ChemComm

Spider Prey-Wrapping Silk is an α-helical coiled-coil / β-sheet Hybrid Nanofiber

Journal:	*ChemComm*
Manuscript ID	CC-COM-06-2018-005246.R1
Article Type:	Communication

Journal Name

COMMUNICATION

Spider prey-wrapping silk is an α-helical coiled-coil / β-Sheet hybrid nanofiber†

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

B. Addison,[a] D. Onofrei,[a] D. Stengel,[a] B. Blass,[b] B. Brenneman,[b] J. Ayon,[b] and G. P. Holland*[a]

Solid-State NMR results on ^{13}C-Ala/Ser and ^{13}C-Val enriched *Argiope argentata* prey-wrapping silk show that native, freshly spun aciniform silk nanofibers are dominated by α-helical (~50% total) and random-coil (~35% total) secondary structures, with minor β-sheet nanocrystalline domains (~15% total). This is the most in-depth study to date characterizing the protein structural conformation of the toughest natural biopolymer: aciniform prey-wrapping silks.

Spiders have evolved to produce up to 7 different types of silk with a wide range of impressive mechanical properties.[1,2] To date, the vast majority of structural data on as-spun silk fibers has been on dragline, or Major Ampullate (MA), silks due to their unmatched strength and ease of study. However, through a unique combination of high strength (~700 MPa) and high extensibility before breaking (~60-80%),[3] lesser-studied aciniform silk, utilized for prey-wrapping and egg-case lining, is actually the toughest of the spider silks and boasts mechanical properties that surpass the toughest man-made materials.[1,4] A complete molecular-level understanding of native aciniform silk fiber is lacking, although significant insight into its protein structure can be gained from the current consensus model for spider dragline fibers and from solution-NMR work conducted on recombinant aciniform spidroin 1 (AcSp1) protein. The high strength and moderate extensibility of dragline fibers is largely attributed to common structural motifs arising from short repetitive protein units; high fiber strength is thought to arise from aligned nanocrystalline β-sheet structures comprised of poly(Ala) and poly(Gly-Ala) and sometimes Ser, while Gly-Gly-X and Gly-Pro-Gly-X-X repeats contribute to fiber elasticity in the form of randomly-oriented

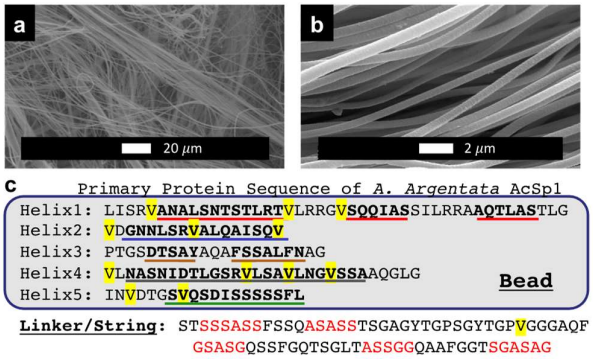


Fig 1. Scanning Electron Microscopy (SEM) images (a, b) of native *A. argentata* prey-wrapping silk bundles. Fibers are roughly 500 nm in diameter. (c) Primary-protein sequence for *A. argentata* AcSp1 repeat unit. Regions determined to be helical in solution are bolded and color-underlined based on structure homology between *A. trifaciata* (PDB Code 2MU3) and *A. argentata* (Supplemental). Val residues are highlighted in yellow and proposed β-sheet-forming motifs from the linker region are shown with red lettering.

