

(34) Van Beek, J. D.; Kümmerlen, J.; Vollrath, F.; Meier, B. H. *Int. J. Biol. Macromolecules* **1999**, *24*, 173−178.

(35) Code for simulation of the NMR spectral line shapes was written using the "n2f" nonlinear least-squares fitting routine from the Port Library [AT&T Bell Laboratories, Murray Hill, NJ; see: Dennis, J. E.; Gay, D. M.; Welsch, R. E. *ACM Trans. Math. Software* **1981**, *7*, 348−368; 369−383] and/or the program MXET1 (copyright 1986 University of California Regents), which was provided to us by the late R. R. Vold.

complex, as each distinct environment might make its own contribution to the overall spectrum. Given the paucity of experimental data, such an approach seems unlikely to yield useful insights. Further complicating the analysis is the possibility that other residues contribute to the observed ^{2}H NMR spectrum. While mass spectral analysis suggests that no other amino acid residue shows appreciable ^{2}H incorporation, even at natural abundance levels (about 1.5 ^{2}H nuclei per 10 000 hydrogen atoms), the background contribution is not negligible. (In fact, in high signal-to-noise ratio experiments, additional signal intensity is observed at quadrupolar frequencies which would be characteristic of ^{2}H in other chemical environments,[36] with the singularities in the vicinity of 50−60 kHz as their dominant feature. Should there be any scrambling of the label in the biosynthetic pathway, the number of such sites contributing to the spectrum would increase.) As a result, only a much simplified model appears justified. We instead assume, as a starting point for the analysis, that all Leu environments in the silk are similar.

While the total line width in the methyl group region is comparable to that observed in the spectrum of crystalline Leu (Figure 5a), the line shape of these ^{2}H nuclei observed in the dry silk differs dramatically (Figure 5b), and the significant increase in intensity observed at the center region of the spectrum is characteristic of dynamics which partially averages the observable quadrupolar coupling. While this effect might arise from multiple phases with distinct dynamic modes, or from a single phase but each chemically distinct Leu environment having its own dynamical environment as well, relaxation measurements suggest otherwise and lead us to attempt to fit the side-chain dynamics to a uniform model. Quadrupole echo signal intensities at room temperature were measured as a function of time after an initial strong π pulse intended to invert all ^{2}H spin magnetization. The recovery of the spin magnetization toward equilibrium showed no breaks, as would be expected for distinct dynamical populations, and instead was best described by a stretched exponential[37] with a "stretching parameter" of 0.83, which is quite typical for single phase, glassy polymers.[38] This suggests that most Leu residues occupy similar conformational and dynamical environments in the dry silk.

Variable-temperature NMR measurements were performed on the dry spider dragline silk as a test of the suggested model. Spectra were recorded from −90 to +90 °C at intervals of 10 °C. Many of the spectral features observed at low temperature (<250 K) can be reproduced with a simple dynamical model inspired by previous studies of the Leu side chain in collagen fibrils.[39] In that work, NMR data were presented which confirmed a slow dynamic equilibrium between the two rotamers observed in earlier X-ray studies,[40] corresponding to a rotation via 120° about the $C_\beta-C_\gamma$ bond converting the g^+t and tg^- conformers of the side chain. This additional dynamical mode—superposed atop the rapid C_3 rotation of the methyl group—

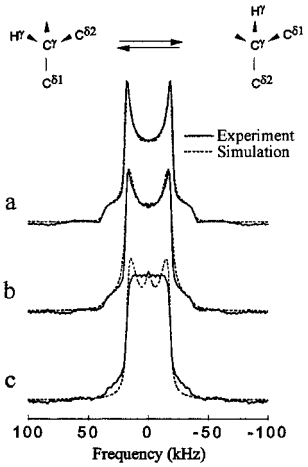

Figure 6. ^{2}H NMR spectra (bottom) of the dry, Leu-labeled dragline silk. Spectra were recorded at (a) −90, (b) 0, and (c) 90 °C. Simulations (dashed lines) are fits which assume jumps about the $C_\beta-C_\gamma$ bond as illustrated, corresponding to sloppy interconversion through 124° rotations between the two most stable conformations of the Leu side chain (top), at jump rates of (a) 8, (b) 16, and (c) 50 kHz. While the model seems appropriate at low temperature, additional dynamical modes are required to fit the observed spectra at higher temperatures.

