Biomol NMR Assign (2012) 6:147–151
DOI 10.1007/s12104-011-9344-z

ARTICLE

^{1}H, ^{13}C and ^{15}N NMR assignments of the aciniform spidroin (AcSp1) repetitive domain of *Argiope trifasciata* wrapping silk

Lingling Xu · Marie-Laurence Tremblay · Qing Meng · Xiang-Qin Liu · Jan K. Rainey

Received: 12 August 2011 / Accepted: 27 September 2011 / Published online: 12 October 2011

Abstract Spider silk is one of nature's most remarkable biomaterials due to extraordinary strength and toughness not found in today's synthetic materials. Of the seven types of silk, wrapping silk (AcSp1) is the most extensible of the types of silks and has no sequence similarity to the other types. Here we report the chemical shifts for the AcSp1 199 amino acid protein repeat unit and its anticipated secondary structure based on secondary chemical shifts.

Keywords Spider silk · Aciniform spidroin · Solution state NMR · Resonance assignments

Biological context

Spider silks are one of nature's most outstanding biomaterials due to a combination of strength, toughness, and elasticity not found in today's synthetic materials (Hardy et al. 2008). Spider silks are also biocompatible and biodegradable, two properties that render them highly desirable for many purposes including medical applications such as tissue scaffolds for tissue regeneration technologies. Araneoid spiders produce up to seven different protein-based silks/glues with diverse physical properties. The wrapping silk has high elasticity and toughness and is used by the spider to wrap and immobilize prey, for building sperm webs and for web decoration. For example, *Argiope trifasciata* wrapping silk is renowned for its ability to absorb energy without failing, being 50% tougher than the tough dragline silk (Lewis 2006).

The AcSp1 protein sequence of *Argiope trifasciata* wrapping silk contains at least fourteen highly homogenized repeats and a C-terminal non-repetitive domain (Lewis 2006). The repeat unit is 200 amino acids long and has no significant sequence similarity to the other silk proteins. High-resolution structural information is critical to understand spider silk fiber formation at the molecular level and to unravel the intra- and intermolecular interactions that result in such extraordinary mechanical properties. Thus far, very little is known about silk fiber formation, with most studies focusing on the major and minor ampullate (or dragline) silks, and only one repetitive domain structure, a 171-aa repeat of the egg case silk protein (TuSp1) of *Nephila antipodiana*, has been determined (Lin et al. 2009). Notably, the TuSp1 structure was determined by solution-state NMR methods. Dragline and egg case silks have completely different protein sequences and mechanical properties in comparison to wrapping silk (Hardy et al. 2008; Lewis 2006). Structural and biophysical data on dragline silk proteins or TuSp1 are therefore likely to be of limited value for understanding of AcSp1. Here, as a first step towards a high-resolution AcSp1 structure, we report the ^{1}H, ^{13}C and ^{15}N assignments alongside the likely regions of regular secondary structure for a recombinant AcSp1 repeat unit.

Lingling Xu and Marie-Laurence Tremblay contributed equally.

L. Xu · Q. Meng
Institute of Biological Sciences and Biotechnology, Donghua University, Shanghai, People's Republic of China

L. Xu · M.-L. Tremblay · X.-Q. Liu (✉) · J. K. Rainey (✉)
Department of Biochemistry & Molecular Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada
e-mail: paul.liu@dal.ca

