Supporting Information

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

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

Table of Contents

Materials and Methods

Silk Collection

Spiders were collected locally from along the coastal scrublands of San Diego, CA and kept in individual enclosures. In order to isotopically enrich the silk proteins in NMR-active nuclei, each the time silk was collected, the spiders were anesthetized using CO_2 gas, restrained and fed a saturated solution of either u-^{13}C Ala or u-^{13}C-Val (Cambridge Isotopes). The spiders were fed one cricket per week to ensure survival. Silk harvesting was done by exploiting the prey-wrapping behavior of Argiope orb-weavers, which produce wrapping silk in the presence of a vibrating stimulus, such as a pair of tweezers or electric tooth brush.[1] In our case a tooth brush resulted in the most wrapping silk produced. The silk was then carefully removed from the head of the toothbrush and placed in a pre-weighed centrifuged tube. Wrapping silk was obtained in this way until sufficient isotopically-enriched silk sample was obtained (~5 mg in the case of ^{13}C-Ala silks) or until the spider died. Unfortunately, our last spider died after we obtained about 1.5 mg of labeled Val silk. This proved to be sufficient.

Electron Microscopy

Silk samples were imaged on an FEI Quana 450 FEG Scanning Electron Microscope at San Diego State University, using 10kV accelerating voltage. Samples were coated with 8nm of platinum using an EMS 150 Sputter Coater.

Solid-State NMR Measurements

All measurements on *Argiope argentata* prey-wrapping silk were collected on a 600 MHz Bruker Avance IIIHD NMR spectrometer equipped with a 1.9 mm triple-resonance HCN MAS probe. Silk samples were center-packed in 1.9 mm Zirconia rotors using small amounts of Teflon tape. For 1D cross-polarization (CP-MAS) and direct-polarization (DD-MAS) experiments, samples were spun at the magic angle at 30 kHz, while for two-dimensional spin-diffusion NMR experiments the samples were spun at 14 kHz MAS. ^{13}C chemical shifts were referenced externally to TMS at 0 ppm by setting the downfield resonance of adamantane to 38.48 ppm.[2] ^{1}H – ^{13}C Cross-polarization MAS experimental conditions were as follows: hC.cp Bruker pulse program, a 2.5 μs initial proton $\pi/2$ pulse, a square 100 kHz cross-polarization contact pulse on the ^{13}C channel matched to the -1 Hartmann-Hahn sideband of a 1.75 ms ramped (~30%) CP spin-pock pulse on the proton channel, 120 kHz Proton decoupling (swftppm-13 Bruker decoupling pulse program) was used during 20 ms of acquisition time (2048 points collected, 300 ppm spectral width), 2048-10240 scan averages were collected, and the relaxation delay was set to 6 seconds so that all protons involved in CP were fully relaxed. For two-dimensional Dipolar-Assisted Rotational Resonance (2D DARR) experiments, data was collected using 14 kHz MAS, hCC.cp Bruker pulse program, 128 scans per slice, 1024 and 200 acquired points in the direct and indirect dimensions, respectively, a 200 ppm spectral width for both F1 and F2, a 2.5 second recycle delay, and 120 kHz proton decoupling (swftppm-13 Bruker decoupling pulse program) was applied during acquisition. To assist dipolar recoupling between ^{13}C nuclei, a 14 kHz (first rotational resonance condition) continuous wave radio frequency pulse was applied for 100 ms. All data were processed using Topspin version 3.5.

For silk samples from the species *Argiope aurantia*, data was collected with similar conditions but on a 400 MHz Varian VNMRS NMR Spectrometer using a 1.6 mm HXY triple-resonance MAS probe, and samples were spun at 35 kHz.

Solid-State NMR Peak Fitting and Analysis

Spectral deconvolution of solids NMR data is pivotal to this study, thus extreme care was taken when extracting precise chemical shifts and eliminating bias from fits. All NMR spectra were processed using 20 Hz exponential line broadening before Fourier transform, then spectra were phased and baseline corrected prior to peak deconvolution using DMFIT version dm2015vs release #20150521.[3] As stated in the main text, precise chemical shifts for Ala, Ser and Val resonances in various secondary structures were extracted from 2D DARR data. For these reasons we required high chemical shift resolution in our 2D data; 200 acquired points in the indirect dimension proved sufficient, although data took a long time to collect. Therefore, it was extremely important that chemical shifts were properly referenced and were stable over time. To compensate for magnetic field drift of the spectrometer, all experiments were referenced to Adamantane immediately before data collection, and additionally the magnetic field drift rate was previously measured over a long period of time, allowing us to apply a manual field correction under software control. Through this method chemical shifts were stable to 0.01 ppm within the course of each experiment.

Generally, the methodology was as follows:

1) Precise chemical shifts (down to 0.1 ppm) for relevant amino acids adopting clear secondary structures were extracted from high resolution 2D DARR spectra.

2) If exact ^{13}C shifts could not be determined from the 2D data, chemical shifts were locked at exactly the probability-based shifts published in Wang et. al, but with 1.8 ppm added due to the difference in referencing DSS vs TMS in the solid state.[4] For Alanine residues, our high-resolution 2D DARR data provided excellent confidence in exact chemical shifts of Ala Cα and Ala Cβ resonances adopting random-coil (RC), α-helical and β-sheet structures, and additionally we could confidently resolve Serine Cα and Cβ in β-sheet structures, thus using tabulated shifts was only needed for Ser Cα and Cβ resonances adopting RC and α-helical structures. Considering DARR-extracted Ser Cα and Cβ β-sheet shifts align exactly with tabulated values,[4] we think this assumption is more valid than allowing peak position to vary significantly during deconvolution.
3) When choosing starting conditions for spectral deconvolution, all signals of interest were initialized using the measured ^{13}C chemical shift, a 3 ppm half-width, a 0.9 %Gaussian / %Lorentzian initial lineshape ratio, and initial amplitudes were set with constant magnitudes to minimize bias by the researcher. These starting conditions were chosen based on one-dimensional extractions of 2D DARR data; signals were clearly more Gaussian than Lorentzian, and line-widths of extracted signals were always in the range of 2.5 – 3.5 ppm for full-width half max.
4) After all initial signals for relevant amino acids were added to the spectrum, we then needed additional background signals to account for natural abundant and partially labeled ^{13}C resonances, especially in the 25 – 40 ppm region. The choice of fitting background signals was arbitrary and up to the researcher performing the peak-fitting. Therefore to reduce researcher bias and to estimate error in reproducibility, four different researchers performed peak-fitting of the same NMR data without sharing fitting parameters.
5) Initially, only peak amplitude was allowed to vary where peak position, width, and shape were held constant. These results did not result in good fits, so next we allowed peak width to vary, but only one peak-type at a time. For example, all signals were allowed to vary peak amplitude, while both amplitudes and widths for Ala Cβ fits for RC, α-helix and β-sheet secondary structures were allowed to vary. This resulted in decent convergence of fits for the selected residue. Any fits that resulted in line-widths greater that 3.5 ppm were rejected.
6) After reasonable convergence was reached, we finally allowed peak position to vary for the final fit, but in addition to peak intensity and width, peak position was only allowed to vary within 0.1 ppm of the starting position. This resulted in excellent fits for Ala, Ser and Val Cβ resonances.
7) Spectral deconvolution for each silk sample was completed by four separate researchers, and results were averaged and deviations measures (Table S2). From these data we concluded that only Ala, Ser and Val Cβ resonances provided accurate quantifications. All other signals were nevertheless deconvoluted, but are only considered qualitatively.