reasonably reproduces experimentally observed spectra (both line shapes and intensities) at low temperatures and slow jump rates (<20 kHz), where the jump angle is assumed to be "sloppy"—that is, varying by about ±5° from the ideal (see Figure 6a and 6b). As the correlation time for the two-site jump is comparable to the spacing between pulses in the echo sequence, line shape averaging is accompanied with some signal loss—accounting quantitatively for most of the apparent discrepancy between labeling levels observed by mass spectroscopy, and the somewhat diminished signal levels we observe via ^{2}H NMR. The observed jump rates would suggest that the barrier to leucyl group reorientation is unusually low and most probably less than a third of that observed in collagen.

At higher temperatures, however, there is significant deviation between the predictions of a two-site jump and the observed spectra (Figure 6c); in particular, the experimental spectra are significantly flatter in their center than what is found in any of our single-mode simulations. Allowing for unequal populations of the two rotamers provides no substantial improvement in the quality of the fit from simulations to the observed spectrum; instead, it would appear that at higher temperatures there is an additional dynamical mode which is not present in collagen.[39] This suggests that, whatever the details of the local modes, the silk molecule is more flexible and/or packs with substantially more local free volume in the vicinity of the Leu residues in silk than in the threefold helices of collagen.[41] This finding is consistent with the hypothesis that the Leu residues in the dragline silk of *N. clavipes* participate in compact turnlike structures,[42] because packing of amino acid side chains in such a turn is relatively inefficient.

4. Side-Chain Dynamics in Supercontracted Silk. After the material was soaked thoroughly in ^{2}H-depleted water (Aldrich Chemical Co., Milwaukee, WI), the solid-state ^{2}H NMR spectrum of silk was again recorded at ambient temperature but in the supercontracted state (Figure 7). Its spectral line shape differs considerably from that of the dry material recorded with identical spectrometer settings (Figure 5b). Convincing spectral

(36) Scrambling of methyl group deuterium to non-methyl group positions had been observed for an alanine-3,3,3-d_3-labeled silk; see, for example, ref 21.

(37) The longitudinal relaxation times (T_1) are found to be orientation-dependent, with values ranging from 34 ± 4 to 42 ± 2 ms measured at the parallel and perpendicular parts of the line shape, respectively. This dependence is consistent with the three-site jump model for methyl group rotation; see, for example: Batchelder, L. S.; Niu, C. H.; Torchia, D. A. *J. Am. Chem. Soc.* **1983**, *105*, 2228−2231.

(38) Schnauss, W.; Fujara, F.; Hartmann, K.; Sillescu, H. *Chem. Phys. Lett.* **1990**, *166*, 381−384.

(39) Batchelder, L. S.; Sullivan, C. E.; Jelinski, L. W.; Torchia, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 386-389.

(40) Janin, J.; Wodak, S.; Levitt, M.; Maigret, B. *J. Mol. Biol.* **1978**, *125*, 357−386.

(41) Ramachandran, G. N.; Ramakrishnan, C. *Biochemistry of Collagen*; Plenum Publishing Co.: New York, 1976; pp 52−57.

(42) Michal, C. A.; Jelinski, L. W. *J. Biomol. NMR* **1998**, *12*, 231−241.