J. K. Rainey
e-mail: jan.rainey@dal.ca

J. K. Rainey
Department of Chemistry, Dalhousie University, Halifax, NS B3H 4R2, Canada

Methods and experiments

Protein expression and purification

A synthetic gene was produced (Integrated DNA Technologies, Coralville, Iowa) to encode a 199 aa protein (W_1) that corresponded to the 200-aa consensus repeat sequence of the *Argiope trifasciata* AcSp1 protein, with the N-terminal serine of the 200-aa repeat omitted for easier cloning. This W_1 coding sequence was fused to a sequence coding for an N-terminal His_6-SUMO tag (SUMO sequence from *Saccharomyces cerevisiae*) and was inserted in a pET plasmid vector (Novagen, Darmstadt, Germany) (Fig. 1). *E. coli* BL21 (DE3) cells were transformed with the recombinant plasmid and grown as a starter culture in LB medium at 37°C to an OD_{600} of ~0.6 before being transferred into M9 minimal medium and grown for an additional 30 min. The M9 minimal medium was supplemented with 1 g/L $^{15}NH_4SO_4$ (Cambridge Isotope Laboratories, Andover, MA) and 2 g/L ^{13}C-D-glucose (Cambridge Isotope Laboratories, Andover, MA) to provide uniform ^{15}N- and ^{13}C-labeling. Cells were induced with isopropyl-β-D-thiogalactoside (IPTG) and incubated overnight at room temperature to express the His_6-SUMO-W_1 fusion protein. The fusion protein was purified using affinity chromatography on Ni–NTA Sepharose (Qiagen, Hilden, Germany) and a SUMO protease was used to cleave the fusion protein into His_6-SUMO and W_1 fragments. After removing imidazole from the protein solution by dialysis, the cleavage products were passed through a second Ni–NTA Sepharose column. The W_1 protein was collected in the flow-through fraction, while the His_6-SUMO fragment and any uncleaved fusion protein remained trapped on the column.

MGHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKT

TPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAH

1 10 20 30 40

REQIGGAGPQGGFGATGGASAGLISRVANALANTSTLRTVLRTGVSQQIA

50 60 70 80 90

SSVVQRAAQSLASTLGVDGNNLARFAVQAVSRLPAGSDTSAYAQAFSSAL

100 110 120 130 140

FNAGVLNASNIDTLGSRVLSALLNGVSSAAQGLGINVDSGSVQSDISSSS

150 160 170 180 190

SFLSTSSSASYSQASASSTSGAGYTGPSGPSTGPSGYPGPLGGGAPFGQ

SGFGG

Fig. 1 AcSp1 W_1 fusion protein sequence. The fusion protein sequence consists of the N-terminal His_6 tag (*yellow*), followed by SUMO (*black*), and AcSp1 W_1 (*blue*). Residues in *red* are unassigned and those in *purple* mark the region where chemical shifts are most perturbed

Nuclear magnetic resonance spectroscopy

Lyophilized W_1 protein (~3 mg/mL) was suspended in 90% deionized water, 10% D_2O containing 20 mM CD_3COO^-, 1 mM 2,2-dimethyl-2-sila-pentane-5-sulfonic acid (DSS) and 1 mM NaN_3 at pH 5.02 ± 0.05 (without correction for effects of deuterium). The protein suspension was vortexed for 1 min and centrifuged at 2,000*g* for 5 min to precipitate protein aggregates. The supernatant containing the soluble protein was decanted, filtered through a 0.45 µm filter (Millipore, Billerica, Massachusetts) and concentrated by use of centrifugal filter units (Ultracel-3K Amicon Ultra, Millipore, Ireland). The sample was transferred into a 5 mm high quality NMR tube (Bruker Biospin, Fällanden, Switzerland). The final concentration of the NMR sample was 0.76 mM and was measured by ultraviolet spectroscopy at 210 nm (estimated ε_{210}: 270,858 M^{-1} cm^{-1}) in a 0.5 cm path length quartz cuvette (Hellma, Müllheim, Germany).

NMR experiments were collected on a 16.4 T Avance III spectrometer (Bruker Canada, Milton, ON) equipped with a 5 mm indirect detection TCI cryoprobe at 303.15 K. Additional ^{1}H-^{15}N HSQC and HNCO experiments were acquired at 11.7 T on an INOVA spectrometer (Varian, Palo Alto, CA) equipped with a 5 mm HCN cold probe. All data were processed using NMRpipe (Delaglio et al. 1995) and analyzed using CcpNmr Analysis 2.2.1 (Vranken et al. 2005). ^{1}H frequencies were referenced to DSS at 0 ppm and ^{13}C and ^{15}N were referenced indirectly to the ^{1}H zero-point DSS frequency (Wishart et al. 1995b). AcSp-1 W_1 backbone resonances were assigned using the following 3D experiments: HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCANH, HNHA, HNHB, and ^{15}N-edited NOESY-HSQC (mixing time: 85 ms). In addition to the latter four experiments, side chain chemical shifts were determined with the aid of 3D HCCH-TOCSY (mixing time: 12 ms), ^{13}C-edited NOESY-HSQC (mixing time: 85 ms), ^{13}C-edited TOCSY-HSQC (mlev spinlock for 60 ms), and a ^{1}H-^{13}C HSQC-NOESY-^{1}H-^{15}N HSQC (mixing time: 85 ms). Arg, Asn and Gln side chain assignments were determined using both a 2D ^{1}H-^{15}N HSQC and the ^{15}N-edited NOESY-HSQC.