domains and elastin-like type II β-turns.[1,2,5] On the other hand, the consensus sequence of AcSp1 is composed of a string of ~14 much longer repeats of about 200 amino acids, flanked by non-repetitive C- and N-terminal regions.[3,6] The AcSp1 sequence contains both Ser-rich and Ala-Ser rich motifs (Fig 1c, red lettering), but aciniform silks are entirely deficient in the traditional poly(Ala), poly(Gly-Ala), Gly-Gly-X and Gly-Pro-Gly-X-X repeats found in dragline spider silks.§ It was assumed through crude sequence-based structure predictions that AcSp1 repeats are likely to be rich in α-helices, and in 2011 this assumption was confirmed when Xu et. al published the liquids-NMR chemical shift assignments of a single 200-AA repeat unit for recombinant AcSp1 from *A. trifaciata* spiders.[7] A complete solution-state NMR structure of the *A. trifaciata* AcSp1 wrapping unit was later solved (PDB code 2MU3), where the authors convincingly showed that AcSp1 protein exists as a multi-domain "beads on a string" structure composed of a well-defined 5-helix globular domain

[a] Department of Chemistry and Biochemistry, San Diego State University, San Diego, CA, 92182-1030
[b] School of Molecular Sciences, Arizona State University, Tempe, AZ, 85287
BA and GH designed experiments, BA wrote the manuscript. All authors were involved in the experiments and had valuable contributions to the work.
†Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

(bead, helices 1-5) and a disordered linker domain (string).[8] Additionally, NMR structural and dynamical data showed that the Ser-rich terminal helix of the globular bundle, denoted helix-5 (Fig. 1c, green), is more dynamic, and thus possibly primed to form β-sheet nanostructures upon fiber aggregation.[8,9] While extremely enlightening, these studies were performed on recombinant silk protein in solution, thus probing the secondary protein structures of native as-spun aciniform silk fibers is largely lacking. Polarized Raman studies suggest that the silk protein dope in the gland is α-helix rich, while both fiber-aligned and α-helical and β-sheet structures are observed after fibrillization,[10,11] however these Raman data are not capable of obtaining amino-acid specific structural data nor directly correlating data to the primary protein sequence.

To gain a more detailed view into the molecular structure of native prey-wrap silks we utilized solid-state NMR techniques. *A. argentata* (this text) and *A. aurantia* (data in Supplemental) spiders were fed ~50 microliters of a saturated solution containing A) ^{13}C-labelled Ala (Ser is also labelled through metabolism of Ala), or B) ^{13}C-labelled Val every few days. Prey-wrap silk was collected by simulating prey using vibrating tweezers or a vibrating electric toothbrush (see Supplemental for detailed methods). These labelling schemes were chosen to highlight the two distinct domains found in the aciniform repeat sequence (Fig. 1c); namely that Ala and Ser are found dispersed throughout both the globular helical domain and the disordered linker region, while Val (and other hydrophobic residues) are found almost exclusively in the helical globular region (bead); notably in helices 1, 2, 4 and 5 but none found in helix 3 and only two found in the linker (Fig. S2).

Alanine: ^{1}H-^{13}C CP-MAS NMR data on ^{13}C-Ala / Ser enriched aciniform silk reveals a dominant Ala Cβ resonance at 16.3 ppm with a minor β-sheet shoulder at 21.0 ppm. The observed Ala Cβ chemical shift is suggestive of helical structure, but differentiating between random-coil (RC) and α-helical structures is more easily visualized by Cα chemical shifts; one expects a 2-3 ppm downfield shift for helical Cα resonances relative to RC.[12,13] Spectral overlap in the 1D data inhibits accurate extraction of chemical shifts, thus we collected two-dimensional (2D) ^{13}C-^{13}C through-space correlation experiments using dipolar-assisted rotational resonance (DARR) recoupling (Fig. 2c).[14,15] From the 2D DARR data we clearly see distinct Ala Cα / Cβ / CO dipolar-coupling cross-peaks. The dominant Cα / Cβ / CO cross-peaks (16.3 / 53.2 / 176.8 ppm) align with Ala adopting α-helical secondary structure. There are additional minor cross peaks at (21.0 / 49.3 / 172.8 ppm) and (17.5 / 50.5 ppm) representing Ala adopting β-sheet and random-coil structures, respectively. While one can visually interpret that the dominant Ala secondary structure is α-helical, we were interested in a more quantitative interpretation. Using precise chemical shift information extracted from 2D DARR data we could deconvolute the 1D CP-MAS data to estimate the percent representations of Ala adopting helical, β-sheet and RC secondary structures. For Ala, the Cβ resonance is the most well resolved and therefore is most likely to give confident fitting results. We find through deconvolution of the Ala Cβ resonance that