Sequence Alignment

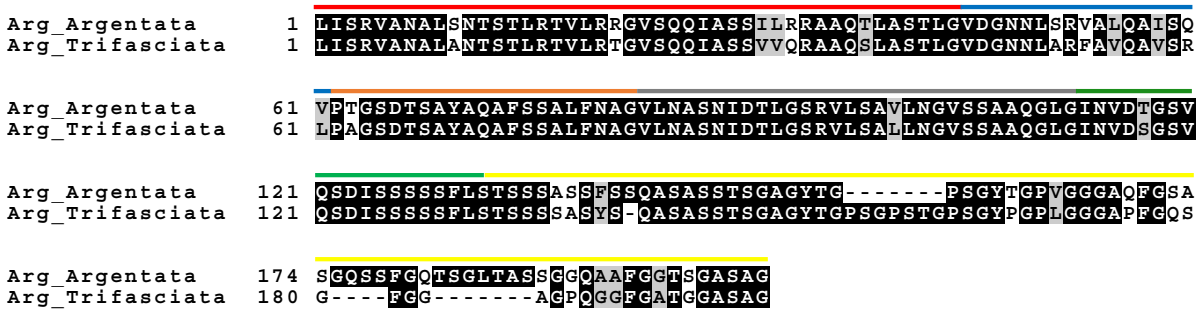


Figure S1: Primary protein sequence alignment comparing Argiope argentata and *Argiope trifasciata* repeat units. The primary protein sequence of AcSp1 shown in Fig. 1c in the main text is displayed with helical domains based on structural homology to the solution-NMR structure of a single recombinant AcSp1 repeat unit (PDB code 2MU3),[5] however the NMR structure is from *Argiope trifasciata* AcSp1 recombinant construct while the sequence information displayed in Fig 1c is from *Argiope argentata*, which is the species of interest in this study. Colored bars represent the solution-NMR "beads-on-a-string" model; red, blue, orange, grey and green bars respectively represent helices 1 – 5 of the globular domain, and the yellow bar represents the disordered linker/string domain. The sequences are 77.3% identical and 85.3% similar, highlighting a very high degree of conservation of the repeating units. The globular domain of AcSp1 repeat unit (residues 1-131) is nearly identical between *A. argentata* and *A. trifasciata* with only a few like-for-like substitutions, while the linker domain displays minor differences. Thus, we assume near-identical "bead-on-a-string" model in solution for *A. argentata* and *A. trifasciata*. Sequence alignment was performed using the EMBOSS Needle alignment tool using the Needleman-Wunsch algorithm.[6] Pairwise scoring was performed with the BLOSMUM62 scoring matrix with a pairwise alignment score of 10 for the first residue in a gap and a score of 0.5 for additional residues and score of 10 removed for end gaps with no penalty applied.

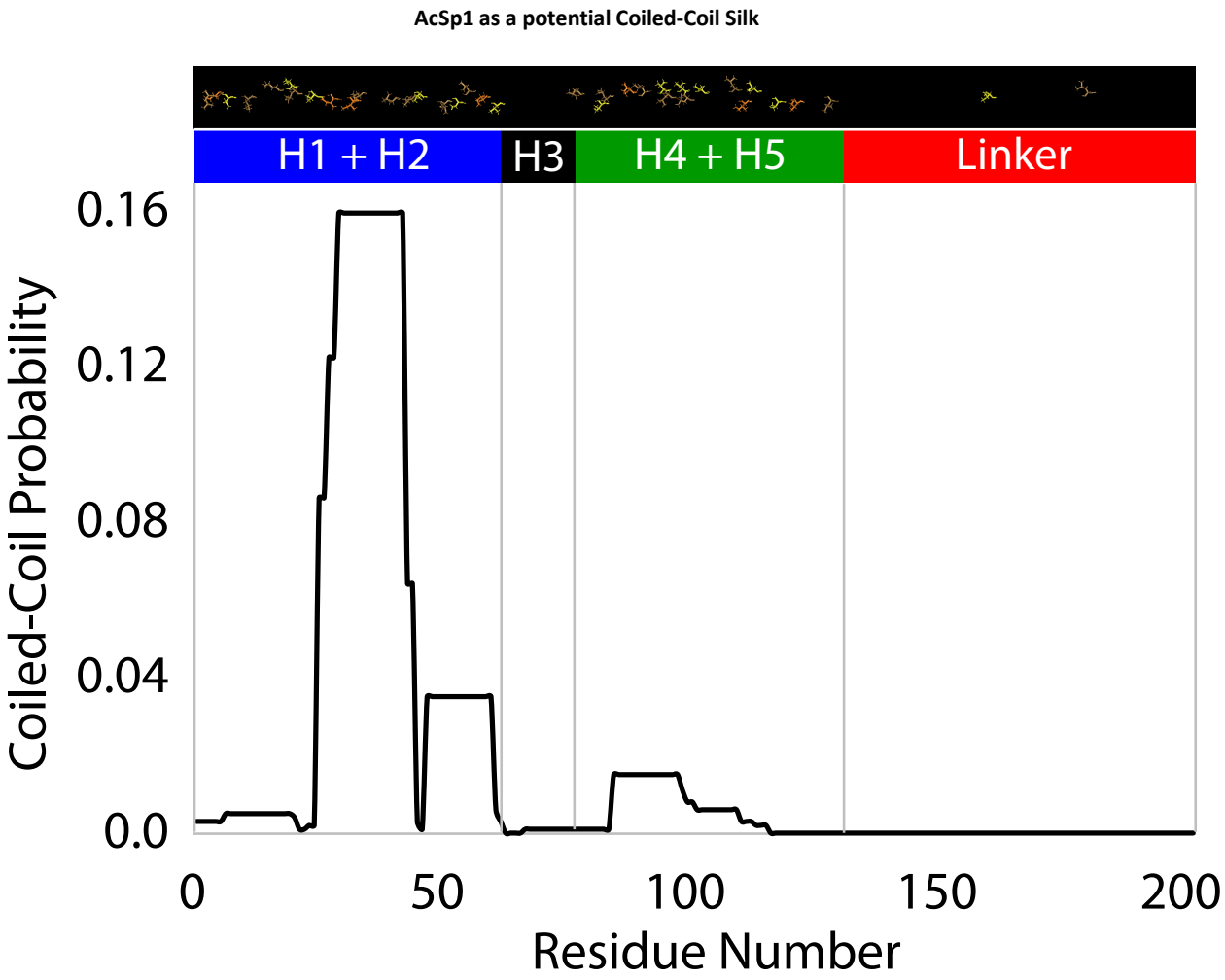


Figure S2: Prediction of coiled-coil probabilities of the entire AcSp1 repeat unit from *Argiope argentata* using the PCOIL algorithm (Web-based MPI Bioinformatics Toolkit, *https://toolkit.tuebingen.mpg.de/#/tools/pcoils*).[7]
The colored bars above the probability chart correspond to the following domains from the primary protein sequence: Helix1 + Helix2 (coil 1, blue bar), Helix 3 (black bar), Helix 4 + Helix 5 (coil 2, green bar), and Linker/String (red bar). Distribution of hydrophobic residues Val (yellow), Ile (sand) and Leu (orange) are also shown at the top. One can clearly see a clustering of hydrophobic residue in Helix1 + Helix2 (coil1) and in Helix 4 + Helix5 (coil2), which we propose interact in some sort of coiled-coil suprahelical structure. This prediction does not result in convincingly high coiled-coil probabilities, so perhaps ideal coiled-coil structures consisting of a tight knobs-into-holes packing structure are only loosely represented. Nevertheless, interhelical interactions are likely due to the closeting of hydrophobic residues.

