THE STRUCTURAL CHARACTERIZATION OF *ARGIOPE TRIFASCIATA* SPIDER WRAPPING SILK BY SOLUTION-STATE NMR

By

Marie-Laurence Tremblay

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
December 2014

DEDICATION

I would like to dedicate this thesis to all my friends that I have lost, gained, and kept, and to my family for their undying support throughout my journey as a university student. Together, they have kept me strong and standing tall during this rocky road that eventually shaped me into the strong and confident person that always dreamed to be, making all of this possible.

- Thank you.

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Biomolecular nuclear magnetic resonance (NMR) spectroscopy frequently employs secondary chemical shifts to estimate local secondary structuring in advance of obtaining a full 3D structure. To assess the effect of variation in dielectric upon secondary chemical shift, two new sets of random coil chemical shifts, one in dimethyl sulfoxide and the other in chloroform:methanol:water, a popular membrane-mimetic solvent system, were produced. Using a new program, CS-CHEMeleon, we demonstrated that the chemical shift-based structure prediction accuracy was much more affected by the secondary structure type than solvent environment.

Spider silks are outstanding biomaterials with remarkable mechanical and physical properties. Little is known about spider silk atomic-level structuring and how this relates to variations in mechanical properties. The spider wrapping silk protein in *Argiope trafisciata* (AcSp1) contains a 200 amino acid sequence ("W" unit) repeated at least 14 times; it self-assembles into the toughest type of silk. The high-resolution structure of a recombinantly produced repeat W unit was solved by solution-state NMR. W_1 is composed of a 5-helix globular core with intrinsically disordered N- and C-terminal tails. With the use of split-intein technology, the effect of repetitive domain tandemization was investigated with W_2 (W_1+W_1) through separate selective labeling of each of the two repeats in W_2 with ^{13}C and/or ^{15}N NMR active isotopes, providing W_2 constructs where only one W unit was enriched with NMR-active isotopes. W_2 backbone chemical shifts demonstrate the conservation of the W_1 structure within W_2 and nuclear spin relaxation analysis demonstrates that motions within W_2 are conserved with tandemerization. Reduced spectral density mapping analysis shows two types of motion, one for globular core and another for the tails or linker, further emphasizing the two separate domains of AcSp1. NMR titrations of W_1 and W_2 with chemical denaturants and with the detergent dodecylphosphocholine (DPC) were performed to locate the residues most amenable to unfolding and, hence, likely responsible for initial structural transition upon fibrillogenesis. On the basis of these data, a structural model for the solution structure of native AcSp1 is described, having a "beads-on-a-string" architecture, and the region likely responsible for seeding fibrillogenesis is proposed.

LIST OF ABBREVIATIONS AND SYMBOLS USED

γ	Gyromagnetic ratio
δ	Chemical shift
$\Delta\delta$	Secondary chemical shift
ε	Dielectric constant
λ	Wavelength
φ	Psi dihedral angle
ϕ	Phi dihedral angle
τ_c	Global/rotational correlation time
τ_e	Effective internal correlation time
τ_f	Fast motion internal correlation time
τ_s	Slow motion internal correlation time
ω_o	Larmor frequency
AcSp1	Aciniform spidroin 1
ADR	Ambiguous distance restraint
AIC	Akaike information criteria
ARIA	Ambiguous restraint for iterative assignment
ASA	Accessible surface area
B_o	External magnetic field
BMRB	Biological magnetic resonance bank
CCS	Combined chemical shift
CD	Circular dichroism spectropolarimetry
CS	Chemical shift
CSI	Chemical shift index
CTD	C-terminal non-repetitive domain
Cyl	Cylindriform
D_C	Translational diffusion coefficient
DMF	Dimethyl formamide
DMSO	Dimethyl sulfoxide
DOSY	Diffusion ordered spectroscopy
DPC	Dodecylphosphocholine

DSS	2,2-dimethyl-2-silapentane-5-sulfonate
Flag	Flagelliform
GdmCl	Guanidinium chloride
GUI	Graphical user interface
hetNOE	$\{^{1}\mathrm{H}\}$-^{15}N heteronuclear nuclear Overhauser effect
HSQC	Heteronuclear single quantum correlation spectroscopy
IDIS	Isotopically discriminating
IPAP	In-phase anti-phase
$J(\omega)$	The spectral density function
MA	Major ampullate
MAS	Magic angle spinning
MI	Minor ampullate
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NMR	Nuclear magnetic resonance
ssNMR	soli-state nuclear magnetic resonance
NRC-IMB	National Research Council Institute for Marine Biology
NTD	N-terminal non-repetitive domain
PDB	Protein Data Bank
PRE	Paramagnetic Relaxation Enhancement
R_1	Longitudinal relaxation rate
R_2	Transverse relaxation rate
R_{ex}	Conformational sampling rate
RCCS	Random coil chemical shift
RDC	Resdiual dipolar coupling
R_{g}	Radius of gyration
RMSD	Root mean square deviation
S^2	Generalized order parameter
SSP	Secondary structure propensity
T_1	Longitudunal relaxation time
T_2	Transverse relaxtion time

TOCSY	Total correlation spectroscopy
W_1	One AcSp1 200 amino acid repeated domain
W_2	Two concatenated AcSp1 200 amino acid repeated domain
W_3 and W_4	Three and four concatenated AcSp1 200 amino acid repeated domain

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Jan Rainey for providing me with an excellent enthusiastic scientific environment and entrusting me with such wonderful biomaterial projects since 2008. I am also extremely grateful to Dr. Paul Liu for introducing our lab to the wonders of spider silk, and to Dr. Lingling Xu, my collaborator, my second half in the spider silk research, and good friend, without whom all this would have been impossible, and to Kathleen Orrell who helped in protein expression and purification. Special thanks to Michèle Auger, Thierry Lefèvre, and Michel Pézolet at Université Laval, QC, for their collaboration on the structure of spider silk using Raman Spectroscopy.

I would like thank my committee members Drs. David Jakeman, David Byers, and Laurent Kreplak for their support, trouble shooting, useful criticism of my project over the years and for keeping me on track when I had too many ideas projects and not enough time to complete them.

I would also like to express my gratitude to all the NMR people on the east coast of North America that have helped me gather NMR data for this thesis: Dr. Mike Lumsden at the Nuclear Magnetic Resonance Research Resource (NMR^3) at Dalhousie University for all the help in setting up solution-state NMR experiments on the 500 MHz magnet over the years, to Dr. Ulrike Werner-Zwanziger for initiating me to solid-state NMR and her valuable assistance in setting up and running proteins solid-state experiments here at Dalhousie University, special thanks also goes to Drs. Ray Syvitski, Nadine Merkley, and Ian Burton at the National Research Council of Canada Institute for Marine Biology for all their help, counsel, and patience in setting up NMR experiments. And also great thanks to Christopher Kirby and Lloyd Kerry at the Crops and Livestock Research Centre at Agriculture and Agri-Food Canada for their use of their 600 MHz solution-state magnet, named Maggie, in Charlottetown, Prince-Edward Island, and to Drs. Zhehong Gan and Ivan Hung at the National High Magnetic Field Laboratory in Tallahassee, Florida, USA for initiating me in protein solid-state NMR experiments on spider silk, and their valuable counsel and ideas on experimental procedures.

I would like to thank the Nova Scotia Health Research Foundation and the Natural Science and Engineering Research Council for their financial support.

My everyday work has been greatly facilitated by working with a great team. Thanks goes to Tyler Reddy, David Langelaan for helping me settle in the lab and teaching me the basics, to Aaron Banks for programming CS-CHEMelon and teaching me about computers and programming, to Muzaddid Sarker with whom I've had numerous insightful discussions about NMR assignments and structure calculations, Kyung Soo Shin, Aditya Pandey, and Nigel Chapman for supporting me throughout the years, to all past and present lab members, and to all the undergraduate students I supervised over the years, constantly reminding me of my passion for science.

CHAPTER 1. INTRODUCTION

1.1. SPIDER SILKS AS BIOMATERIALS

Naturally derived polymers exhibit a variety of physical and mechanical properties (Donald et al. 2006) and despite the individually weak components from which they assemble. Nature has found many ways to arrange these simple building blocks into highly proficient materials whose properties outweigh synthetic, man-made replicas. Natural fibres self-assemble at ambient temperature and pressure, providing a large advantage over synthetic fibres that require harsh chemicals for fabrication, non-ambient temperature and pressure conditions for fabrication (Ko and Jovicic 2004). Some examples of proteinaceous fibres are of wool, silk, and cellulose fibres such as cotton and linen (Hsia et al. 2011). The relative biocompatibility of these fibres, in combination with their stability, unique mechanical properties, and options for genetic control, provide a foundation upon which to exploit these natural proteins for use in every life (Wong Po Foo and Kaplan 2002).

Silk fibres are fine, lustrous protein filaments excreted from animals (Hsia et al. 2011). Other than spiders, silks are produced by the mulberry silkworm *Bombyx mori*, and by other arthropods including mites, moths, butterflies, wasps, and bees (Zhao and Asakura 2001; Foo et al. 2006). Each type of silk differs in physical and mechanical properties, but spider silks outshine all other types of silk in their mechanical properties. They are outstanding biomaterials with strength comparable to steel, are tougher than the strongest synthetic material Kevlar, and are as light as cotton or nylon (Vollrath and Knight 2001; Lewis 2006; Omenetto and Kaplan 2010).

Spiders produce up to seven types of silks that serve many different functions crucial for spider survival, including cocoon fabrication, egg protection, prey wrapping, web construction and decoration, and safety lines (Foo et al. 2006). Their excellent biocompatibility (Altman et al. 2003), minimal immunogenicity (MacIntosh et al. 2008), and limited bacterial adhesion (Zhang et al. 2008) mean that this natural biopolymer is suitable for a variety of applications. To this day, most silk applications are derived from silkworm silk simply because silkworms are docile and easy to farm. Spiders, although their silks have more appealing mechanical properties and diversity (Lewis 2006), are difficult to farm due to their aggressive and territorial behaviour (Kundu et al. 2014), increasing the difficulty of industrialization. Rather than farming spiders, genetically cloning spider silk genes into more tractable organisms seems to be the current mainstream solution for industrialization of spider silk. But the resulting silk fibres possess less desirable mechanical properties than silk spun from spiders. It is therefore of paramount importance that we investigate the mechanism of spider silk fibre formation and understand the origins of strength and toughness.