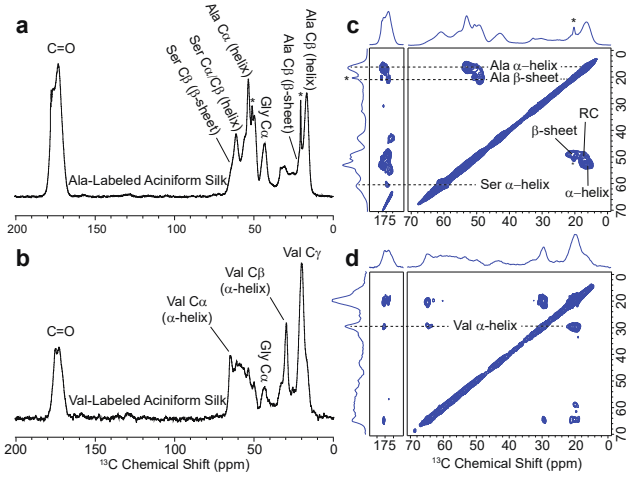


Fig 2. ^{1}H-^{13}C CP-MAS spectra (a, b) and 2D ^{13}C-^{13}C DARR (c, d) spectra of freshly-spun *A. argentata* aciniform silks that were isotopically enriched with either ^{13}C-Ala (a, c) or ^{13}C-Val (b, d). Signal from crystalline ^{13}C-Ala as a minor contaminant is indicated with an asterisk.

approximately 40% of all Ala residues adopt α-helical structures, with roughly 25-29% residing in β-sheet domains and the remaining 32-35% is unstructured or random coil (Fig. 3).‡

Serine: Similar to Ala, Ser amino acids are found distributed throughout both the globular (bead) domain and the disordered linker in the primary protein sequence of the AcSp1 repeat unit. Our solid-state NMR data tells a similar story to Ala, namely that ^{13}C chemical shifts of the dominant Ser resonances in native aciniform silks align with a true α-helical secondary structure with only minor β-sheet and RC subdomains present. We note that in the 2D DARR data, at the contour level displayed in Fig. 2c we do not see any Ser Cα / Cβ cross-peaks for β-sheet environments. However, if we increase the contour level to just below the noise threshold those correlations are observed. Hence, it is clear from a simple visual interpretation of the 2D DARR spectra that Ser β-sheet content is minimal compared to α-helical. To gain a more quantitative picture we peak-fit the seryl resonances. Ser is more difficult to peak fit than Ala due to substantial spectral overlap; Ser Cα and Cβ resonances in an α-helix, β-sheet, and RC conformations all reside between 55 - 65 ppm. However, since the chemical shifts for Ser β-sheet structures could be extracted from ^{13}C-^{13}C DARR data, we could at minimum estimate β-sheet versus non-β-sheet structures by deconvoluting the Ser Cβ resonance, which is the most downfield and best resolved and therefore most likely to yield accurate results (see Supplemental for further discussion on peak fitting). It is clear that for native prey-wrapping silk Ser is mostly α-helical (~50-55%) with minor β-sheet (~25-30%) and RC (~15-20%) content. More confidently, we can state that roughly 25-30% of Ser residues exist in β-sheet structures, while the remaining 70-75% is either helical (dominant) or unstructured (minor). We note that the most dominant seryl side-chain / carbonyl DARR cross-peak is at 174.0 ppm, align with α-helical secondary structures, further supporting that the dominant Ser environment is α-helical.