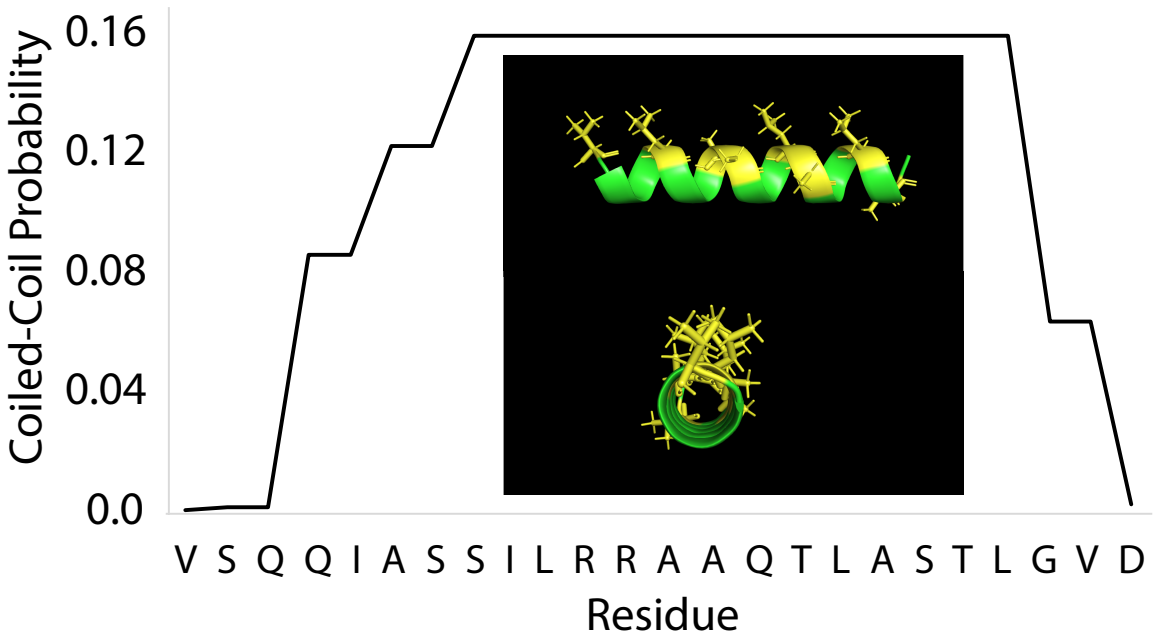


Figure S3: Prediction of coiled-coil probabilities of residues 23 – 46 from the AcSp1 repeat unit from *Argiope argentata*, along with a cartoon structure of the sequence modeled as an ideal α-helix. Hydrophobic residues Val, Ile, Leu and Ala are shown in yellow, while the rest of the sequence is green. Here we can see a clear alternating hydrophobic - polar (HHPP)x pattern, resulting in opposing hydrophobic and polar faces. Hydrophobic faces are expected to engage in inter-helix stabilizing van der Waals interactions in the native fiber, although the precise nature of this interaction is currently unclear.

Supplementary NMR Data

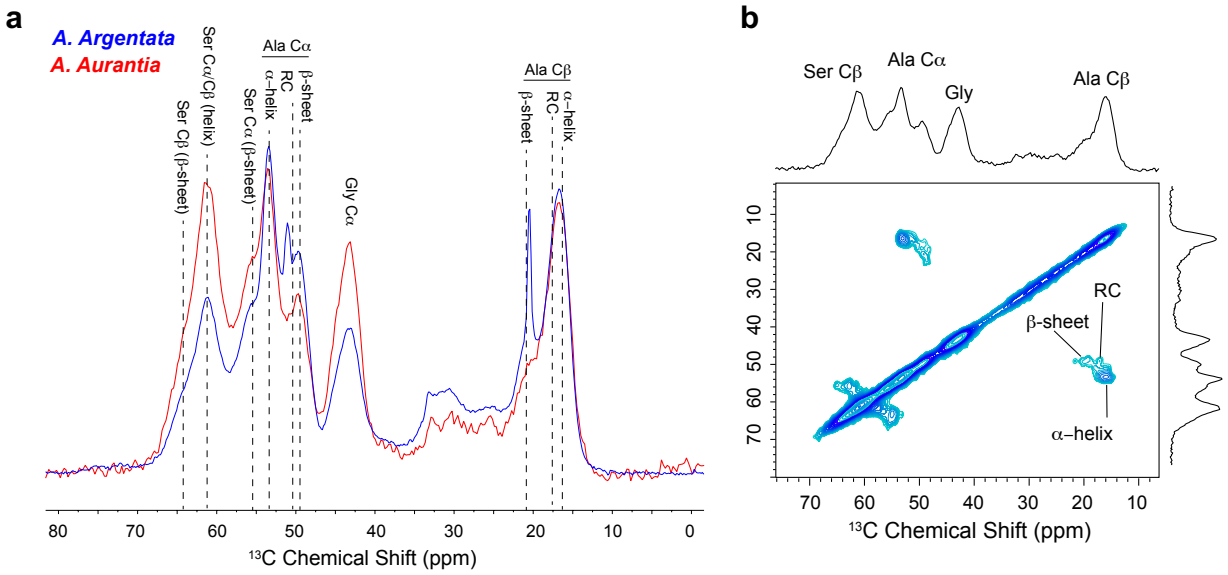
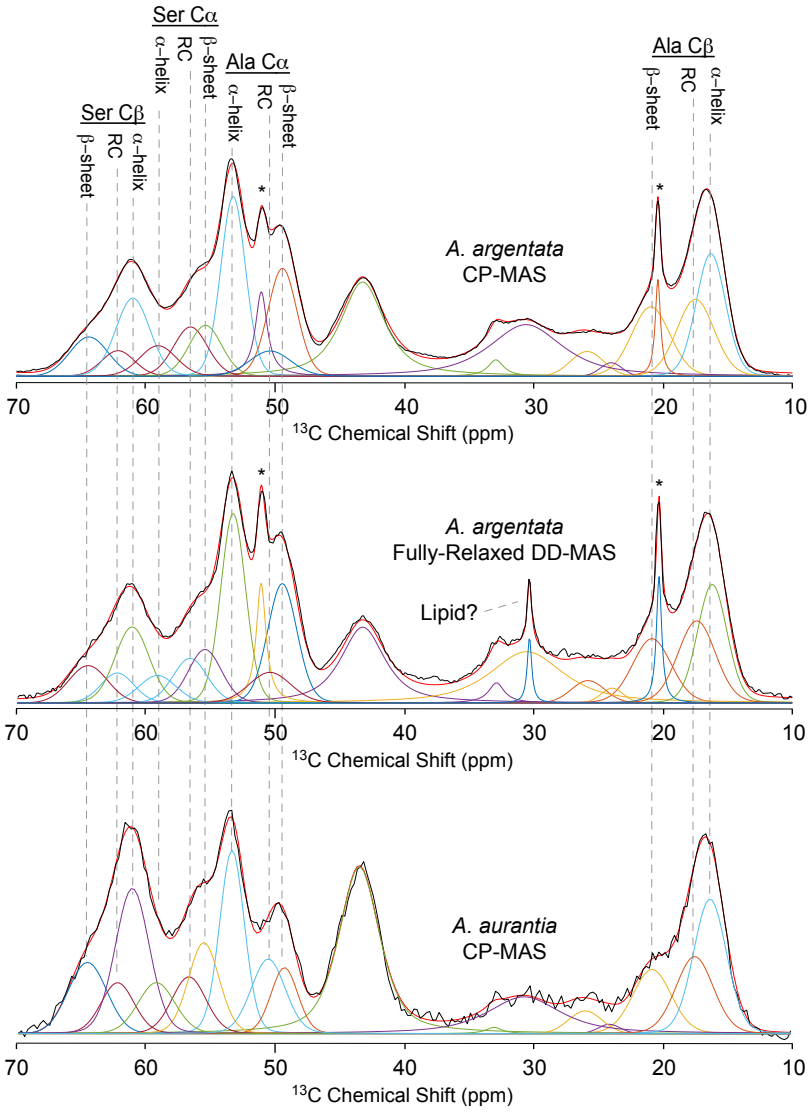