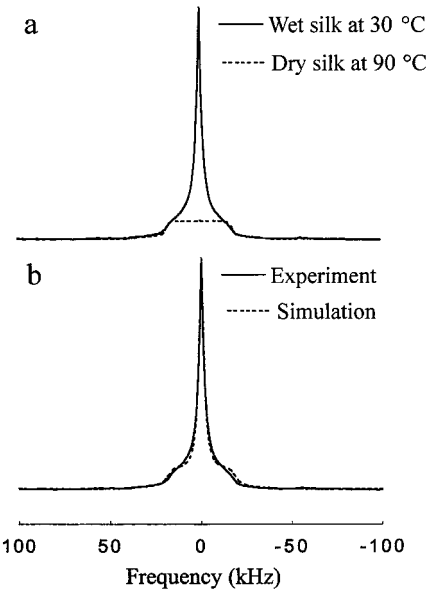

Figure 7. Ambient-temperature ^{2}H NMR spectrum of the wet Leu-labeled dragline silk. (a) Comparison to the spectrum of the dry silk at 90 °C. (b) Comparison to a simulated line shape obtained under the assumption of the presence of two dynamically distinct populations of leucine, one of which is identical to the high-temperature population found in dry silk, and the other of which is Lorentzian with 5-kHz line width.

simulations in the wetted silk require that we assume two distinct populations of Leu with dramatically different dynamical properties (Figure 7b). As the narrowed component is Lorentzian in shape, it must be ascribed to Leu residues whose side chains reorient both rapidly and nearly isotropically. The remainder of the spectral intensity observed in the wetted silk is nearly superimposable with the high-temperature spectrum of the dry silk, suggesting that the dynamics in the second population of Leu residues are not very different whether wet or dry. The best agreement between the experimental and simulated line shapes was achieved where a Lorentzian line of width 5 kHz is added to the high-temperature spectrum (90 °C) of the dry silk. Even those Leu residues whose motion remains restricted when they are wet appear to be more mobile than those in the dry silk, and therefore the free volume available to these Leu residues is increased in the wet silk. This is consistent with the observation that when wet silk supercontracts, the observed shrinkage to about half its original length is accompanied by a more than 3-fold increase in the fiber's cross-sectional area.[10] Swelling in the plane transverse to the fiber axis opens up more space for the reorientation of Leu side chains.

Quantification of the Leu fraction in each component is complicated by the additional loss of roughly one-third of the NMR signal intensity in the wet silk (as compared to that of the same sample when dry) at 30 °C. While it is common for dynamic processes to be associated with signal intensity losses in ^{2}H NMR, and always difficult to unambiguously assign the signals which have not been observed, we nonetheless believe that all the "missing" signals are associated with Leu in the nearly isotropically averaged component. In the fast dynamics limit found in this system, changes in the rate of the two-site hop motion lead to no appreciable signal loss. On the other hand, the signal from Leu residues participating in isotropic motion forms almost no echo in the quadrupole echo pulse sequence,[43] as the isotropic motion averages away the quadrupolar interaction. Instead, the observable intensity in this spectral component decays exponentially with time after the first excitation pulse. This necessarily reduces the signal intensity assigned to the nearly isotropic component of the spectrum.

Figure 8. Model of the dynamics and phase structure of MaSp1 protein in the major ampullate gland silk of the spider *N. clavipes*. The β-sheet region, immobile in wet silk, is depicted as white characters on a black background; the amorphous region, accessible to water and fast moving in wet silk, is shown as bold characters; and the interphase, having increased dynamics in wet silk, is in dark characters on a gray background. The amorphous region is most likely to be responsible for the supercontraction effect.