1.2. THE 7 TYPES OF SPIDER SILK

1.2.1. General Properties Of Spider Silks

Female spiders produce up to seven types of silk, each named after the gland in which it is generated, and each varying in physical and mechanical properties (Lewis 2006; Heidebrecht and Scheibel 2013). Spider silk proteins, commonly referred to as spidroins (contraction of 'spider' and 'fibroin'), are extremely large proteins (250-500 kDa) rich in glycines and alanines, which can amount to half of the protein sequence (Bini et al. 2004). The typical spidroin architecture is predominantly composed of a repetitive core, accounting for approximately 90% of the total protein sequence, flanked by non-repetitive N- and C-terminal domains (referred to as NTD and CTD, respectively) of varying sizes (Andersen 1970; Hagn et al. 2010a; Eisoldt et al. 2012).

As the name suggests, the repetitive core of spidroins consists of repeated protein motifs and ranging from ~3-200 amino acids in size. Iterations of four simple amino acid sequences generally comprise the majority of the spider silk motifs: (A_n), (GA_n), $(GGX)_n$, and $(GPGX)_n$, where X can be any of a subset of residues depending on the silk type (Hayashi et al. 2004; Garb and Hayashi 2005). These repeated motifs are typically associated with specific secondary structures and are arranged in various combinations generally correlating to observed mechanical properties (Hayashi et al. 1999; Gosline et al. 1999; Brown et al. 2011). For example, A_n and $(GA)_n$ are involved in β-sheet nanocrystals known to provide strength to the fibre (Keten et al. 2010), while the $(GGX)_n$ (X= L, Q, R, Y) and $(GPGXX)_n$ motifs (XX = GA, GS, GY, or QQ) adopt a disordered conformation important for fibre elasticity (Gosline et al. 1986).

The formation of insoluble spider silk fibres from soluble spidroins involves a secondary structural change from an α-helix/random coil rich state to β-sheet/random coil rich fibre, with retention of α-helical content in some types of silks (Figure 1). The relative proportions of secondary structure elements are highly dependent on the type of silk (Lefèvre et al. 2011). X-ray scattering studies demonstrate that only a small fraction of the material in spider silk fibres is crystalline (~15%) (Riekel et al. 1999; Riekel and Vollrath 2001; Glišovic et al. 2008; Izdebski et al. 2010). Solid state NMR (ssNMR) studies show unambiguously that ~82% of A_n and 28 % of $(GA)_n$ repeats in dragline silk are located in crystalline β-sheets (Holland et al. 2008b), and $GPGX_n$ forms β-spirals, which are a result of repeated polyproline-II (PPII) helices (Izdebski et al. 2010; Asakura et al. 2013a). The origin of the spider silk's strength, stiffness, and toughness comes from the network of hydrogen bonds within the β-sheets nanocrystals embedded in an soft polymeric region (Keten et al. 2010).

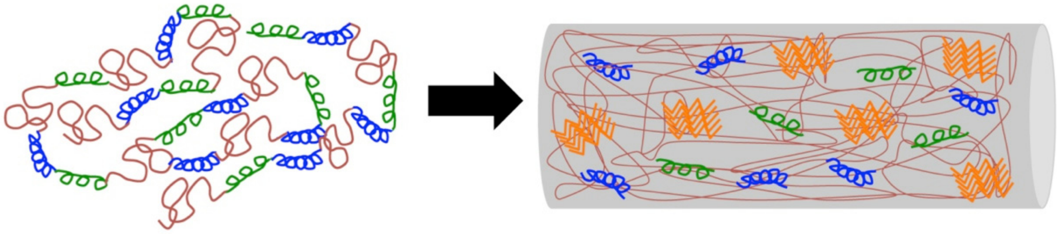


Figure 1 | Spidroin secondary structure transition for solution to fibrous form. The soluble form contains a mixture of helices (3_{10} (green),α (blue) and random coil (red)) for all types of silk and the fibre might contain a mixture of those same elements with the addition of β-sheets (orange).

1.2.2. Major Ampullate Silk

The major ampullate (MA) spidroin, generally referred to as dragline silk, is the most thoroughly investigated of all types of spider silk due to its high strength, with comparable values to steel (Table 1) (Lewis 2006; Slotta et al. 2007; Creager et al. 2010). MA spidroin functions as the spider's lifeline when it needs to escape from predators and is also used to build the frame of the orb web (Hayashi et al. 1999; Exler et al. 2007; Geisler et al. 2008; Heidebrecht and Scheibel 2013). MA silk is a semi-crystalline polymer comprised of rigid alanine-rich β-sheet nanocrystals of 2x5x7 nm (van Beek et al. 2002) surrounded by a soft glycine-rich matrix responsible for the fibre's elasticity. MA proteins are ~200-350 kDa and are divided into two classes: one with a low proline content (MaSp1) and the other is proline-rich (MaSp2) (Candelas and Cintron 1981; Ayoub et al. 2007). MA proteins are natively unfolded and only show trace content of α-helices and 3_{10}-helices. The MA silk repetitive domain consists of small motifs of A_n, $(GA)_n$, $(GGX)_n$, and $(GPGXX)_n$, where X is typically Tyr, Leu or Gln, repeated at least 100 times, and flanked by highly conserved NTD and CTD (Bini et al. 2004; Challis et al. 2006; Foo et al. 2006; Lewis 2006; Rising et al. 2010). The proline containing GPGQQ and GPGXX motifs are responsible for β-turns and γ-turns while $(GGX)_n$ folds into a 3_{10}-helix (Hijirida et al. 1996; Thiel et al. 1997; Hayashi et al. 1999; Liu et al. 2008). The aromatic residues, usually found in the GGX motifs, are also found incorporated into the 3_{10}-helical soft matrix and not into β-sheets (Izdebski et al. 2010). The amount of disorder and minimal structuring of the disordered matrix within the fibre, in combination with fibre hydration, are fundamental elements required for elastic silk (Parkhe et al. 1997; Shi et al. 2014).

1.2.3. Minor Ampullate Silk

Minor ampullate (MI) silk is ~ 250 kDa in size, shares similar mechanical properties to MA silk, and is also composed of two major proteins: MiSp1 and MiSp2 (Lewis 2006; Hardy et al. 2008). The minor ampullate is used to build the spiral that stabilizes the scaffold (made of MA silk) of the web. In contrast, the MI silk protein sequence is quite different from MA silk; it typically has no prolines, the glutamine content is quite low, and the dominating motifs typically comprise 10 consecutive repeats of GGXGGY (X = Tyr, Gln, or Ala, Y = any amino acid) followed by $(GA)_n(A)_y$ (n = 1-3 and y = 2-5) (Colgin and Lewis 1998; Heidebrecht and Scheibel 2013). MI silk contains multiple copies of a highly conserved 137 amino acid non-repetitive serine-rich ‘linker’ sequence that acts as a spacer between the repetitive regions (Colgin and Lewis 1998). Unlike MA silk, MI silk A_n stretches are shorter and don’t contribute as significantly to the crystalline β-sheet content (Parkhe et al. 1997). The alanine residues in fibres are also less oriented than those in MA fibres (Lefèvre et al. 2011). In solution, similarly to MA silk, soluble MI proteins are nearly unfolded with small amounts of α-helices and 3_{10}-helices (Lefèvre et al. 2011).

Table 1 | Mechanical properties of silks and other materials.

Material	Young's modulus (GPa)	Strength (MPa)	Extensibility (%)	Toughness (MJ/m^3)	Reference
Steel	200	1500	0.8	6	(Gosline et al. 1999)
Kevlar 49	130	360	2.7	50	(Gosline et al. 1999)
Rubber	0.001	0.05	850	100	(Lewis 2006)
Nylon 6.6	5	950	18	80	(Heidebrecht and Scheibel 2013)
Carbon fibre	300	4000	1.3	25	(Heidebrecht and Scheibel 2013)
Elastin	0.001	2	15	2	(Heidebrecht and Scheibel 2013)
Tendon collagen	1.5	150	12	7.5	(Gosline et al. 1999)
Bombyx Mori	7	600	18	70	(Lewis 2006)
Major Ampullate *Nephila clavipes*	22	1300	12	80	(Gosline et al. 1999)
Major Ampullate *Argiope bruennichi*	11.8	1320	22	134	(Zhao et al. 2006)
Major Ampullate *Araneus diadematus*	10	1100	27	160	(Zhao et al. 2006)
Major Ampullate *Argiope trifasciata*	9.3	1290	22	145	(Hayashi et al. 2004)
Major Ampullate *Latrodectus hesperus*	nd	1000	34	nd	(Heidebrecht and Scheibel 2013)
Minor Ampullate *Argiope trifasciata*	8.5	342	54	148	(Heidebrecht and Scheibel 2013)
Aciniform *Argiope trifasciata*	9.6	687	86	367	(Hayashi et al. 2004)
Flagelliform *Araneus diadematus*	0.003	500	270	150	(Gosline et al. 1999)
Cylindrical *Argiope bruennichi*	9.1	390	40	128.6	(Zhao et al. 2006)

1.2.4. Flagelliform

Flagelliform (flag) silk, also known as viscid silk, has a high molecular weight close to 500 kDa (Hayashi and Lewis 1998). The sticky and elastic properties of flag silk are used in the capture spiral of the orb web as a way to dissipate the kinetic energy produced by flying insects upon impact on the web (Bini et al. 2004; Ohgo et al. 2006). Flag silk contains equal amounts of proline and alanine, with reduced amounts of alanine, and greater amount of valine compared to MA and MI silks (Hayashi and Lewis 1998). The most common repeats are the GPGGX (X=Ala, Ser, Tyr, or Val) forming the elastic β-spirals and the $(GGX)_n$ motif responsible for the 3_{10}-helices within the fibre

(Dicko et al. 2004a; Ohgo et al. 2006). Reminiscent to MI silk, flag silk contains a 34 amino acid ‘spacer’ sequence that appears between the highly repetitive small motifs. This spacer contains many charged and hydrophilic amino acids (Hayashi and Lewis 1998). Flag silk in its fibrous form is unique because it contains no crystalline fraction and the protein has no preferred molecular orientation (Craig and Riekel 2002). Due to the large and periodic distribution of prolines in the sequence, there is no β-sheet present (Lefèvre et al. 2011).