Figure S4: Solid-state NMR data on (A) ^{1}H-^{13}C CP-MAS data on native prey-wrapping silks from *Argiope argentata* (blue, 600 MHz Bruker AVIIIHD NMR Spectrometer, 30 kHz MAS) and *Argiope aurantia* (red, 400 MHz Varian VNMRS NMR Spectrometer, 35 kHz MAS) spiders. Data were collected on different instruments with slightly different Cross-Polarization conditions, thus it is no surprise that intensities from various ^{13}C sites are not superimposable. However, we see here that chemical shifts and lineshape are identical. This is clearer when CP-MAS data are deconvoluted using DMFIT (Figure S5). (B) ^{13}C-^{13}C DARR data (150 ms DARR mixing time) on native *Argiope aurantia* prey-wrapping silks. Both 1D and 2D spectra are extremely similar between the two species. AcSp1 sequence information is not published for *A. aurantia* silks, thus *A. argentata* data was used for the main discussion.



A. argentata CP-MAS

Residue	Structure	Shift (ppm)	% Comp.
Ala Cβ	α–helix	16.34	39%
	β–sheet	20.98	29%
	RC	17.53	31%
Ser Cβ	α–helix	61.01	54%
	β–sheet	64.40	29%
	RC	62.15	17%

A. argentata DD-MAS

Residue	Structure	Shift (ppm)	% Comp.
Ala Cβ	α–helix	16.34	39%
	β–sheet	20.98	27%
	RC	17.53	34%
Ser Cβ	α–helix	61.01	52%
	β–sheet	64.40	27%
	RC	62.15	20%

A. aurantia CP-MAS

Residue	Structure	Shift (ppm)	% Comp.
Ala Cβ	α–helix	16.34	39%
	β–sheet	20.98	26%
	RC	17.53	35%
Ser Cβ	α–helix	61.01	54%
	β–sheet	64.40	29%
	RC	62.15	17%

Figure S5: Peak fitting results of Ala and Ser Cβ resonances from solid-state NMR collected on native prey-wrapping silks from *Argiope argentata* (top, CP-MAS, middle, DD-MAS) and *Argiope aurantia* (bottom, CP-MAS) spiders. Peak positions of Ala and Ser Cα and Cβ residues adopting RC, α-helical and β-sheet structures are indicated with dotted lines, and the crystalline Ala contaminant is indicated with an asterisk. *A. argentata* data was collected with a 600 MHz Bruker AVIII NMR spectrometer at 30 kHz MAS, while *A. aurantia* data was collected with a 400 MHz Varian VNMRS NMR spectrometer at 35 kHz MAS. For the data presented above, fitting results are those of a single researcher. Averages and standard-deviations of peak-fitting results from the same data performed by 4 separate researchers are summarized in Table S2. Chemical shifts and peak fitting results are extremely similar between the two species. Notably, *Argiope argentata* and *Argiope aurantia* prey-wrapping silks yield nearly identical fitting results for Ser and Ala Cβ resonances, although the presence of crystalline Ala contaminant in *A. argentata* silk adds uncertainty. This contaminant was observed repeatedly even after adjusting silk harvesting methods, so we believe it was deposited onto the silk by the spider, most likely from a biting event during or after attack-wrapping. Only the Ala Cβ and Ser Cβ resonances are considered for quantitative analysis because they are the most well resolved and therefore most likely to yield accurate quantifications (See Table S2). Also, to ensure that our Cross-Polarization NMR data in fact yield quantitative results, we collected fully-relaxed ^{13}C Direct-Detect (DD-MAS, 30 second recycle delay) data on *A. argentata* silk fibers, which indeed yielded nearly identical results.

	Carbon-13 Chemical Shift (ppm, from TMS)				
Residue	**Acin. Silk (*A. aurantia*)**	**Acin. Silk (*A. argentada*)**	**α-helix**	**β-sheet**	**RC**
Ala Cα	53.2 [49.3]	53.2 [49.3]	52.3 - 52.8	48.2 - 49.3	50.8
Ala Cβ	16.3 [21.0, 17.5]	16.3 [21.0, 17.5]	14.8 - 16.6	19.9 - 20.7	17.4
Ala CO	176.6	176.8 [172.8]	176.2-176.8	171.6 - 172.4	176.1
Ser Cα	61 - 63, [55.4]	60.5 - 61.5 [55.4]	60.3 - 62.3	55.5 - 58.0	56.6
Ser Cβ	61 - 63 [64.4]	59 - 61.5 [64.4]	62.0 - 63.8	63.5 - 67.3	62.1
Ser CO	174.0 [170.9]	174.0 [171.3]	174.6	172.9	171.5
Val Cα	NA	64.8 [?]	64.3	59.0	60.1
Val Cβ	NA	29.6 [?]	28.7	32.4	31.2
Val CO	NA	175.6 [?]	174.9	171.8	174.6