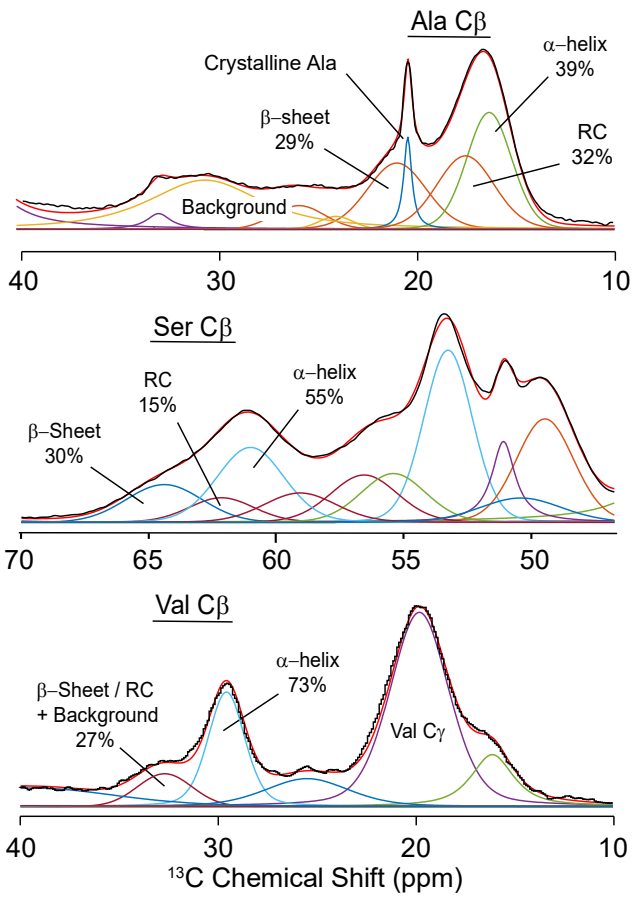


Fig 3. Spectral deconvolutions of Ala, Ser and Val Cβ resonances from ^{1}H-^{13}C CP-MAS data collected on isotopically-enriched freshly-spun *A. argentata* aciniform silks.

Valine: There are 13 total Val residues in the *A. argentata* AcSp1 repeat unit, 12 of which are found in the globular bead region and only one found in the linker (Fig. 1c). With Ala and Ser amino acids well represented in both the bead and string regions, we collected data on ^{13}C-Val enriched prey-wrapping silk to better understand the divide between the bead and linker after fibrillization (Fig. 2). Again, we extracted exact ^{13}C chemical shifts for the dominant Val Cα, Cβ, and CO resonances (Table S1), which as anticipated, align very clearly with α-helical secondary structure. We utilized the Val Cβ resonance to quantify helical versus non-helical structures because it is the most well resolved (Fig. 3c). Peak fitting results suggest that at minimum, 73% (likely 80-90%, see Supplemental)[§§] of all Val residues adopt an α-helical environment, with a minor RC and/or β-sheet component. We could not confidently assign the likely protein structure of the Val Cβ shoulder because no clear DARR correlations with Cα or CO are observed, thus the minor Val environment is assumed to be a combination of random-coil and/or β-sheet structures in addition to background signal from natural-abundant and partially-labelled signals (Fig. S6).