Assignments and data deposition

AcSp1 W_1 protein samples prepared with centrifugation and filtering steps were remarkably well behaved, proving exceptionally long lived (>9 months, to date) for collection of consistently high quality NMR data, as illustrated in the ^{1}H-^{15}N HSQC in Fig. 2. Backbone ^{1}H, ^{13}C and ^{15}N resonances were 97.5% unambiguously assigned. Most backbone amide peaks in this 199 residue protein are remarkably well dispersed, given the abundance of Ser (41), Gly (31),

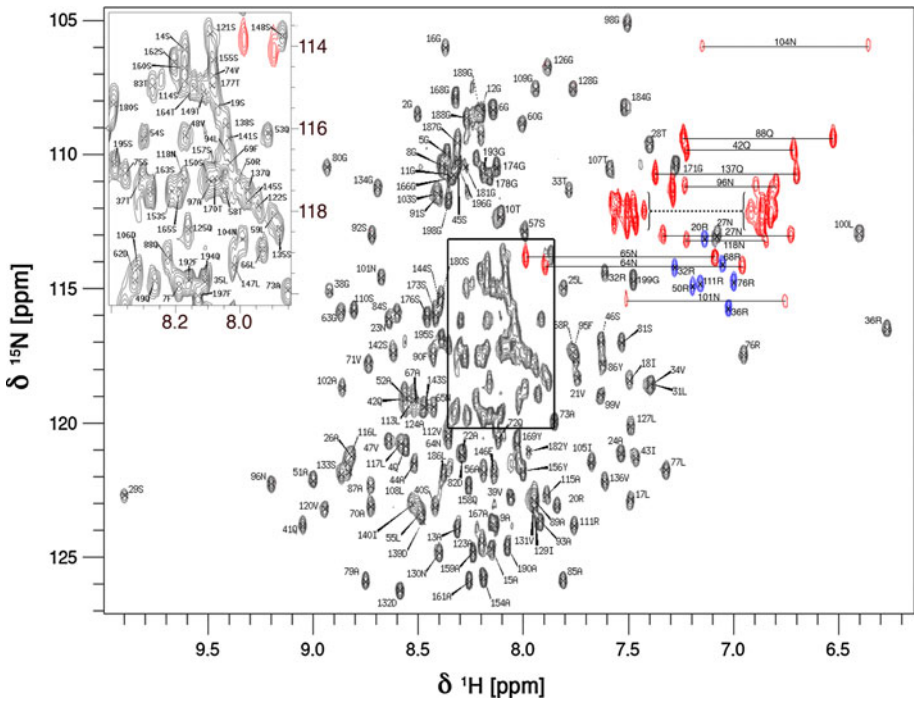

Fig. 2 2D ^{1}H-^{15}N HSQC spectrum of uniformly ^{13}C- and ^{15}N-labeled AcSp1 at pH 5.0 in 90%H_2O/10% D_2O. *Black contours* are backbone amides; *red* are Asn/Gln side chain amides (assigned pairs connected by *solid lines*; the *dotted line* represents a group of overlapped, unassigned side chain amides); and, *blue* are the Arg N^{ε}-H^{ε} resonances aliased by +30.01 ppm

and Ala (29), indicating that the protein is mostly structured and folded.

The only unassignable residues were T30, S152, and S153. Both S152 and S153 remain unassigned due to spectral overlap of the ^{1}H ^{13}C and ^{15}N shifts in the backbone spectra. These serines are located in a polyserine region leading to signal overlap for those residues. T30, located in a region where some H^N chemical shifts are significantly perturbed, still remains unassigned even after an exhaustive search, suggesting that this residue might be undergoing chemical exchange leading to line broadening and disappearance of the signal in the 3D experiments. S29 and R36 exhibit extreme H^N chemical shifts, at 9.9 ppm for S29 and 6.3 ppm for R36 (Fig. 2). According to BioMagResBank (BMRB) statistics, chemical shifts in these ranges have been reported for <3% of the submitted assignments for those residues (Ulrich et al. 2008). Correspondingly, L31 and V34 also exhibit significantly perturbed H^N chemical shifts relative to their random chemical shifts (Δ0.76 and Δ0.64 ppm respectively) (Wishart et al. 1995a). Two other amino acids also have perturbed H^N chemical shifts: L100 and R76 with Δ1.71 ppm and Δ1.28 ppm respectively. There are ~27 unassigned, generally weaker crosspeaks in the ^{1}H-^{15}N HSQC that are also observed in some of the 3D experiments. The majority are weaker in intensity, with a few being comparable in intensity to crosspeaks for assigned spin systems. Based on inferences from connectivity in HNHA, HNHB and CBCANH, these unassigned amino acids consist of 7 Gly, 8 Ser, 3 Ala and 9 other amino acids. None of these spin systems are sequentially connected in the backbone walk experiments, at least at the level of signal-to-noise achieved, and none exhibit any NOE crosspeaks making further analysis impossible at this time. These spin systems may arise from an as yet unassigned minor secondary conformation of the protein.