Table S1: ^{13}C Chemical shifts observed for Ala, Ser and Val residues from native aciniform (prey-wrapping) silks collected from either *Argiope aurantia* or *Argiope argentata* spiders, compared to shifts from polypeptides adopting known secondary structures.[8–10] and from average chemical shifts of many published solution NMR structures.[4] Major components are indicated without brackets while chemical shifts for minor components are in brackets. The data clearly indicates that the dominant secondary structures for Ala, Ser and Val residues is helical, with minor β-sheet and/or random-coil subdomains. Precise chemical shifts were extracted from two-dimensional ^{13}C-^{13}C Dipolar-Assisted Rotational Resonance (DARR) data on isotopically-enriched prey-wrapping silk. All chemical shifts were referenced indirectly to TMS by setting the downfield resonance of adamantane to 38.48 ppm.[2]

			A. argentata Native Ala/Ser CP-MAS			*A. argentata* Native Ala/Ser DD-MAS			*A. aurantia* Native Ala/Ser CPMAS		
Residue	Structure	Shift (ppm)	FWHM	FWHM STD	% Comp	% Comp STD	FWHM	FWHM STD	% Comp	% Comp STD	FWHM	FWHM STD	% Comp	% Comp STD
Ala Cβ	RC	17.5	3.34	0.22	32.0%	1.3%	3.44	0.04	33.3%	0.7%	3.44	0.04	32.6%	3.7%
Ala Cβ	α-helix	16.3	2.70	0.02	39.8%	2.0%	2.70	0.02	38.1%	1.2%	2.86	0.17	43.1%	2.4%
Ala Cβ	β-sheet	21.0	3.33	0.27	28.2%	1.0%	3.45	0.17	28.6%	1.8%	3.40	0.27	24.4%	2.1%
Ala Cα	RC	50.4	2.79	1.03	9.9%	5.7%	2.80	1.31	12.0%	6.0%	3.13	0.29	29.1%	6.7%
Ala Cα	α-helix	53.3	2.32	0.12	52.6%	3.7%	2.31	0.13	57.3%	15.1%	2.16	0.17	48.7%	5.3%
Ala Cα	β-sheet	49.5	2.59	0.10	37.5%	4.5%	2.65	0.03	30.7%	17.3%	2.50	0.08	22.2%	2.9%
Ser Cα	RC	56.6	3.03	0.13	49.8%	17.1%	3.09	0.07	48.0%	18.6%	3.35	0.43	29.5%	6.5%
Ser Cα	α-helix	59.1	2.91	0.35	18.2%	6.5%	3.00	0.42	20.8%	4.0%	3.00	0.42	24.8%	5.2%
Ser Cα	β-sheet	55.4	2.36	0.78	32.1%	17.0%	2.51	0.85	31.2%	14.6%	2.98	0.35	45.6%	9.4%
Ser Cβ	RC	62.2	2.79	0.64	16.5%	5.3%	3.08	0.15	20.6%	1.9%	3.08	0.15	19.5%	3.0%
Ser Cβ	α-helix	61.0	2.96	0.06	55.5%	5.1%	3.04	0.08	51.1%	0.8%	3.04	0.08	53.0%	1.1%
Ser Cβ	β-sheet	64.4	3.44	0.15	28.0%	2.4%	3.41	0.12	28.3%	1.9%	3.41	0.12	27.6%	2.5%

Table S2: Averages and Standard-Deviations (STD) are reported for peak widths (full-width half max, FWHM, in ppm) and percent composition (% Comp) for each residue using DMFIT software performed independently by four different researchers (BA, DO, DS, GH). All spectra were fit using identical NMR chemical shifts (extracted from 2D DARR data). Even when not allowing peak position to vary at all, there are significant remaining variables that can dramatically influence fitting results if one is not careful, including peak intensity, peak width, peak lineshape (percent Gaussian vs percent Lorentzian), choice of background signals, and even starting conditions. The averages and standard deviations suggest that Ala and Ser C quantifications cannot be trusted and should therefore only yield qualitative interpretation. However, since Ala C and Ser C resonances are well resolved, four researchers independently arrived at very similar fitting results. All discussion on residue quantifications therefore only consider C resonances.

Table S2: Average spectral deconvolution results for solid-state NMR data on isotopically-enriched native prey-wrapping silks.