1.2.5. Pyriform

Pyriform spidroin (PySp1 and PySp2) is a protein glue that forms the attachment disks holding the web joints together and tethering the dragline silk to surfaces (Hardy et al. 2008; Hajer et al. 2009; Heidebrecht and Scheibel 2013). Pyriform silk contains a low proportion of small amino acids (Gly and Ala) and instead contains the highest percentage of polar amino acids (S+T, Q, E, R+K) of all the silk types (46 and 49% for PySp1 and PySp2, respectively). *Latrodectus hesperus* PySp1 consists of blocks of AAARAQAQAEARAKAE and AAARAQAQAE and is completely void of the common sub-repeat modules found in other silks (Blasingame et al. 2009). Two new silk repetitive motifs from *Argiope trifasciata*, *Nephila clavipes*, and *Nephilengys cruentata* (Perry et al. 2010; Geurts et al. 2010) were sequenced and attributed to PySp2: an alternating proline-rich motif (PXPXP), and a sequences dominated by the amino acids Ser, Gln, and Ala. In all species, the alternating polar and non-polar residue motifs render this type of silk ideal for both surface and silk binding (Perry et al. 2010).

Structurally, *N. clavipes* pyriform silk displays an estimated 45% α-helical character in solution with the remainder being disordered (Lefèvre et al. 2011). In its fibrillous form (Wolff et al. 2015), a secondary structure transition occurs, decreasing the amount of disorder and increasing the β-sheet content to achieve a final ~1:1:1 ratio of α-helix:β-sheet:random coil (Lefèvre et al. 2011).

1.2.6. Cylindriform Spidroin

Cylindriform silk proteins (cyl) are part of the tough outer egg case protecting the offspring, and are unique to mature female spiders (Tian and Lewis 2005; Zhao et al. 2006). Cyl silk of *A. diadematus* spiders is composed of A_n and $(GGX)_n$ blocks , where

X = Ala, Leu, Gln, or Tyr, yielding protein structures rich in β-sheet similar to MA silk (Barghout et al. 1999; Hardy et al. 2008). Soluble cyl proteins are folded inside the gland and contain about 51-57% α-helix (Dicko et al. 2004b; Lefèvre et al. 2011). The other half of the protein is estimated to be composed of disordered regions, 3_{10}-helices, and turns. The β-sheets within cyl fibres are oriented, and the spinning process involves a complete α-helix → β-sheet transition, like MA and MI silks (Parkhe et al. 1997; Lefèvre et al. 2011).

1.2.7. Tubuliform Spidroin

TuSp1 forms part of the tough outer egg case protecting the offspring, along with cyl silk, and is again only found in mature female spiders (Tian and Lewis 2005; Zhao et al. 2006; Lin et al. 2009). The amino acid motifs found in other types of silk are not present in TuSp1. TuSp1 is instead characterized by a high serine content and low glycine content, differentiating TuSp1 from all other types of silks being the second most polar spidroins (Geurts et al. 2010). The representative motifs are S_n, $(SA)_n$, $(SQ)_n$, and GX (X = Gln, Asn, Ile, Leu, Ala, Val, Tyr, Phe, or Asp). Raman spectroscopy (Lefèvre et al. 2011) and X-ray diffraction patterns (Parkhe et al. 1997) both demonstrate that TuSp1 has a high content of crystalline β-sheet within its fibre. Mechanical testing shows that tubuliform silk has relatively high tensile strength with a fairly low elasticity (Table 1), similar to MI silk (Stauffer et al. 1994).

The structures of two repeat domains of TuSp1, RP1 and RP2 from *Nephila antipodiana*, were the first spider silk structures of a large repetitive domain available and were solved by solution-state NMR by the Yang group (Lin et al. 2009) (Figure 2 and Figure 3).

RP1 sequence:

```
1     SGATSQAASQ SASSSYSSAF AQAASSALAT SSAISRAFAS VSSASAASSL
51    AYNIGLSAAR SLGIASDTAL AGALAQAVGG VGAGASASAY ANAIARAAGQ
101   FLATQGVLNA GNASALAGSF ARALSASAES QSFAQSQAYQ QASAFQQAAA
151   QSAAQSASRA - 160
```

RP2 sequence:

```
1     SYSSAFAQAA SSSLATSSAI SRAFASVSSA SAASSLAYNI GLSAARSLGI
51    ASDTALAGAL AQAVGGVGAG ASASAYANAI ARAAGQFLAT QGVLNAVNAS
101   SLGSALANAL SDSAANSAVS GNYLGVSQN — 129
```

Figure 2 | *Nephila antipodiana* TuSp RP1 and 2 sequence reported in Lin *et al* (Lin et al. 2009).

Lin *et al.* found that RP1 consists of disordered N- and C-termini (31 and 20 residues respectively) flanking the folded domain, which forms a six-helix bundle (Lin et al. 2009). About 46% of the hydrophobic residues located in the structural region were solvent exposed. Some form a long hydrophobic groove on the surface while the rest are scattered throughout the surface. RP2 is similarly structured to RP1 with a folded core consisting of six orthogonal helices capped by disordered N- and C-termini with identical hydrophobic amino acid distributions. RP1 and RP2 structurally overlap at 95%. Neither RP1 nor RP2 form homodimers but, in the presence of each other, form ~2% of hetero-oligomers in solution.

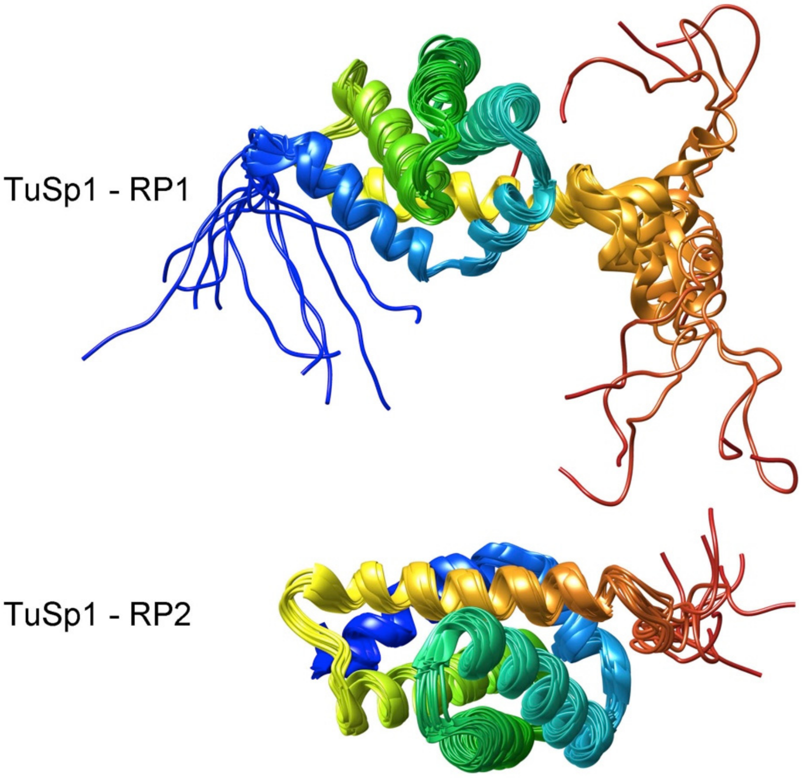


Figure 3 | Rainbow colour representation of the solution-state TuSp1 RP1 and RP2 structures by the Yang's group (Lin et al. 2009). The gradient from blue to orange represents the N-terminal to the C-terminal.

1.2.8. Aciniform Spidroin

Aciniform spidroin (AcSp1), also known as wrapping silk, is composed primarily of the protein AcSp1 and is the toughest spider silk (Table 1) (Hayashi et al. 2004; Lewis 2006). Female orb web weaving spiders use aciniform silk for the soft lining of the egg case, with the outer egg case composed of tubuliform and cylindriform silks (Lewis 2006; Heidebrecht and Scheibel 2013). AcSp1 is also used for prey wrapping (Lewis 2006; Chaw et al. 2014), helps in pyriform silk attachments (Vasanthavada et al. 2007), and serves as web stabilizers, known as stabilimenta (Chaw et al. 2014).

AcSp1 is the unconventional silk protein; it lacks the small repetitive motifs found in all other types of silks and rather comprises of a homogenous 200 amino acid sequence (referred to as W units herein, W for wrapping silk) repeated at least 14 times flanked by a ~99 amino acid CTD and a putative NTD in *Argiope trifasciata* (Hayashi et al. 2004) and a total of 20 times flanked by an NTD and CTD in *Argiope argentata*

(Chaw et al. 2014). The sequence of the W unit from *A. trifasciata* is presented in Figure 4 with amino acid frequency by type of residue shown in Figure 5.

```
1    AGPQGGFGAT GGASAGLISR VANALANTST LRTVLRTGVS QQIASSVVQR
51   AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
101  NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSSSSFLSTS
151  SSSASYSQAS ASSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGGS
```

Figure 4 | *Argiope trifasciata* AcSp1 200 amino acid repetitive domain sequence.