To make sense of these results we directly compared our quantifications to the primary protein sequence of the *A. argentata* AcSp1 repeat unit. Solid-state NMR data on silk biopolymers has repeatedly been used to quantify secondary structure content at the amino-acid level, and also to quantitatively correlate the primary amino acid sequence to said secondary structures.[16–19] When taken together, our peak-fitting results point to a protein structural model of native aciniform silk in which A) the majority of the silk fibroin exists as α-helices (~45-50%) or loosely-structured regions (~35%), B) the helical domains likely form some higher-order coiled-coil suprahelical structure, C) β-sheet aggregation occurs upon fibrilization, accounting for about 15% of the total fiber, likely from possible sheet-forming sequences in the linker domain, and D) the Poly(Ser) region of helix-5 remains helical or partly unstructured in the native silk. The arguments for this proposed model go as follows. To generate a complete model of AcSp1 in its final fibrous form we first consider that ~75-90% of all Val residues adopt true α-helical structures. Interestingly, the solution-NMR structure of the AcSp1 W unit (PDB code 2MU3)[8] identifies 6 Val residues in the "bead" region adopting a true helical conformation, 6 additional Val residues in the bead region that are on the ends of individual helices and are loosely helical or unstructured, and finally one clearly unstructured Val in the linker domain (Fig. 1c).[8] Considering this observation, it is clear that the loosely structured Val residues on the edges of helices are enticed to adopt well defined helices upon fiber formation. We also note that hydrophobic residues Val, Leu and Ile exist predominantly in helices 1, 2, 4 and 5, often in patterns expected for coiled-coil suprahelical structures,[20] while such hydrophobic-rich residues are nearly non-existent in helix 3 and in the linker domain (Fig. S2). With this data we can hypothesize which specific regions in the AcSp1 primary protein sequence adopt α-helical structures in the final fiber; we propose that helices 1, 2, 4 and 5 of the 5-helix bundle (bead) remain in a helical conformation as the fiber is pulled from the spider (Fig. 4), most likely as a some coiled-coil suprahelical structure where hydrophobic residues are buried in the coiled-coil core (Fig. S2, S3). We fully expect these helices to be fiber-aligned, as previously suggested.[10,11] In addition to Val information we can map onto the protein sequence our fitting results for Ala (~40% helical, 25-30% β-sheet) and Ser (~50-55% helical, 25-30% β-sheet) residues. From the primary protein sequence of *A. argentata* prey-wrapping silk repeat unit we identified 5 possible short stretches in the linker region and one from helix-5 that might be prone to form β-sheets similar to well-known sheet-forming motifs found in other common silks: SSSASS, ASASS, GSASG, ASSGG, SGASAG, and SSSSS. Indeed, these motifs (red lettering in Fig. 4) show a high propensity for β-sheet aggregation in our preliminary Monte-Carlo Molecular Dynamics (MD) simulations (Fig. S8, S9). From these regions we can estimate 7 Ala residues out of a total of 28 (25%) might form β-sheet nanostructures upon fibrillization, which agrees nicely with our experimental data (~25-30%). However, if we expect that all of the possible β-sheet prone Ser-rich motifs including the Poly(Ser) run in helix-5 indeed form β-sheets, then roughly 40% of all seryl residues should be β-sheet, 30% helical and 30% RC (Fig. S7). This prediction does not agree with our experimental results. We therefore revisited the repeat sequence

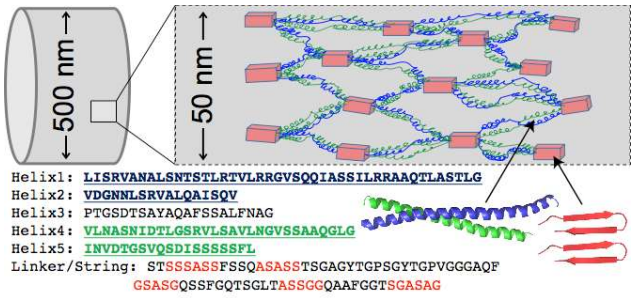


Fig 4. Proposed hierarchical molecular protein structure of aciniform prey-wrapping silks as a hybrid α-helical coiled-coil and nanocrystalline β-sheet fibroin. Helix 1 and 2 form a helical coil (blue), Helix 3 is unstructured, helices 4 and 5 form another coil (green), and the Ser / Ala-rich regions of the linker form pleated β-sheet (red) subunits.