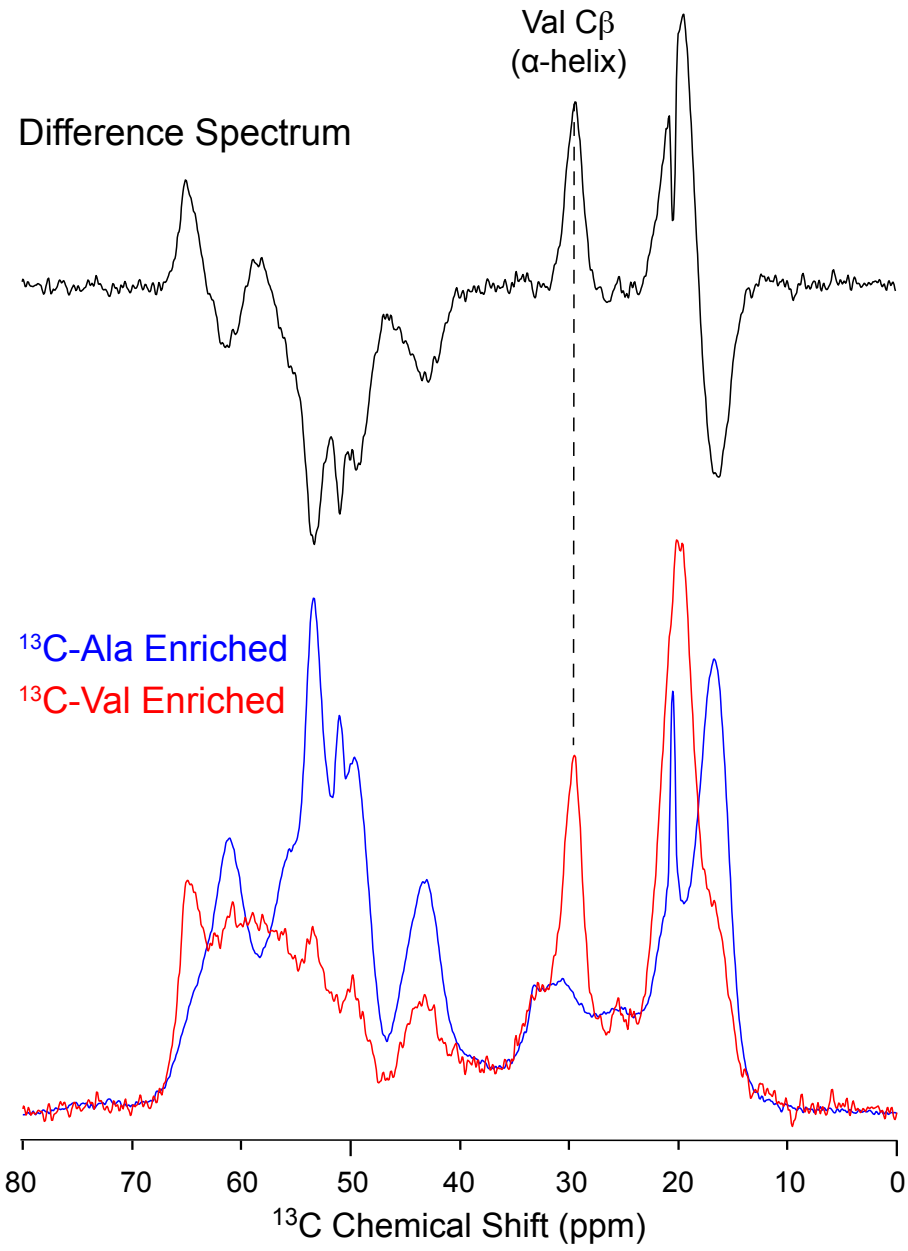


Figure S6: ^{13}C CP-MAS of ^{13}C-Ala/Ser (blue) and ^{13}C-Val (red) labeled *Argiope argentata* prey-wrapping silk, and the difference spectrum (black) when data are scaled to account for the background signal between 25-35 ppm. The shoulder at 32.4 ppm in the ^{13}C-Val labeled sample is likely from both background signal (natural-abundance and from other partially-enriched residues) and from Val residues adopting either random-coil and/or β-sheet secondary structures. The peak-fitting results suggest that at minimum, Val residues are 75% α-helical. If the Val Cβ shoulder near 32 ppm is entirely from background signal, then Val residues could be upwards of 90-100% α-helical in the native silks. Notably, no dipolar cross-peaks were visible in the 2D DARR experiment for the shoulder near 32 ppm, suggesting that signal might be mostly background. The reality is likely somewhere in between. If we predict a coiled-coil model, we would expect 11-12 out of 13 Val residues to adopt a-helical structures (85-92%) and 1-2 of 13 Val residues in the linker domain would likely adopt RC structures (8-15%).

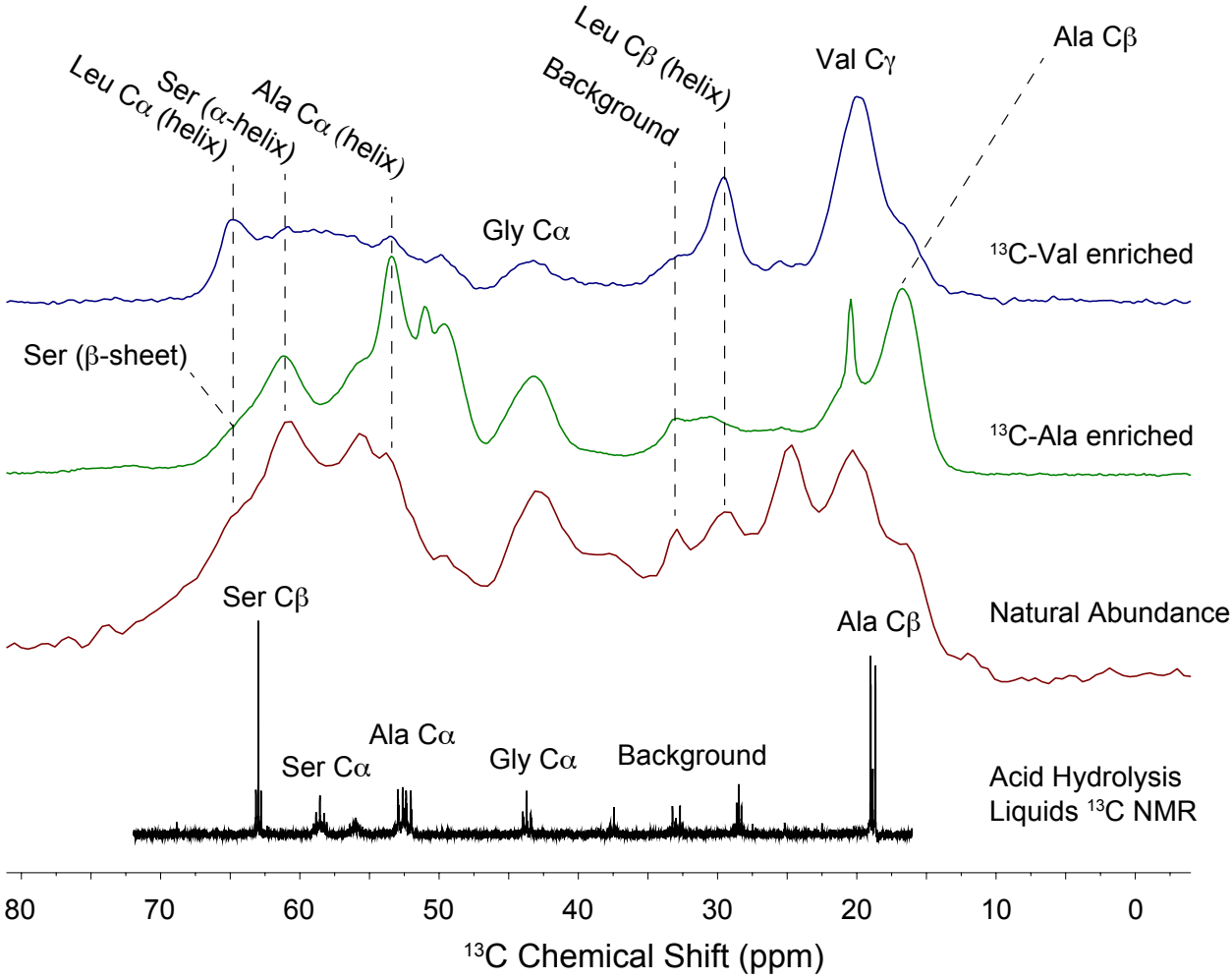


Figure S7: ^{13}C CP-MAS of natural abundance prey-wrapping silk compared to silk that had been labeled with ^{13}C-Ala/Ser or ^{13}C-Val. Spectra are also compared to standard ^{13}C solution-state NMR data on acid hydrolyzed ^{13}C-Ala-labeled silk. Labeled samples are from *Argiope argentata* spiders while natural abundance spectrum was obtained from *Argiope aurantia* spiders. Here we see further evidence that the shoulder at 32.4 ppm in the ^{13}C-Val labeled sample is likely from both background signal (natural-abundance and partial labeling) and from Val residues adopting either random-coil and/or β-sheet secondary structures. Additionally, we see that natural abundant signals from Leu, Val and Ile may overlap with Ser and Ala peak fitting results. For example, the strong Val C signal at 20 ppm overlaps with Ala C in a β-sheet secondary structure, indicating that Ala C peak deconvolution results are possibly a high estimate. However, since Ala and Ser Cβ signals are about 20-30% isotopically enriched, and there are no background signals in these regions in the ^{13}C NMR spectrum of acid-hydrolyzed silk, clearly the background signal for Ala and Ser deconvolutions are minimal. Amino acid incorporation was measured through solution-NMR analysis of acid hydrolysis according to methods by Shi, 2013.[11] Data is not shown.