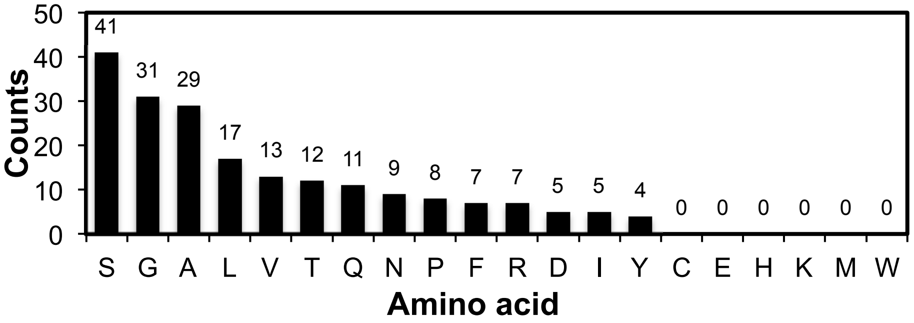


Figure 5 | Frequency of amino acids in *Argiope trafisciata* AcSp1 W_1.

The repeat units are highly conserved and show 100% identity to each other at the protein level and an astonishing 99.9% identity at the DNA level (Chaw et al. 2014). Arguably, the most common sub-repeats in aciniform silks are TGPSG, repeated twice, and a small polyserine region, which only accounts for 8.5% of the total sequence (Hayashi et al. 2004). In addition, the polyalanine and $(GA)_n$ stretches responsible for the incredible strength in MA and MI silks are absent in AcSp1 and therefore cannot account for AcSp1 mechanical properties.

In addition, AcSp1 was sequenced from the aciniform gland in *Latrodectus hesperus* from the *Theridiidae* family rather than the *Araneidae* family (Ayoub et al. 2013). *L. Hesperus* AcSp1 contains 16 repeats of two differing repetitive domains of ~187 amino acids, with 45% identity and 62% similarity (Vasanthavada et al. 2007; La Mattina et al. 2008). *A. trifasciata* AcSp1 shares ~44% sequence identity with *L. hesperus* AcSp1-like domain (Vasanthavada et al. 2007). Sequence alignment between the *A. trifasciata* and *A. argentata* AcSp1 shows an overall 79% identity, 84% over the predicted folded domain and 54% over the predicted disordered regions (prediction based on (Xu et al. 2012b)).

To date, there is one atomic-level structure of a truncated AcSp1 repetitive domain (AcSp-tr) of *N. antipodiana* (*Nephilidae* family) published by the Yang group (Wang et al. 2012). The variability in the AcSp1 protein between species seems to be quite large; *N. antipodiana* AcSp1 has only 37% similarity to *A. trifasciata* AcSp1. The sequence of the AcSp-tr is as follows: the italic residues represent the His$_{6x}$ tag for purification and part of the protein, the bold residues represent the truncated disordered portion not structurally evaluated:

1 *MHHHHHH*SGS LGDQLTSTLA SALTKTNTLK AVSASKPSAN VAVAIVTSGL
51 KKALGALRIN AGVSSQLTSA VSQAVANVRP GSSPAVYAKA IAAPSVQILV
101 SSGSVNNNNA KQVASTLSEN LVREMANTAR RYRVNVPEAS VQADVSLVTS
151 MTSTFVISSQ TSVQMGG**FPG SDSGFPGGDA GYPGGDAGFP AGGDYPEGGA**
201 **PSDLGGPSG**

Figure 6 | *N. antipodiana* AcSp1 repetitive domain sequence reported in Wang *et al*. (Wang et al. 2012).

AcSp1-tr is composed of a compact seven-helix bundle and a predicted C-terminal flexible tail (Wang et al. 2012). The barrel-shaped core of the protein is held tightly together by strong hydrophobic interactions (Figure 7). Hydrophobic residues, such as Val, Leu, and Ile, line the inside of the 'barrel'. The surface of AcSp1-tr is very hydrophilic, without large hydrophobic patches.

Structural evaluation by Raman spectroscopy revealed that *N. clavipes* AcSp1 contains ~50:50 α-helical disordered structure and transitions into an insoluble fibre with a mixture of ~24:30:46 of α-helix: β-sheet: disordered structure (Lefèvre et al. 2011). The β-sheets are moderately oriented. Unlike the MA, MI, and cyl silks, the transformation of the AcSp1 fibre into an oriented β-sheet is 'incomplete', as 24% of the α-helical content is retained in the fibre. The apparent minimum sequence requirement for *A. trifasciata* to form fibres is two consecutive 200 amino acid repeat units (termed W$_2$ for 2 repeats) (Xu et al. 2012a) whereas Wang *et al.* (Wang et al. 2012) report that only one repeat unit from *N. antipodiana* is required for fibre formation.

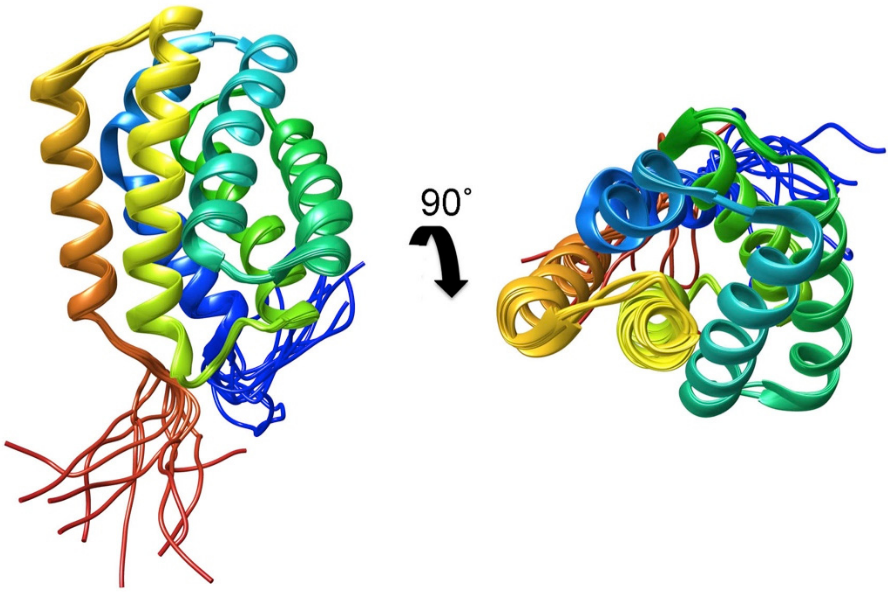


Figure 7 | Rainbow coloured representation of the solution-state AcSp1-tr from the Yang group (Wang et al. 2012) with pdb accession 2LYI. The gradient from blue to red represents the N-terminal to the C-terminal.

1.3. THE SPIDER SPINNING APPARATUS

Thus far, the MA gland has been the most highly characterized spider silk gland (Coddington 1989; Nentwig 2013; Tokareva et al. 2014), although the MI gland has also recently been characterized in some depth (Andersson et al. 2014). The knowledge gathered from the MA gland anatomy is often attributed as being representative of the *general* spider silk spinning process.

The MA gland is composed of three distinct anatomical zones (Figure 8): the spinning gland, where the proteins are accumulated and stored; the spinning duct, responsible for the alignment of soluble proteins into silk fibres; and, the spinneret which secretes the final silk fibre (Rey and Herrera-Valencia 2011; Tokareva et al. 2014). Each type of silk is stored at a very high concentration in the abdomen (30-50 wt.%) without onset of aggregation (Exler et al. 2007; Eisoldt et al. 2012). The high solubility is facilitated by the presence of the NTD and/or CTD interactions (see *1.3.1*) and by a high concentration of sodium chloride. High salt concentration decreases spidroin-spidroin interactions and lets spidroin-solvent interactions dominate, inhibiting silk oligomer formation within the gland (Exler et al. 2007).

Silk proteins are generally believed to form oligomeric macrostructures in solution to aid protein orientation during fibrillogenesis. Two competing theories have been put forward: one suggesting a protein micellar intermediate (Jin and Kaplan 2003) and the other consisting a liquid crystalline packing (Vollrath and Knight 2001). Transmission and scanning electron microscopies, along with atomic force microscopy and fluorescence spectroscopy, provide evidence of the micellar state for various recombinant spidroin proteins (Lammel et al. 2008; Lin et al. 2009; Breslauer et al. 2010; Silvers et al. 2010; Hofer et al. 2012; Xu et al. 2013). Evidence for the liquid crystal state was investigated by *in vivo* X-ray scattering on the gland (Riekel et al. 2000), transmission electron microscopy (Knight and Vollrath 1999) and modeling studies (De Luca and Rey 2006; Li et al. 2008; Cui et al. 2009). It is believed that at lower concentrations, spidroin self-assemble into micelles, while at higher concentrations, assemble into a liquid crystal creating hexagonal columns (Knight and Vollrath 1999; Barón 2001; Heim et al. 2009). Therefore, these theories are not mutually exclusive.

The silk proteins begin assembling for fibrillogenesis in the spinning duct. The spinning duct is folded into several S-shaped bends that progressively narrow (Vollrath and Knight 1999). In the spinning duct, the proteins align in either a nematic liquid crystal phase or in an elongated micellar arrangement (Heim et al. 2009). Sodium chloride ions soon start to exchange with more kosmotropic ions like potassium and phosphate (Foo et al. 2006; Geisler et al. 2008; Hardy et al. 2008). The pH also decreases from ~7 in the gland to ~5 late in the spinning duct (Andersson et al. 2014). As the duct narrows, the protein concentration slowly increases by dehydration of the protein. Finally, shear forces generated by the excretion of the silk fibre through the spinneret generate the silk fibre.