90% of side chain ^{1}H resonances and 76% of side chain ^{13}C and ^{15}N resonances were assigned. The remaining unassigned side chain nuclei are the aromatic Phe and Tyr side chains, and Gln side-chain amides.

Secondary chemical shifts were obtained by subtraction of the random coil values in water (Wishart et al. 1995a) from the observed chemical shifts. These were in turn used to determine the chemical shift index (Wishart et al. 1992) at each residue using the consensus between secondary chemical shift values calculated for C^{α}, H^{α}, C^{β}, and C' (Fig. 3). The secondary structure was also predicted using the algorithm DANGLE, as implemented in CcpNmr Analysis (Cheung et al. 2010; Vranken et al. 2005) The consensus of secondary structure assignments of DANGLE and CSI indicates that AcSp W_1 is composed of six major helical regions (Fig. 3). The last quarter of the protein (residues 150–199) appears unstructured on the basis of chemical shifts, but this does not come as a surprise given that 6/8 Pro and 14/31 Gly are located in that region. A major α-helical portion seems to be located between residues 102–151, denoted as helix 5 and 6 in Fig. 3, and includes a kink at residue L108 and a loop between residues L126-D132. The ^{1}H, ^{13}C, and ^{15}N backbone and side chain resonances were deposited in the BMRB with accession number 17899.