and proposed multiple iterations of how the AcSp1 repeat might exist in its final fiber (Fig. S7). A sequence-based structure prediction that best agrees with our experimental NMR data is one in which the β-sheet aggregation occurs in the linker domain while the Poly(Ser) motif in helix-5 remains helical or loosely-structured (Fig. 4). If we also consider the density of hydrophobic residues Val, Leu and Ile in the globular region (found in helices 1, 2, 4, 5 but not in helix 3), and also the location of charged residues Arg (higher density in helices 1 and 2) and Asp (higher density in helices 4, 5), we can propose an antiparallel coiled-coil motif in which helices 1 and 2 form a single α-helical coil, helix 3 is more randomly-oriented, helices 4 and 5 form a second antiparallel coil, and finally, the Ser / Ala rich motifs at the beginning and end of the linker domains align to form antiparallel pleated sheets. This model is illustrated in Fig. 4. Significant work remains to support or modify the proposed hybrid nanofiber molecular architecture.

In conclusion, through our solid-state NMR data we have shown that Argiope spider prey-wrapping silks, which are actually the toughest biopolymer known yet are surprisingly understudied, are dominated by α-helical secondary structures (~40-50% of total fiber) with only minor β-sheet content (15% total). This is in loose agreement with Raman spectroscopy studies on aciniform fibers from both *A. aurantia* and *N. clavipes* spiders (~25-35% α-helices, ~30% β-sheets),[8,10,11] although in contrast to the authors conclusions, here we find that α-helix-to-β-sheet conversion is minimal, and prey-wrapping silk β-sheet content is low compared to other spider silks. Spectral deconvolutions of NMR data show that β-sheet nucleation indeed occurs during fibrillization, but this is likely from Ser- and Ala-rich domains found in the linker region while the helical domains remain largely in-tact. True α-helical silks are observed in other insect silks,[21] but the helical nature of spider prey-wrapping silk is a new discovery. Aciniform silks are capable of extending about twice as far before breaking compared to dragline fibers, a property that is most certainly a direct product of their high helical content.

This work was supported by grants from DOD-AFOSR (FA9550-17-1-0282) and DOD-AFOSR-DURIP (FA9550-17-1-0409) (GPH). DO was supported by a University Graduate Fellowship at SDSU.

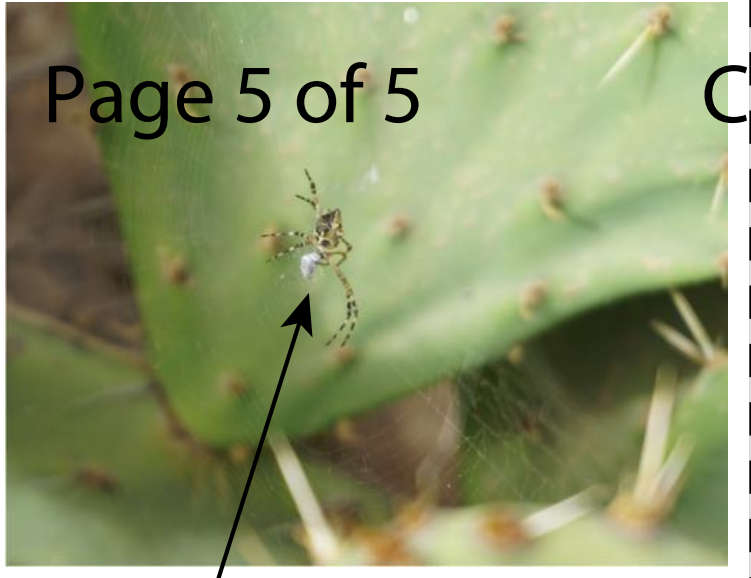
Conflicts of interest

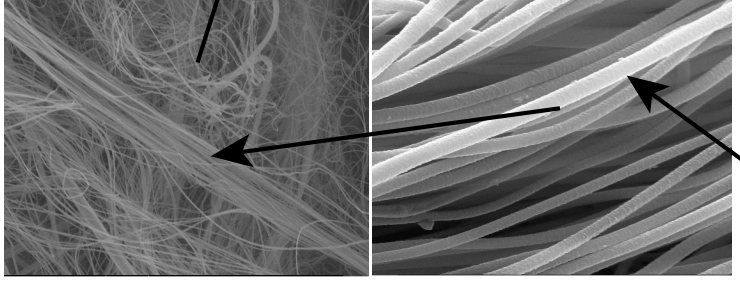
There are no conflicts to declare.