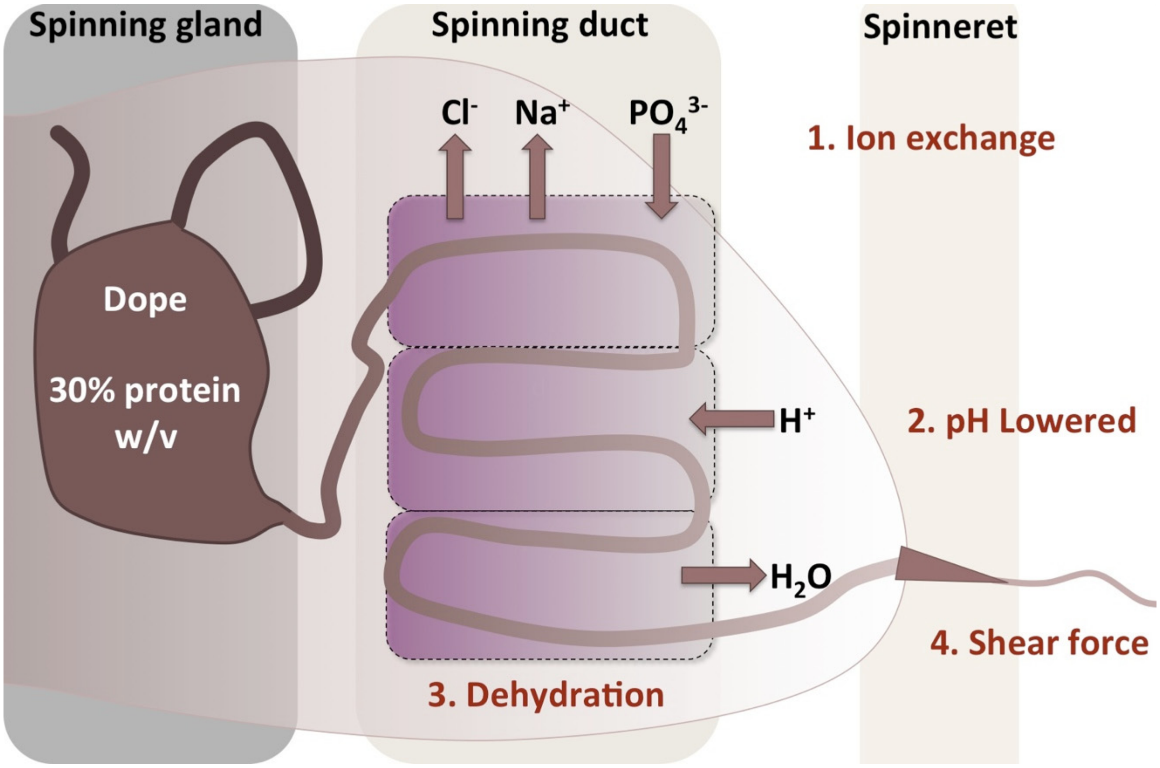


Figure 8 | Illustration of the spider's MA spinning gland separated in three parts.

1.3.1. The Involvement Of The Terminal Non-Repetitive Domains

Both the CTD and NTD have important functions in silk protein storage and fibre assembly. The CTD is a dimer in solution and aids in oligomer or micelle formation (Hagn et al. 2010a; Silvers et al. 2010; Jaudzems et al. 2012). During fibre assembly, CTDs facilitate the salting out process that keep silk proteins soluble within the gland and help control the alignment of the β-sheet repetitive sequence elements during fibre dehydration and shear forces (Hagn et al. 2010b).

The CTD dimer resembles a barrel-like shell enclosing two helices covalently joined by a disulfide bridge and further stabilized by salt bridges. The barrel-like dimeric structure shields hydrophobic residues from the solvent whilst the hydrophilic residues are positioned towards the solvent and ensure solubility of the protein.

The NTDs are highly sensitive to pH and are proposed to function as the pH-regulating relay (Askarieh et al. 2010; Silvers et al. 2010; Jaudzems et al. 2012). The NTD functions as the pH regulator because a change from pH 7 to 6 modifies the quaternary structure of NTDs from mostly monomeric at pH ~7 to a stable antiparallel non-covalent dimer at pH ~6-5 (Jaudzems et al. 2012; Andersson et al. 2014). Kronqvist *et al.* (Kronqvist et al. 2014) reported that sequential protonation of residues E79, E119, and E84 in the NTD of *Euprosthenops australis* MaSp1 are responsible for dimerization, while D40, R60, and K65 mediate inter-subunit electrostatic interactions. Stable dimerization of the NTD combined with the covalently-linked CTD dimer formation will result in spidroin multimerization, paving the way for the pathway to fibre formation (Hedhammar et al. 2008; Gaines et al. 2010).

While MA and MI silks need their NTD and CTD to form fibres (Ittah et al. 2006; Hedhammar et al. 2008; Hagn et al. 2010a; Hagn 2012; Gao et al. 2013), PySp2, TuSp, and AcSp1 can assemble into fibres without an NTD or CTD (Geurts et al. 2010; Gnesa et al. 2012; Xu et al. 2012a). It has been shown that the addition of a CTD to TuSp (Gnesa et al. 2012) and AcSp1 (Dr. Lingling Xu's data, *manuscript in review*) increases the extensibility and strength of the fibre, thus improving its mechanical properties.

1.3.2. Supercontraction Of Silk

A unique ability of spider silks is to undergo supercontraction at high humidity. Supercontraction is a material's ability to shrink in the presence of water. Supercontraction provides a spider web with strength and elasticity (Plaza et al. 2009) while decreasing the fibres' molecular stiffness and molecular order (Shi et al. 2014). If dragline silk is unrestrained in the presence of water, the fibre can shrink up to 60% in length and doubles in diameter (Bell et al. 2002; Liu et al. 2005; Agnarsson et al. 2009; Asakura et al. 2013a; Asakura et al. 2013b). If restrained, dragline silk generates stressed of up to 50 MPa (Agnarsson et al. 2009).

The amorphous proline-rich regions very efficiently incorporate water (Liu et al. 2008). When water is incorporated into wetted spider silk, the hydrogen-bonding network becomes disrupted and the molecular mobility and dynamics increase leading to a less organized arrangement (Yang et al. 2000; Shi et al. 2014). The spidroin motif GPGXX of MA silk, in particular, is primarily responsible for supercontraction (Yang et al. 2000). Results from ^{13}C cross-polarization magic angle spinning (CP-MAS) and ^{1}H solid-state NMR on MA silk show that significant chain mobility occurs in a portion of the fibre containing the following residues: Gly, Gln, Ser, Tyr, and Leu, but the crystalline portion remain rigid and undisturbed in the presence of water (Holland et al. 2008c).

1.4. STRUCTURAL STUDIES ON SPIDER SILK BY NMR

1.4.1. Basic Concepts In NMR Spectroscopy

NMR relies on the quantum effects induced by an external magnetic field (B_o) to the spin angular momentum of an atomic nucleus (Keeler 2002; Rule and Hitchens 2006). The spin angular momentum is an intrinsic property of a nucleus and requires a value greater than zero for an NMR active isotopes. The most commonly employed isotopes in biomolecular NMR have a spin quantum number of ½: ^{1}H, ^{13}C, and ^{15}N, with ^{1}H, ^{13}C and ^{15}N having 99.9% 1.1%, and 0.3% natural abundance respectively.

Nuclei that have spin angular momentum also possess a magnetic moment with magnitude directly proportional to the strength of applied external magnetic field behaving like tiny little bar magnets (Keeler 2002). When B_o is applied along an arbitrary defined *z*-axis, the z-component of the spin angular momentum is quantized parallel to or antiparallel to the magnetic field, dividing the spin populations that are governed by the Boltzmann distribution with a low energy level state called α state and a high energy level state termed the β state. The energy required for transition between the two states is given by:

$$\Delta E = \gamma \hbar B_o \tag{1.1}$$

where ΔE is the energy level difference between the states, γ is the gyromagnetic ratio of the nuclei, $\hbar$ is plank's constant, and B_o is the strength of the magnetic field in Tesla. The spins rotate around B_o at a frequency $\omega_{o,}$ called the Larmor frequency, which is equal to $-\gamma B_o$, in radians/second (Keeler 2002). The Larmor frequency is the resonance frequency that gives rise to the NMR phenomenon. Although ω_o is directly dependent on the NMR spectrometer magnetic field, the Larmor frequencies of nuclei in a molecular structure are also affected by the effects of partial shielding by electrons under the influence of the chemical shielding tensor (σ).

$$\Delta E = \gamma \hbar (1 - \sigma) B_o \tag{1.2}$$

The free induction decay (FID) is the observable NMR signal generated from non-equilibrium nuclear spin magnetization precessing about the z-axis induced by applying pulses of radiofrequency perpendicular to B_o (Rule and Hitchens 2006). The FID is then Fourier transformed from a time domain into a frequency domain in Hertz for visualization.

Chemical shifts (δ) are the nuclear resonance frequencies measured in NMR spectroscopy, normalized for strength of magnetic field relative to a reference standard (Keeler 2002; Rule and Hitchens 2006). Since nuclei resonance frequencies are dependent on the spectrometer frequency, a method to standardize across all spectrometers is to report and reference chemical shifts as parts per million (ppm). A chemical shift is defined by:

$$\delta\,(\text{ppm}) = \frac{\upsilon - \upsilon_{ref}}{\upsilon_{ref}} 10^{6} \qquad (1.3)$$

where υ is the Larmor frequency of the nucleus in Hertz and υ_{ref} is the frequency of the reference standard.

1.4.2. Structural Studies Of Spider Silk Motifs Within The Fibre

Only high-resolution solution-state spider silk structures have been discussed so far in this chapter. There have also been many efforts using ssNMR to determine and assign the functions of the small repetitive motifs of MA and MI silk in defining fibrous spider silk mechanical properties (Zhao and Asakura 2001; Asakura et al. 2013a; Asakura et al. 2013b).