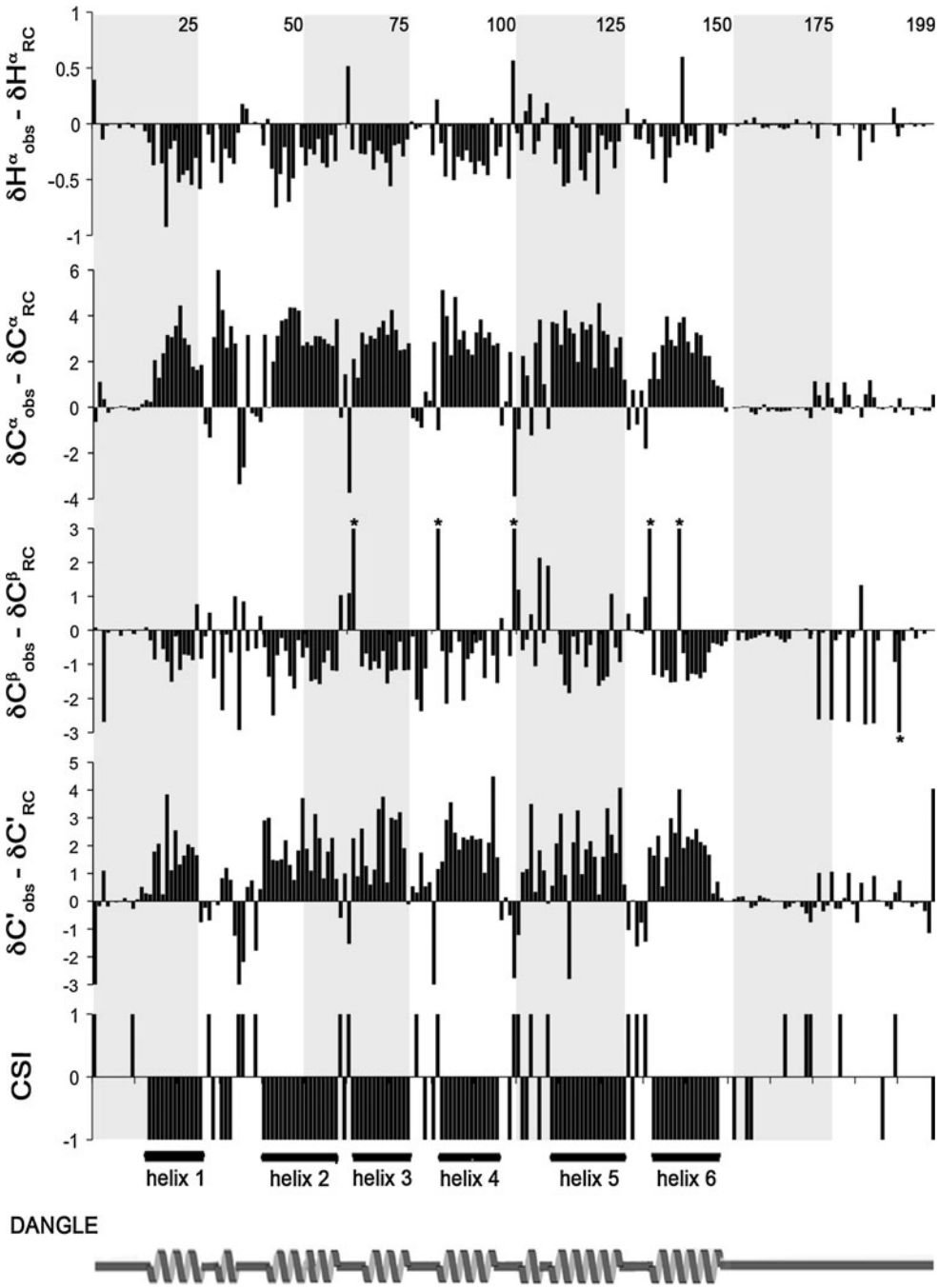

Fig. 3 Chemical shift index (CSI, Wishart et al. 1992) histogram illustrating the predicted secondary structures of W_1 according H^{α}, C^{α}, C^{β}, and C′. The CSI plot and DANGLE secondary structure prediction was generated by Analysis (Vranken et al. 2005; Cheung et al. 2010). The *stars above certain bars* indicate deviations greater then the range shown. Refer to Fig. 1 for the protein sequence and respective numbering of amino acids

Acknowledgments We thank Yanfei Wang for technical support in the gene preparation. This work was supported by a research grant from the Canadian Institutes of Health research to XQL; a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to JKR; research grants to QM from the National High Technology Research and Development Program 863 (NO 2006AA03Z451), the National Natural Science Foundation of China (NO 31070698), and the Shanghai key projects of basic research (NO 10JC1400300); and, a Nova Scotia Health Research Foundation Studentship Award to MLT. 700 MHz NMR data were collected at the Biomolecular Magnetic Resonance Facility (BMRF) housed at the National Research Council of Canada's Institute for Marine Biosciences (NRC-IMB), as facilitated by the Nuclear Magnetic Resonance Research Resource (NMR-3). The He-cooled probe for the 700 MHz at the NRC-IMB was provided by Dalhousie University through an Atlantic Canada Opportunities Agency Grant. 500 MHz NMR experiments were recorded at the Québec/Eastern Canada High Field NMR Facility (QANUC), supported by grants from the Canada Foundation for Innovation, the Natural Sciences and Engineering Research Council of Canada, the Québec ministère de la recherche en science et technologie, and McGill University. Thanks to Dr. Tara Sprules for her expert data acquisition at QANUC.

References

Cheung MS, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: a Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. J Magn Reson 202(2):223–233. doi:10.1016/j.jmr.2009.11.008

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6(3):277–293

Hardy JG, Römer LM, Scheibel TR (2008) Polymeric materials based on silk proteins. Polymer 49(20):4309–4327. doi:10.1016/j.polymer.2008.08.006

Lewis RV (2006) Spider silk: ancient ideas for new biomaterials. Chem Rev 106(9):3762–3774. doi:10.1021/cr010194g

Lin Z, Huang W, Zhang J, Fan JS, Yang D (2009) Solution structure of eggcase silk protein and its implications for silk fiber formation. Proc Natl Acad Sci USA 106(22):8906–8911. doi:10.1073/pnas.0813255106

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2008) BioMagResBank. Nucleic Acids Res 361(Database issue): D402–D408. doi:10.1093/nar/gkm957

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59(4):687–696. doi:10.1002/prot.20449

Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry 31(6):1647–1651

Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD (1995a) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids I Investigations of nearest-neighbor effects. J Biomol NMR 5(1):67–81

Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD (1995b) ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J Biomol NMR 6(2):135–140