Discussion

In this section we propose a model for dynamics in the wet silk, dividing the consensus amino acid sequence in the protein MaSp1 according to time scales of dynamics in the wet and dry forms. (We ignore possible contributions to the silk from MaSp2. Pro, which is found almost exclusively in MaSp2, is largely absent in both our NMR spectra and amino acid analyses of the protein, and thus MaSp2 would appear to comprise <5% of the total protein in our samples.[44]) This division is made largely on the basis of the evidence described above, which allows us to characterize specific amino acid residues as belonging to either the rapidly moving fraction (based on T_1 evidence at frequencies of order 10^8 Hz) or to sites whose dynamics is too slow to impact on T_1. In the important case of the abundant residue Gly, additional experiments allow us to distinguish between the slowly moving (frequencies of order 10^4) and the immobile (based on line shape data). As previous work has established that the β-sheet Ala residues in silk are nearly static, both when the silk is dry and when it is wet,[21] we associate the immobilized Gly residues with the amino acid sites physically closest to the poly(Ala) runs found in the silk. This model is summarized in Figure 8, where residues identified with motion at frequencies less than 20 kHz are depicted as white on black (the poly(Ala) β-sheet region).[21] In good agreement with the ^{13}C results above, this assigns 37% of the total Gly residues in the silk to the static region.

Information on the dynamical properties of the less abundant residues is derived both from the T_1-weighted ^{13}C spectra and, in the case of Leu, from the ^{2}H NMR results described above. The latter provides an identification of the most mobile portion of the silk, if we assume that the rapid side-chain dynamics is correlated to sites of rapid main chain dynamics. In the

(43) Spiess, H. W.; Sillescu, H. *J. Magn. Reson.* **1981**, *42*, 381−389.

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

consensus repetitive sequence of the MaSp1 protein of *N. clavipes* dragline silk, nearly all Leu residues appear in only one of two environments (Figure 8). The highly conserved YGGLGS(N)QGAGR blocks (bold) contain 45.2% of all Leu residues, while nearly all the rest (51.6%) are associated with LGGQ blocks found in the less conserved portions of the protein (normal characters in gray background).[45] Intensity measurements alone appear insufficient to conclusively identify those sites undergoing nearly isotropic motion in the wet silk to either one or the other. Chemical considerations, however, would suggest that the more mobile Leu side chains are in the highly conserved blocks. While the hydrophilic Gln(Q) appears in both sequences, the LGS(N)Q segments represent the only portions of the protein sequence of MaSp1 with consecutive hydrophilic amino acid residues (SQ or NQ). Close to these block are found both Tyr (Y) and Arg (R), which are also hydrophilic. That these amino acid blocks are perfectly conserved in the silk protein is, additionally, a suggestion of functional importance. The LGXQ blocks have previously been studied by REDOR methods, which demonstrate that in the dry state these units form compact turnlike structures.[42] When wetted, these hydrogen-bonded structures may be disrupted, which would decrease the relative alignment of β-sheets connected via the flexible linkers. As the fiber length is determined by the alignment of the β-sheets, supercontraction would appear to be related to the loss of structural integrity in the connectors, and specifically with the increased dynamical mobility associated with what were compact turns when dry. This is in agreement with the loss of orientational order of the β-sheets observed by X-ray under wet conditions[46] and consistent with the observed swelling of the fiber which accompanies supercontraction. The remaining Leu residues are surrounded, instead, by largely hydrophobic residues and show no change in dynamical mode.

The spectra shown in Figure 3 provide a basis for the quantitative separation of other residues in the silk backbone between the rapidly moving ($\omega_c \approx 10^{8\pm1}$ Hz) and largely immobilized portions of the wet silk. A comparison of the intensities of well-isolated chemical shifts in the wet and dry systems suggest assigning a further 30% of the total Gly, and about half of all Tyr, Leu, Ser, Gln, and Arg residues, to the dynamically mobile region of the polymer containing the isotropically reorienting Leu, as indicated in the region of outline characters of Figure 8. These observations collectively confirm the role of the YGGLGS(N)QGAGR units, which represent the amorphous, loosely packed region of the protein, in the phenomenon of supercontraction.