The NMR chemical shift is both conformation and environmentally dependent (Rule and Hitchens 2006), hence an alanine in a β-sheet nanocrystal will have a different chemical shift than in random coil. Hydrogen-bonding also has an impact on the chemical shift (Zhao and Asakura 2001; Asakura et al. 2013b). Initial studies used 1D carbon spectra cross-polarization magic angle spinning (CP-MAS) ssNMR experiments to demonstrate that the crystalline A_n sequence forms the two-component nature of β-sheet nanocrystals with two components in MA (Simmons et al. 1994) and MI (Liivak et al. 1997) silk; one component is well-oriented and the other poorly oriented (Simmons et al. 1996). Although A_n, is found mostly in crystalline β-sheets, it can also form 3_{10}- and α-helices.

ssNMR has also played an integral part in characterizing the molecular structure of non-crystalline glycine-rich domains (GGX, GPGXX) in MA and MI silk fibres (van Beek et al. 2002; Eles and Michal 2004a). The soft, disordered matrix surrounding β-sheet nanocrystals was once thought to be amorphous (Gosline et al. 1984; Gosline et al. 1986), but more recently proven to contain 3_{10}-helices (Kümmerlen et al. 1996; Valluzzi et al. 1999; van Beek et al. 2003; Holland et al. 2008b). Evaluation of dragline silk backbone torsion angles ϕ/ψ by double-quantum spectroscopy (DOQSY) strongly suggested that some glycine are found in β-sheets and others in a 3_{10}-helix conformation (van Beek et al. 2002; van Beek et al. 2003). Both regions were shown to be strongly oriented parallel to the fibre long-axis by Direction Exchange with Correlation for Orientation Distribution Evaluation and Reconstruction (DECODER) experiments.

With the advent of multidimensional NMR solid-state experiments, it has been possible to acquire valuable information on the 3D structuring of spider silk motifs within the fibre. ^{13}C-^{13}C dipolar assisted rotational resonance (DARR) experiments (Takegoshi et al. 2001) with ^{13}C-alanine enriched spider dragline silk were collected at various mixing times to determine short- and long-range contacts between ^{13}C nuclei (Holland et al. 2008a). The refocused INADEQUATE (Incredible Natural Abundance DoublE QUAntum Transfer Experiment) NMR experiment is also useful for acquiring 2D, through-bond, homonuclear correlation spectra in disordered solids and useful for assigning chemical shifts (Holland et al. 2008b). High-resolution ssNMR is achieved with high MAS spinning speeds of >30 kHz coupled with high fields (e.g. 800 and 930 MHz) allows investigation of the hydrogen-bonded structure of the small silk-like $(GA)_n$ (Yamauchi et al. 2000; Suzuki et al. 2009) and A_3 peptides (Yazawa et al. 2012). Accurate chemical shift assignments of the ^1H in the tripeptide A_3 were made for anti-parallel and parallel β-sheets and, from these, a hydrogen-bond network for both forms was devised (Yazawa et al. 2012).

NMR has also been used to correlate the mechanical property differences observed with different spider silk spinning rates and fibre stretching to differences in structure and motif orientation (Eles and Michal 2004b; Cloutier et al. 2011). Spinning speeds during fibre production, as well as fibre stretching, greatly affect the physical and mechanical properties of the silks (Cloutier et al. 2011). Observing fibres at different reeling speeds with different stretching by ssNMR revealed that the secondary structure content is unaffected; rather, the molecular arrangement within a fibre is altered, with less organization at faster reeling speeds (Glišovic et al. 2008). NMR relaxation-based dynamic measurements for alanine and glycine, used to probe molecular motion within the fibre, demonstrates an increase from MA to MI to silkworm silk, showing that the whole fibre motions are silk type-specific and highly dependent on chain packing and intermolecular interactions.

1.5. PROJECT GOALS

AcSp1 from *Argiope trafisciata* consists of a 200 amino acid sequence repeated consecutively at least 14 times and containing none of the characteristic short repetitive motifs found in the other types of spider silk (Lewis 2006; Ayoub et al. 2013). Therefore, any structural information or self-assembly mechanisms derived for the other types of silks are likely not trivially relatable to AcSp1. At the start of this project, there was no structure available for AcSp1; therefore, I set out to structurally characterize the AcSp1 protein from the solution-state through to the fibrous state by NMR. My project goals were:

1) Determine the solution structure of the 200 amino acid repeat unit
2) Analyze dynamics within the protein using spin relaxation measurements
3) Study the structure of AcSp1 within a fibre via solid-state NMR and locate the regions of AcSp1 with α-helical versus β-sheet content.
4) Propose a model for the transitioning of the AcSp1 into fibres via the liquid crystal state or protein micellar state.

To date, only 11 high-resolution spider silk structures have been published (Table 2), three of which are of repetitive domains (RP1 and RP2 from TuSp1 and RP from AcSp1) (Lin et al. 2009; Wang et al. 2012) described in the earlier sections, while the others from are of CTD and NTD regions (Askarieh et al. 2010; Hagn et al. 2010a; Gao et al. 2013; Wang et al. 2014). Our understanding of the structural mechanisms involved in fibre assembly and properties are very limited, implying a need for additional structural data for spider silks.

Within this thesis I first start by describing the relationship of the NMR chemical shift to protein secondary structure, its dependence on solvent dielectric and on protein secondary structure. I then follow with the description of the first full NMR structure of an AcSp1 repeat unit (W_1) from *Argiope trifasciata*. Because W_1 in solution does not form fibres, but W_2 or larger does (Xu et al. 2012a), I then compare the W_1 structure to that in a concatenated protein of two repetitive domains (W_2). To do so, intein *trans*-splicing technology (Southworth et al. 1998; Wu et al. 1998) was employed to selectively label each W domain separately with ^{13}C and/or ^{15}N isotopes. Following structural comparison, I use NMR spin relaxation analysis to compare and contrast polypeptide backbone dynamics of W_1 and W_2 in solution. Finally, the protein segments essential for structural transition during fibre formation were mapped in the W repeat unit by titration with denaturants and a detergent. Placing my work in context, fibres can be pulled directly from the NMR samples of W_2. Therefore, the solution-state structure and dynamics data that I present for AcSp1 are fully functionally-relevant and provide the first direct atomic-level insight into the fibre-forming state of AcSp1.

Table 2 | Summary of the published spider silk structures available in the Protein Data Bank (PDB).

PDB	Protein	Species	Residue count	Method	Resolution	Reference
3LR2 3LR6 3LR8	MaSp1	*Euprosthenops australis*	137	X-ray	1.7	(Askarieh et al. 2010)
4FBS	NTD of MaSp1	*Euprosthenops australis*	137	X-ray	1.7	(Jaudzems et al. 2012)
2KHM	CTD of MaSp1	*Araneus diadematus*	140	NMR		(Hagn et al. 2010a)
2MFZ	CTD of MiSp	*Araneus ventricosus*	125	NMR		(Andersson et al. 2014)
2LPI	NTD A72RMutant of MaSp1	*Euprosthenops australis*	137	NMR		(Jaudzems et al. 2012)
2LPJ	NTD of MaSp1 at pH 7.2	*Euprosthenops australis*	137	NMR		(Jaudzems et al. 2012)
2LTH	NTD of MaSp1 at pH 5.5	*Euprosthenops australis*	137	NMR		(Kronqvist et al. 2014)
2K3N	TuSp1-RP1	*Nephila antipodiana*	160	NMR		(Lin et al. 2009)
2K3O	TuSp1-RP2	*Nephila antipodiana*	129	NMR		(Lin et al. 2009)
2MAB	CTD of AcSp1	*Nephila antipodiana*	109	NMR		(Wang et al. 2014)
2MOM	CTD and NTD	*Nephila antipodiana*	107	NMR		(Gao et al. 2013)

CHAPTER 2. THE PREDICTIVE ACCURACY OF CHEMICAL SHIFT SECONDARY STRUCTURE ASSESSMENT (BASED ON TREMBLAY ET AL. (2010))

2.1. INTRODUCTION

2.1.1. *Relation between chemical shifts and protein secondary structure*

In biomolecular NMR, chemical shifts are often highly amino acid specific and can provide a strong indication of polypeptide secondary structure (Wishart et al. 1992; Thanabal et al. 1994; Wishart and Sykes 1994). Evaluation of chemical shifts typically relies on subtraction of representative random coil chemical shifts (RCCS) from an observed chemical shift in order to provide a secondary chemical shift (Δδ) independent of the identity of the amino acid. The Δδ value experienced by H^{α}, H^{β}, C^{α}, C^{β} and C' backbone nuclei due to secondary structuring in proteins correlates with the torsional angle ψ, which dictates the position of the $^{\alpha}H$-$^{\alpha}C$ bond in relation to the C' (Dalgarno et al. 1983). Δδ values are frequently used either directly as restraints in structure calculations (Kuszewski et al. 1996a) or in calculation of chemical shift index (CSI) parameters used for identifying regions of secondary structure (Wishart et al. 1995b). In structural biology, chemical shifts are most often used to highlight regions of protein secondary structure and are implemented during structure calculation protocols in parallel with other distance, orientation or dihedral angle based restraints (Evans 1995).

2.1.2. *Programs Using Chemical Shifts As Secondary Structure Predictors*

Since the first report that chemical shifts are systematically affected by polypeptide secondary structure by Dalgarno *et al.* (Dalgarno et al. 1983), the relationship between chemical shifts and secondary structure has been extensively studied through empirical correlations (Zuiderweg et al. 1989; Szilágyi and Jardetzky 1989; Williamson 1990; Osapay and Case 1991; Spera and Bax 1991; Williamson et al. 1992; Le and Oldfield 1994; Beger and Bolton 1997; Iwadate et al. 1999) determined through probabilistic methods (Lukin et al. 1997; Wang and Jardetzky 2002b), quantum mechanically (Jiao et al. 1993; Case 1998; Case 2000; Xu and Case 2001), and derived experimentally (Bundi et al. 1975; Wüthrich 1986; Thanabal et al. 1994; Merutka et al.