Sequence-Based Structure Predictions

After aggregation: PREDICTION 1
Helix1: **LISRVANALSNTSTLRTVLRRGVSQQIASSILRRAAQTLASTLG**
Helix2: **VDGNNLSRVALQAISQ**V
Helix3: PTGSDTSAYAQAFSSALFNAG
Helix4: **VLNASNIDTLGSRVLSAVLNGVSSAAQGLG**
Helix5: **INVDTGSVQSDISSSSSFL**
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

Residue	Total	% RC	% Helical	% Sheet
Ala	28	32	43	25
Ser	47	28	43	30
Val	13	15	85	0
All	204	33	53	13

After aggregation: PREDICTION 2
Helix1: **LISRVANALSNTSTLRTVLRRGVSQQIASSILRRAAQTLASTLG**
Helix2: **VDGNNLSRVALQAISQ**V
Helix3: PTGSDTSAYAQAFSSALFNAG
Helix4: **VLNASNIDTLGSRVLSAVLNGVSSAAQGLG**
Helix5: **INVDTGSV**QSDISSSSSFL
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

Residue	Total	% RC	% Helical	% Sheet
Ala	28	32	43	25
Ser	47	28	32	40
Val	13	15	85	0
All	204	36	48	16

After aggregation: PREDICTION 3
Helix1: **LISRVANALSNTSTLRTVLRRGVSQQIASSILRRAAQTLASTLG**
Helix2: **VDGNNLSRVALQAISQ**V
Helix3: PTGSDTSAYAQAFSSALFNAG
Helix4: **VLNASNIDTLGSRVLSAVLNGVSSAAQGLG**
Helix5: **INVDTGSVQSDISSSSSFL**
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

Residue	Total	% RC	% Helical	% Sheet
Ala	28	36	43	21
Ser	47	32	43	26
Val	13	15	85	0
All	204	36	53	11

After aggregation: PREDICTION 4
Helix1: **LISRVANALSNTSTLRTVLRRGVSQQIASSILRRAAQTLASTLG**
Helix2: **VDGNNLSRVALQAIS**QV
Helix3: PTGSDTSAYAQAFSSALFNAG
Helix4: VLN**ASNIDTLGSRVLSAVLNGVSSAAQGLG**
Helix5: **INVDTGS**VQSDISSSSSFL
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

Residue	Total	% RC	% Helical	% Sheet
Ala	28	36	43	21
Ser	47	32	32	36
Val	13	23	77	0
All	204	41	46	13

After aggregation: PREDICTION 5
Helix1: LISR**VANALSNTSTLRTVL**RRG**VSQQIASS**ILRRA**AQTLAS**TLG
Helix2: **VDGNNLSRVALQAISQV**
Helix3: PTGS**DTSAY**AQA**FSSALFN**AG
Helix4: **VLNASNIDTLGSRVLSAVLNGVSSA**AQGLG
Helix5: INVDTGSVQ**SDISSSSSFL**
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

Residue	Total	% RC	% Helical	% Sheet
Ala	28	32	43	25
Ser	47	23	47	30
Val	13	23	77	0
All	204	41	46	13

Before Fibrilization, from Solution-NMR Structure
Helix1: LISRV**ANALSNTSTLRT**VLRRGV**SQQIAS**SILRRA**AQTLAS**TLG
Helix2: VD**GNNLSRVALQAISQV**
Helix3: PTGS**DTSAY**AQA**FSSALFN**AG
Helix4: VL**NASNIDTLGSRVLSAVLNGVSSA**AQGLG
Helix5: INVDTG**SVQSDISSSSSFL**
Linker/String: STSSSASSFSSQASASSTSGAGYTGPSGYTGPVGGGAQFGSASGQSSFGQTSGLTASSGGQAAFGGTSGASAG

Residue	Total	% RC	% Helical	% Sheet
Ala	28	57	43	0
Ser	47	53	47	0
Val	13	54	46	0
All	204	56	44	0

Figure S8: Five sequence-based structure predictions of the *Argiope argentata* wrapping silk repeat unit after fiber aggregation (above horizontal line), with solution-NMR conformation added for context (below horizontal line, see Fig. 1c in main text). Residues that are bolded and underlined are helical, while residues in red lettering are proposed to form β-sheets upon fibrilization. For each variant, tables are shown summarizing the total number of Ala, Ser, and Val residues adopting random-coil (RC), α-helical and β-sheet structures, as well as a sum total of all residues. A green checkbox or a red cross indicates which predictions agree or disagree with experimental NMR data, and the size of those icons loosely indicates our confidence in agreement or disagreement.

Our solid-state NMR data requires that the following conditions must be met for the molecular protein structure of native prey-wrapping silks after fibrilization: 1) Roughly 25-30% of all Ala and Ser residues should adopt β-sheet secondary structures; 2) For the remaining 70-75%, α-helical structures should outweigh random-coil; and 3) Val residues are roughly 75 – 90% α-helical, with only minor random-coil and/or β-sheet structures present. Predictions 1 - 4 in Figure S7 propose a coiled-coil model with varying degrees of β-sheet aggregation in the linker or helix-5 region. Prediction 5, which also agrees with quantitative NMR results for Ala, Ser and Val residues, is essentially the solution-NMR structure but with helices extended to include leading and trailing Val residues; this was necessary for Val quantifications to agree. While other possibilities may exist, the sequence-based

structure prediction that best fits the experimental NMR data is Prediction 1, where Poly(Ser) motif remains α-helical and β-sheet aggregation occurs in the linker domain. These predictions can be further improved as we collect quantitative NMR data for other relevant amino acids. These efforts are ongoing in our laboratory, but are not yet completed at time of publication.