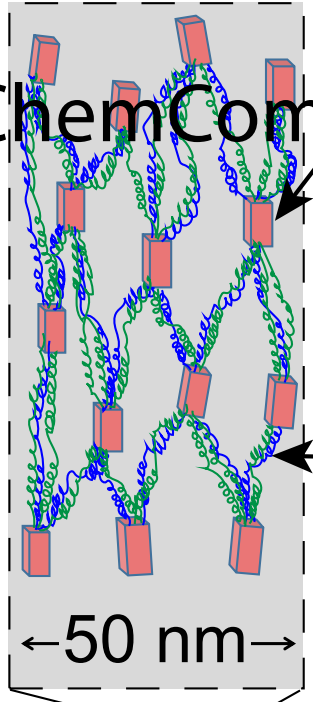
Notes and references

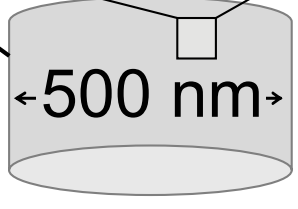
‡ See Supplemental for further discussion of peak-fitting results and error estimation.
§ Repeats found in major ampullate spidroin 1 (MaSp1) and/or MaSp2 sequences.
§§ 73% Is likely a lower-bound estimate due to overlap from natural abundant and partially labelled signals.

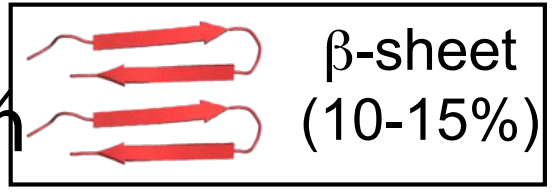
1 R. V Lewis, *Chem. Rev.*, 2006, **106**, 3762–3774.
2 L. Eisoldt, A. Smith and T. Scheibel, *Mater. Today*, 2011, **14**, 80–86.
3 C. Y. Hayashi, T. A. Blackledge and R. V. Lewis, *Mol. Biol. Evol.*, 2004, **21**, 1950–1959.
4 J. Gosline, M. Lillie, E. Carrington, P. Guerette, C. Ortlepp and K. Savage, *Philos. Trans. R. Soc. B Biol. Sci.*, 2002, **357**, 121–132.
5 J. L. Yarger, B. R. Cherry and A. Van Der Vaart, *Nat. Rev. Mater.*, 2018, **3**, 1–11.
6 R. Chaw, Y. Zhao, J. Wei, N. A. Ayoub, R. Allen, K. Atrushi and C. Y. Hayashi, *BMC Evol. Biol.*, 2014, **14**, 1–12.
7 L. Xu, M. L. Tremblay, Q. Meng, X. Q. Liu and J. K. Rainey, *Biomol. NMR Assign.*, 2012, **6**, 147–151.
8 M. L. Tremblay, L. Xu, T. Lefèvre, M. Sarker, K. E. Orrell, J. Leclerc, Q. Meng, M. Pézolet, M. Auger, X. Q. Liu and J. K. Rainey, *Sci. Rep.*, 2015, **5**, 1–15.
9 M. Sarker, K. E. Orrell, L. Xu, M. L. Tremblay, J. J. Bak, X. Q. Liu and J. K. Rainey, *Biochemistry*, 2016, **55**, 3048–3059.
10 M.-E. Rousseau, T. Lefèvre and M. Pézolet, *Biomacromolecules*, 2009, **10**, 2945–2953.
11 T. Lefèvre, S. Boudreault, C. Cloutier and M. Pézolet, *J. Mol. Biol.*, 2011, **405**, 238–253.
12 D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes, *J. Biomol. NMR*, 1995, **5**, 67–81.
13 Y. Wang and O. Jardetzky, *Protein Sci.*, 2002, **11**, 852–61.
14 K. Takegoshi, S. Nakamura and T. Terao, *Chem. Phys. Lett.*, 2001, **344**, 631–637.
15 K. Takegoshi, S. Nakamura and T. Terao, *J. Chem. Phys.*, 2003, **118**, 2325–2341.
16 G. P. Holland, J. E. Jenkins, M. S. Creager, R. V Lewis and J. L. Yarger, *Chem. Commun.*, 2008, **0**, 5568–5570.
17 J. E. Jenkins, M. S. Creager, R. V. Lewis, G. P. Holland and J. L. Yarger, *Biomacromolecules*, 2010, **11**, 192–200.
18 J. B. Addison, T. M. Osborn Popp, W. S. Weber, J. S. Edgerly, G. P. Holland and J. L. Yarger, *RSC Adv.*, 2014, **40**, 41301–41313.
19 J. E. Jenkins, S. Sampath, E. Butler, J. Kim, R. W. Henning, G. P. Holland and J. L. Yarger, *Biomacromolecules*, 2013, **14**, 3472–3483.
20 D. N. Woolfson, in *Fibrous Proteins: Structures and Mechanisms*, eds. D. A. D. Perry and J. M. Squire, Springer International Publishing, 2017, vol. 82, pp. 35–61.
21 T. D. Sutherland, S. Weisman, A. A. Walker and S. T. Mudie, *Biopolymers*, 2012, **97**, 446–454.