1995; Wishart et al. 1995a; Schwarzinger et al. 2000; Schwarzinger et al. 2001). In cases with complete chemical shift assignment, including ^{13}C and ^{15}N nuclei, bioinformatics approaches such as TALOS (Cornilescu et al. 1999), TALOS+ (Shen et al. 2009a) and PREDITOR (Berjanskii et al. 2006) use databases to produce comparable protein fragments, in some cases using homology considerations, to define ϕ and ψ dihedral angles. DANGLE extends these approaches through inclusion of Bayesian inference to improve sampling of less populated regions of Ramachandran space (Cheung et al. 2010). Algorithms which build upon these concepts for complete protein structural restraint generation by chemical shifts alone include CHESHIRE (Cavalli et al. 2007), CS-ROSETTA (Shen et al. 2008) and CS23D (Wishart et al. 2008), with CS-ROSETTA having been more recently extended to support cases of incomplete chemical shift assignment (Shen et al. 2009b). Although these methods do not explicitly account for variation in dielectric constants (ε) or environment between protein interior and exterior, the chemical shift comparison databases inherently consist of structural fragments in a wide variety of environments.

With incomplete chemical shift assignments, i.e. if isotope labeling is unfeasible, or if the direct comparability of the protein database employed for dihedral angle prediction to the protein studied is questionable, $\Delta\delta$ and CSI type approaches based on comparison to RCCS datasets are employed. Our ability to accurately assess secondary structure from RCCS is dependent on solvent conditions (Thanabal et al. 1994; Plaxco et al. 1997; Avbelj and Baldwin 2004), effects of nearest neighbours (Wishart et al. 1995a; Wishart et al. 1995b; Schwarzinger et al. 2001), temperature (Merutka et al. 1995; Wishart et al. 1995b; Tonan and Ikawa 2003), and spectral referencing (Wishart et al. 1995a). In studies examining empirical relationships between chemical shifts and secondary structure, typically derived from datasets of globular proteins studied in aqueous conditions, it is difficult to untangle the effects of environmental differences between nuclei exposed to aqueous solution or buried in the globular core of a protein vs. the effects of polypeptide backbone structural differences. Furthermore, all currently available experimentally measured RCCS datasets have been determined in aqueous solution.

2.1.3. Effect Of Solvent On Chemical Shifts

The solvent environment can perturb chemical shifts of solutes by a number of mechanisms, including effects of ε. Dielectric constant affects chemical shifts by altering the transmittance of the electrostatic field (Søndergaard et al. 2008). Other mechanisms by which solvent environment can affect chemical shifts include modification of preferred bond torsion angles, differences in hydrogen bonding with the solvent, van der Waals interactions, ring currents, and electric charge (Hass et al. 2008). Solvent-induced modification of protonation state at ionizable sites would also be expected to perturb chemical shifts of nearby nuclei (Kim et al. 2009). These other properties are solvent-specific and are not always correlated to the dielectric of the solvent. Given that chemical shifts are influenced by solvent (Shenderovich et al. 2001; Tonan and Ikawa 2003), there is a strong incentive to ensure that the environment used to determine RCCS matches that in which the polypeptide or protein is being studied. In particular, membrane proteins or fibrillar self-assembled proteins are found in liquid crystalline environments with substantially lower ε than water (Donald et al. 2006). Furthermore, the hydrophobic core of a globular protein is better represented as a region of low ε in comparison to water (García-Moreno et al. 1997).

2.1.4. Project Description

For this chapter, I present RCCS values determined in non-aqueous environments and examine the effects of perturbation of solvent environment on our ability to directly correlate Δδ to secondary structure. A set of 21 random coil peptides of sequence GGXAGG, where X is any of the 20 naturally occurring amino acids or the modified amino acid 4-hydroxyproline, was studied in two different media with ε lower than water. Dimethyl sulfoxide (DMSO; ε = 47.5) was used as a mimic of the bilayer/water interface (ε ~40) (Brockman 1994; Koehorst et al. 2008). A ternary solvent mixture (hereafter referred to as the trisolvent system; theoretical ε ~37.8) composed of chloroform, methanol, and water in a 4:4:1 ratio (by volume) was also investigated due to its membrane mimetic properties (Slepkov et al. 2005). Notably, it is not clear which solvent perturbation effect(s) of those mentioned above would be most significant in the DMSO or the trisolvent system. However, since previous structural studies have successfully demonstrated both of these solvents to be reasonable membrane mimetics (reviewed in Rainey *et al.* (Rainey et al. 2006)), providing a good approximation of both the lower ε and decreased availability of H-bonding donors and acceptors in a membrane environment, RCCS values determined in these environments should also be applicable in a membrane environment. The python program CS-CHEMeleon, web-mounted at http://structbio.biochem.dal.ca/jrainey/CSChem, written by a former summer student under my guidance, was implemented to allow rapid analysis of sizeable sets of globular and membrane protein structures solved by NMR methods with published chemical shifts. This approach allowed comprehensive statistical comparison of the relative ability to predict protein secondary structure in membrane and non-membrane proteins using Δδ values derived from our sets vs. three other sets of RCCS values, providing the first differential analysis (to our knowledge) between these classes of proteins.

2.2. MATERIALS AND METHODS

2.2.1. Materials

9-Fluorenylmethoxycarbonyl (Fmoc) protected amino acids, Rink Amide AM Sure Resin (0.65mmol/g loading) and coupling reagents were obtained from AAPPTec (Louisville, KY), except for Fmoc protected ^{15}N-labelled Gly (Cambridge Isotope Laboratories, Andover, MA) and Ala (C/D/N Isotopes, Pointe-Claire, QC). N,N-dimethylformamide (sequencing grade) and acetonitrile (high performance liquid chromatography (HPLC) grade) were obtained from Fisher Scientific (Ottawa, ON). The deuterated solvents DMSO-d_6 and methanol-d_3 (CD_3OH) were acquired from C/D/N isotopes (Pointe-Claire, QC) while chloroform ($CDCl_3$) was obtained from Sigma-Aldrich (Oakville, ON). The chemical shift standard 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) was obtained from Wilmad (Buena, NJ). All other chemicals were purchased as biotechnology, high performance liquid chromatography (HPLC) or reagent grade, as appropriate, from Sigma Aldrich (Oakville, ON). All reagents and chemicals were used without further purification, unless otherwise specified. NMR samples were prepared in either 5 mm O.D. Wilmad 535-PP-7 or -8 tubes (DMSO and aqueous samples) or 535-TR-8 screw cap tubes (trisolvent mixture).

2.2.2. *Peptide synthesis and purification*

Peptides with sequence Ac-Gly-Gly-X-Ala-Gly-Gly-NH_2 (X being any of the 20 L-amino acids or the modified amino acid 4-hydroxyproline) were synthesized at ~0.2 mmol scale on a semi-automatic solid-phase peptide synthesizer (Endeavor 90, AAPTec) on Rink AM resin. Protocols were as outlined in Langelaan *et al.* (Langelaan et al. 2009) with the exception that peptides were N-terminally acetylated with anhydrous acetic anhydride (5 eq. to resin) and the cleavage cocktails used were appropriate for the side-chain protecting groups in a given peptide (Guy and Fields 1997). Reverse phase HPLC (Beckman System Gold, Fullerton, CA) purification was performed using a C_{18} column (5μm particle, 120 Å pore size, 10x250 mm AAPPTec Spirit) at a flow rate of 3.0 mL/min. A linear water/acetonitrile (A/B) gradient was used from 98%A/2%B to 60%A/30%B over 25 min. Peptide identities and purities were confirmed by NMR spectroscopy.

2.2.3. *Circular Dichroism (CD) Spectropolarimetry*

Far-ultraviolet (far-UV) CD spectropolarimetry (J-810, Jasco, Easton, MD) was performed in 2,2,2-trifluoroethanol (Sigma, 99% non-deuterated NMR grade) for 5 peptides (peptide concentration determined by UV absorbance at 210 nm using a 1.0 cm path length quartz cuvette (Hellma, Müllheim, Germany) on a diode array spectrophotometer (Hewlett Packard, 8452A) for ellipticity normalization; X = G, I, M, P, and V) at 25°C (controlled with a NESLAB RTE-111 bath, Thermo Scientific, Newington, NH). Three repetitions (190-260 nm, 1 nm steps, 20 nm/min) were performed and averaged for all trials of each peptide in a 0.1 mm path length quartz cuvette (Hellma, Müllheim, Germany). All spectra were blank subtracted and a moving average of 3 was performed in Microsoft Excel. All data was converted to mean residue ellipticity [θ] as described previously (Langelaan et al. 2009).

2.2.4. *NMR Spectroscopy*

Experiments were performed at 298 K either on the Nuclear Magnetic Resonance Research Resource (NMR^3, Dalhousie University) 11.7 T Avance II spectrometer (Bruker Canada, Milton, ON) equipped with a 5 mm broadband observed (BBO) probe or the National Research Council Institute for Marine Biosciences (NRC-IMB, Halifax, NS) 16.4 T Avance III (Bruker Canada) spectrometer equipped with a 5 mm indirect detection TCI cryoprobe. Samples (10 mM peptide, 5 mM DSS, 600 μL) were prepared in DMSO-d_6 or $CDCl_3$:CD_3OH:H_2O (4:4:1 by volume; mixture pH 5.5±1). An aqueous sample of the Hyp peptide was using the conditions of Wishart *et al.* (Wishart et al. 1995a). Spectra collected in DMSO are reported indirectly referenced to a value of 0 ppm for aqueous DSS using intermediate trimethylsilane (TMS) shifts (^{1}H (Hoffman 2006), ^{13}C (Wishart et al. 1995b)). Shifts in the trisolvent system (and of the Hyp in aqueous solution) are reported relative to internal DSS at 0 ppm. The dielectric constant of the trisolvent system (ε_s) was estimated using a combination of equations published by Abraham *et al.* (Abraham et al. 1966) and Amirjahed and Blake (Amirjahed and Blake 1975) :

$$\varepsilon_s = \sum \frac{\varepsilon_n - 1}{\varepsilon_n + 2} M_n \qquad (2.1)$$

where ε_n is the dielectric constant of solvent n with mole fraction M_n and the sum is carried out over all components of the solvent mixture.