Approximately one-third of the Gly, and half of all non-Gly, non-Ala residues, remain unassigned to either the crystalline or amorphous regions of the silk backbone. We assume that all these residues compose a transitional region in the sequence, with dynamics intermediate between those of the slow- and fast-moving regions. The existence of such a coupling region has been predicted previously and was suggested to be critical to the silks' mechanical properties.[3]

We assign the remainder of these residues to dynamical regimes using the recent observations of Valluzzi et al.,[47] who demonstrated that a simple, linear correlation exists between the size of the amino acid residues in β-sheets (as represented by their average volume in solution) and the average intersheet distance. In *N. clavipes*, the intersheet spacing was found to be 5.3 Å,[48] both wet and dry, which is slightly smaller than that in poly(Ala). As 37% of the Gly residues are found in the β-sheet along with the vast majority of the Ala, some larger amino acids must also be involved. Where we incorporate into the β-sheet region all of the residues marked as white letters on black background, which includes half of the large Gln residues (and which would correspond to the static Gln residues observed by ^{13}C NMR, Peak 2 in Figure 2), the average residue volume is in good agreement with their correlation. The remaining amino acid residues—which include all of the Gly which belong neither to the static block nor to the floppy-when-wet block, as their dynamical rates are between 20 kHz and 100 MHz—we assign to the proposed transitional regime which is shown with gray background in Figure 8.

Conclusions

Given an understanding of the residue-specific changes in dynamics associated with supercontraction, modifications to those the naturally occurring sequences using the techniques of genetic engineering should allow for the design of protein sequences optimized for any of a wide range of mechanical properties. For example, to mimic synthetic fibers such as Spandex, a protein fiber is required which is amorphous and floppy both wet and dry—unlike what is found in *N. Clavipes* silk. Similarly, practical exploitation of a silklike protein may require a sequence which maintains a high modulus and does not supercontract in water—which could be achieved if the disordering dynamics in amorphous region remains inhibited, even in the presence of water. Controlling dynamics in the presence and absence of solvent is critical to achieving the desired mechanical properties in these fibers.

The present investigation suggests that only a small, highly conserved block of amino acids in the major ampullate gland silk of the spider *N. clavipes* plays a major role in the interaction of the fibers with water and results in supercontraction. We predict that exchange of the YGGLGS(N)QGAGR blocks of the native fibroin by the less hydrophilic, but structurally analogous YGGLGGQGAGQ, would produce fibers with stiffness, strength, and elasticity similar to those of the natural counterpart but much less perturbed in the presence of water. On the other hand, the YGGLGS(N)QGAGR sequence could be used to produce a biosensor that would be useful in humidity-critical applications. Such materials will have a wide range of potential applications.

Acknowledgment. The authors acknowledge the financial support of the National Science Foundation Grants MCB-9601018 and DMR-9708062, and through seed funding provided by the Cornell Center for Materials Research under NSF Grant DMR-9632275. Z.Y. thanks funding through a Whitaker scholarship. O.L. acknowledges the support of the National Institute of Health Training Grant in Molecular Physics of Biological Systems (NIH T32GM08267). A.S. thanks the Deutsche Forschungsgemeinschaft for a postdoctoral scholarship. The authors are grateful to A. Simmons for insightful discussions and C. Michal for his help in building the temperature control unit. Mass spectra were provided by Dr. Athula Attygalle. C. Wehman, R. Do, and S. B. Lee assisted with sample collection and caring for the spiders. Constructive comments from an anonymous reviewer are greatly appreciated.

JA0017099

(45) There is one Leu found outside the LGXQ sequences.

(46) Work, R. W.; Morosoff, N. *Text. Res. J.* **1982**, *52*, 349−356.

(47) Valluzzi, R.; Szela, S.; Avtges, P.; Kirschner, D.; Kaplan, D. *J. Phys. Chem. B* **1999**, *103*, 11382−11392.

(48) Becker, M. A.; Mahoney, D. V.; Lenhert, P. G.; Eby, R. K.; Kaplan, D.; Adams, W. W. In *Silk Polymers: Materials Science and Biotechnology*; Kaplan, D., Adams, W. W., Farmer, B., Viney, C., Eds.; ACS Symposium Serious 544; American Chemical Society: Washington, DC, 1994; pp 185−195.