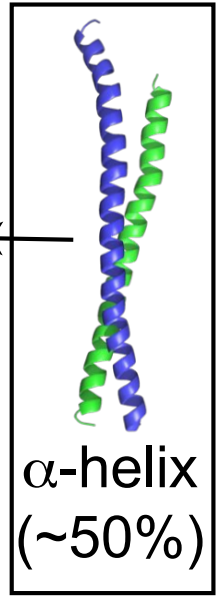


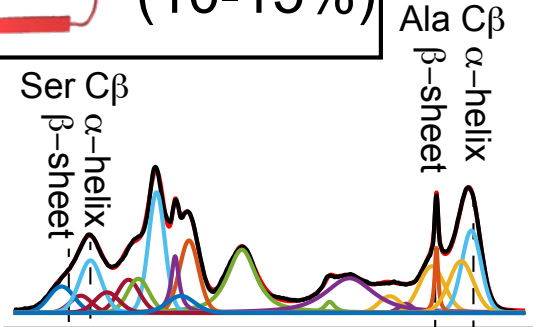


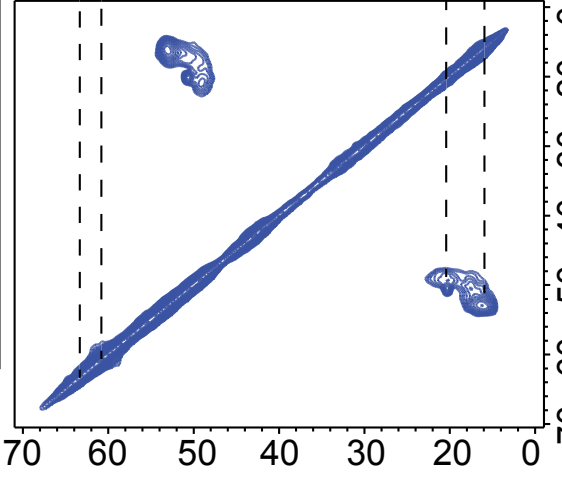

←50 nm→


←500 nm→


β-sheet
(10-15%)


α-helix
(~50%)


Ser Cβ
β–sheet
α–helix
Ala Cβ
β–sheet
α–helix


70
60
50
40
30
20
0
0
20
30
40
50
60
70