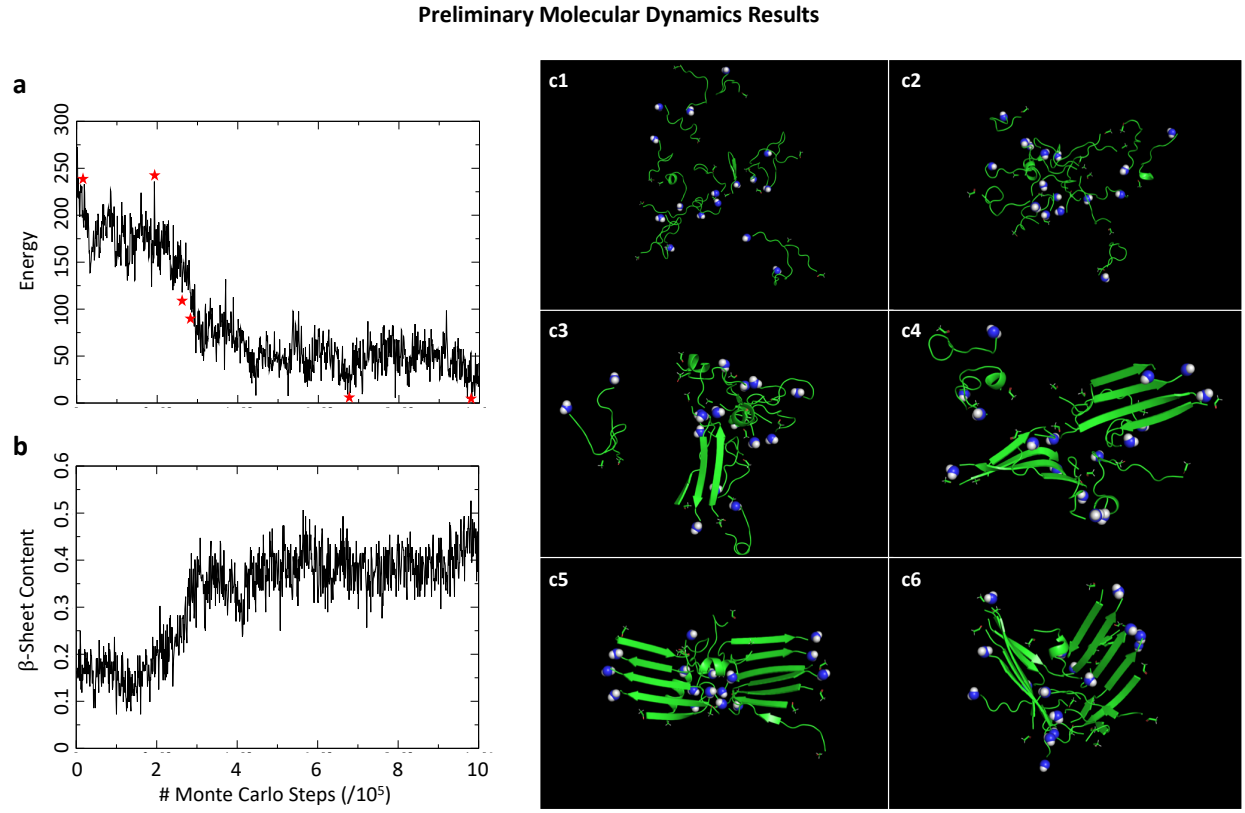


Figure S9: Preliminary Monte-Carlo Simulation results on possible β-sheet forming domains from the AcSp1 linker, in this case 8x Acetyl-ST**SSSASS**FSS-NH2 (chain 1) and 8x Acetyl-AAFGGT**SGASAG**-NH2 (chain 2), representing the beginning and ending β-sheet prone motifs in the linker domain. Acetyl and NH_2 capping groups were included to eliminate unwanted charge-induced stabilization. Monte-Carlo evolution of the energies (a) and β-sheet content (b) are shown. Red stars indicate timepoints for snapshots in c1 – c6. In this particular simulation, β-sheet nucleation occurs between two strands chain 1 and chain 2 (c3), which then seeds larger scale aggregation. Monte-Carlo simulations were conducted using the software package PROFASI version 1.5, using 16 total chains (see above) in a 60 Angstrom box at 298 Kelvin. 1,000,000 Monte-Carlo cycles were executed, corresponding to 1,000,000,000 elementary Monte-Carlo updates. The simulation was performed utilizing the San Diego State Computing Cluster. All images were rendered in Pymol version 2.1.1. By no means should these computational results be considered conclusive or exhaustive, but they do support our hypothesis that the Ala-Ser rich motifs in the linker domain are responsible for β-sheet aggregation upon AcSp1 fibrilization.

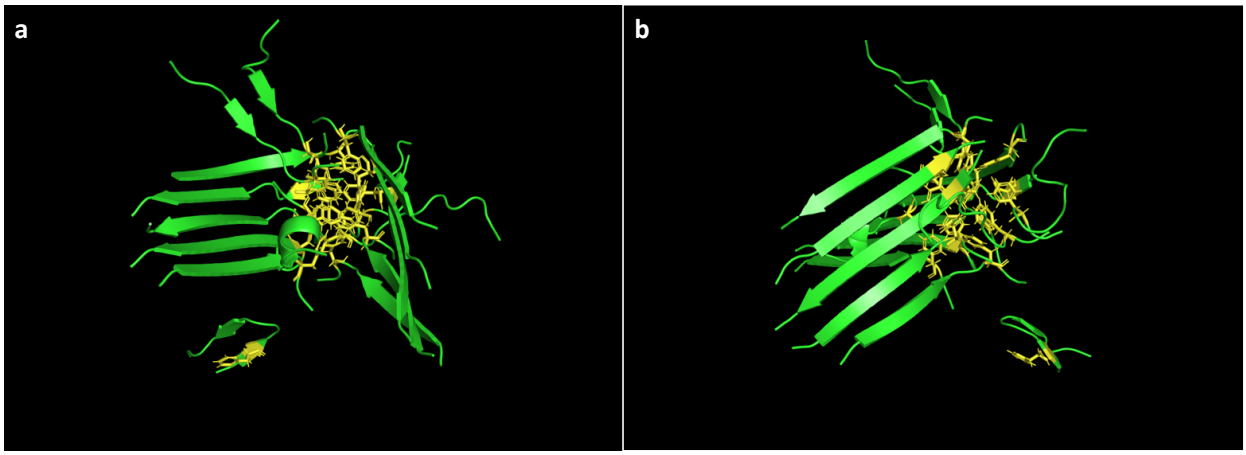


Figure S10: Representative minimum energy structure demonstrating AcSp1 linker aggregation from the MC simulation displayed above (Figure S8 Frame c6), with Phe residues highlighted in yellow. Images (a) and (b) are rotated 90 degrees about the y axis. We can see both parallel and anti-parallel β-sheets from this simulation, although in all cases, we see clustering of aromatic Phe residues. Interestingly, potential β-sheet forming sequences in the linker domain are flanked by aromatic Phe residues. These results may suggest that clustering of Phe residues on the terminal ends of Ala-Ser rich runs could be involved in β-sheet nucleation and/or stabilization. Images were rendered in Pymol version 2.1.1.

References

1 M. H. Robinson, *Am. Zool.*, 1969, **9**, 161–173.
2 C. R. Morcombe and K. W. Zilm, *J. Magn. Reson.*, 2003, **162**, 479–486.
3 D. Massiot, F. Fayon, M. Capron, I. King, S. Le Calvé, B. Alonso, J. O. Durand, B. Bujoli, Z. Gan and G. Hoatson, *Magn. Reson. Chem.*, 2002, **40**, 70–76.
4 Y. Wang and O. Jardetzky, *Protein Sci.*, 2002, **11**, 852–61.
5 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.
6 S. B. Needleman and C. D. Wunsch, *J. Mol. Biol.*, 1970, **48**, 443–453.
7 M. Gruber, J. Söding and A. N. Lupas, *J. Struct. Biol.*, 2006, **155**, 140–145.
8 A. Shoji, T. Ozaki, H. Saita, R. Tabeta and I. Ando, *Macromolecules*, 1984, **17**, 1472–1479.
9 H. R. Kricheldor and D. Müller, *Macromolecules*, 1983, **16**, 615–623.
10 D. S. Wishart, C. G. Bigam, A. Holm, R. S. Hodges and B. D. Sykes, *J. Biomol. NMR*, 1995, **5**, 67–81.
11 X. Shi, J. L. Yarger and G. P. Holland, *Anal. Bioanal. Chem.*, 2013, **405**, 3997–4008.