Unless otherwise specified, experiments were performed at 11.7 T (details in Table 3). In cases where 1D ^{1}H NMR experiments were ambiguous, 1D nuclear Overhauser effect spectroscopy experiments (0.5 or 1 sec mixing time) with 64 scans were performed by irradiation of the X amino acid H^α proton for identification of amide protons. A combination of distortionless enhancement by polarization transfer with modification for the detection of quaternary nuclei (DEPTQ-135; (Burger and Bigler 1998)) and 2D ^{13}C-^{1}H sensitivity-enhanced heteronuclear single quantum coherence (HSQC; (Farrow et al. 1994)) allowed accurate ^{13}C shift assignment for all 21 peptides in all conditions. The ^{15}N-labelled Gly and Ala peptides were analyzed at 16.4 T. Sensitivity-enhanced ^{15}N-^{1}H HSQCs (Farrow et al. 1994) were used to assign the labeled X-position amino acid ^{15}N and ^{1}H chemical shifts. 2D total correlation spectroscopy (TOCSY; 60 ms DIPSI-2 mixing time) was used to assign H^a and H^β for Ala and H^α for Gly. All 1D experiments were processed and analyzed using TopSpin 1.6 (Bruker) and 2D experiments were processed using NMRpipe (Delaglio et al. 1995).

Table 3 | Details of NMR experiments performed in Chapter 2.

Experiment	Bruker pulse program	Number of scans	Relaxation delay (s)	Sweepwidth (Hz)	Acquisition time (s)	Acquired points & increments
1D	zg30 or zgcpgppr	64	2.00	7002	3.28	32768
DEPTQ135	deptqsp	4000	1.00	33784	0.74	50670
^{13}C-^{1}H HSQC	hspcetgpsisp.2.	16	1.52	F1 33784 F2 4807	0.11	1024x128
^{15}N-^{1}H HSQC	hspcetfpf3gpsi	4	1.00	F1 3225 F2 8417	0.12	2048x64
2D^{1}H-^{1}H TOCSY	dipsi2ph	8	1.50	F1 5122 F2 7183	0.28	4096x32

2.2.5. Comparative Evaluation Of Random Coil Chemical Shift Tables

The DMSO and trisolvent system chemical shifts presented in this chapter were compared to two sets of published aqueous chemical shifts (Wishart et al. 1995a; Schwarzinger et al. 2000; Wang and Jardetzky 2002a) and one derived by probability-based methods (Wang and Jardetzky 2002a) for their ability to assess secondary structure in proteins. All 33 transmembrane (TM) protein NMR-STAR files accessible in the BioMagResBank in 2009 (BMRB: http://www.bmrb.wisc.edu/ (Ulrich et al. 2008)) along with their matching structural data from the protein data bank of transmembrane proteins (PDB_TM: http://pdbtm.enzim.hu/ (Tusnády et al. 2004)) were acquired (See Appendix A for tables). Chemical shifts and structures for a randomly selected set of 107 non-membrane proteins whose structures were determined in aqueous conditions (AQ) were acquired from the BMRB and the Protein Data Bank (www.pdb.org (Berman et al. 2000)), respectively. All PDB files with less than 10 models were eliminated; 2 cases (*i.e.* 4 structures out of 107) of identical proteins having differing structures and chemical shift were retained (PDB entries 2KAX and 2KAY; 1YZA and 1YZC), and an instance of protein structures for a pair of isoforms having different lengths and C-terminal extensions was retained (PDB entries 2YSE and 2ZAJ). Statistical examination was performed on concatenated datasets containing all proteins and structural models of each class (TM or AQ) rather than on individual NMR-STAR or PDB files. NMR-STAR files not referenced to DSS, with the assumption that those with unspecified reference were referenced to DSS, were indirectly referenced to DSS when the standard was trimethylsilyl-2,2,3,3-tetradeuteropropionic acid (TSP), TMS, or H_2O for ^1H and TSP, TMS, or dioxane for ^{13}C (Wishart et al. 1995b; Hoffman 2003; Hoffman 2006).

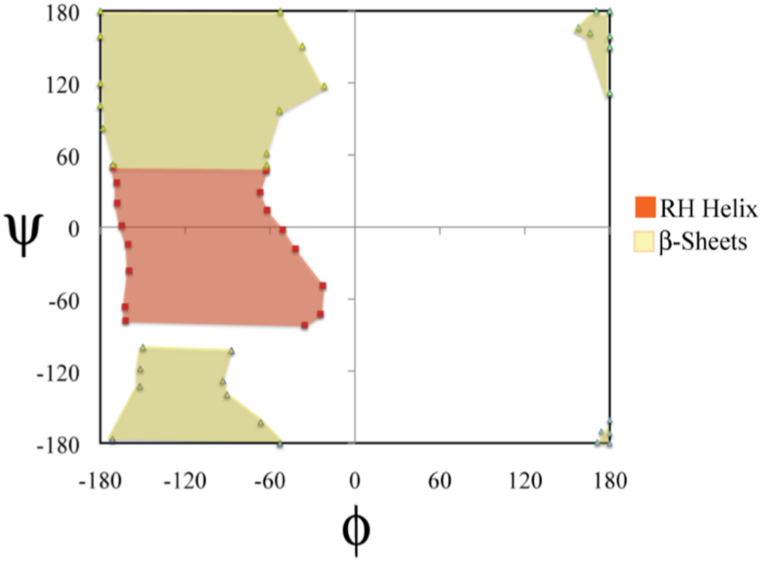


Figure 9 | Ramachandran plot showing the boundaries used to define helical and sheet structures in CS-CHEMeleon. Angles were chosen using the structural boundaries defined by Lovell *et al.* (Lovell et al. 2003) and slopes were calculated between these angles to define a shape that delimits each structure type (α-helices (RH helix)).

For the purpose of this study, the structure at a given amino acid residue “i” was determined from the average ϕ and ψ dihedral angles of the NMR-derived structural ensemble. For cases of localized averaging, a 5-residue sliding window average (residues i-2 to i+2) averaged over all ensemble members was employed. The consensus was deemed the “true” structure of the peptide or protein. Backbone dihedral angles were calculated and the regions defining various secondary structures were geometrically derived from those of Lovell *et al.* (Lovell et al. 2003). The reliability of local structural averaging was verified against 10 randomly selected proteins from the AQ dataset by qualitative comparison of our secondary structure output (helix, sheet, or coil) to that of PROMOTIF v1.0 (Hutchinson and Thornton 1996)}.

The chemical shift based secondary structure prediction (*i.e.,* helix, coil or sheet) for each residue was determined by comparison of the magnitude and sign of the Δδ relative to a specified threshold for a given nucleus, with Δδ values calculated including the correction factors of Schwarzinger *et al.* (Schwarzinger et al. 2001) for nearest neighbour effects. Optimized thresholds for nuclei were determined iteratively for each RCCS table, with calculation of agreement using thresholds of -1 ppm to +1 ppm (0.01 ppm interval) for H^{α} and of -2 to +2 (0.1 ppm interval) for C^{α} and C^{β} to obtain the best agreement between PDB file based dihedral angles and Δδ for the full AQ dataset (the TM dataset contained too few amino acids). Consensus Δδ-based predictions were defined as follows, based on the nuclei out of H^{α}, C^{α}, and C^{β} that are reported in an NMR-STAR file. If the NMR-STAR file contained chemical shifts for at lease two of those nuclei, the consensus prediction required agreement of at least 2 Δδ-based threshold tests. If there was no consensus, then the amino acid was designated as part of a coiled structure. Local averaging, when employed, was carried out identically to the local structural averaging using a 5-residue sliding window. The numbers of correct and incorrect predictions were compared for each threshold during iteration and the 'correct-incorrect' predictions were normalized to a percentage. The optimal threshold for that atom's structure type was at the highest normalization value. The standard thresholds of 0.1 ppm for ^{1}H and 0.7 ppm for ^{13}C introduced by Wishart *et al.* (Wishart et al. 1992) were used alongside the optimized thresholds determined for each RCCS dataset during analysis. The accuracy of RCCS assessment of secondary structure was expressed as a total percentage of correctly assessed structure for both the AQ and TM datasets, subdivided by RCCS set, atom type and secondary structure threshold. Glycine and proline were not used for comparison since they have different Ramachandran angle preferences from the other amino acids in the same secondary structures (Lovell et al. 2003) and also have a greatly elevated Δδ threshold (Wishart et al. 1992).

2.2.6. CS-CHEMeleon Implementation

For data analysis, a python 2.5 program named CS-CHEMeleon was written by Aaron Banks, a previous summer student, with my guidance for the design and functionality of the program. This has been configured to run from a web-based graphical user interface (freely available at http://structbio.biochem.dal.ca/jrainey/CSChem). Using a chemical shift file uploaded in NMR-STAR format (Ulrich et al. 2008), Δδ values are calculated directly using the

RCCS dataset(s) specified by the user. By default, two experimentally determined aqueous chemical shift datasets (Wishart et al. 1995a; Schwarzinger et al. 2000), one aqueous probability-based RCCS table (Wang and Jardetzky 2002b), and the DMSO and trisolvent tables presented herein can be used. Any desired alternative RCCS table may also be uploaded and used. $\Delta\delta$ values may be determined with or without accounting for nearest neighbours, using the correction factors published by Schwarzinger *et al.* (Schwarzinger et al. 2001). Evaluation and comparison of $\Delta\delta$ values may be performed graphically within the web browser and/or $\Delta\delta$ values may be downloaded in ASCII format for offline analysis. The user also has the option to only assess secondary structure for residues inside or outside the membrane as defined in the associated extensible mark-up language PDB_TM file when investigating TM proteins.

CS-CHEMeleon also calculates ϕ and ψ dihedral angles upon demand for an uploaded NMR structural ensemble (or single structure) in PDB file format, allowing comparison between the secondary structure given by the region of the Ramachandran plot (Lovell et al. 2003) and the predicted secondary structure from a given RCCS table using the structural thresholds of Wishart *et al.* (Wishart et al. 1995a) or the optimized iterated thresholds presented here. The user can also make use of local sliding window structural averaging (in either the NMR-STAR, PDB or both files) if the assessment yields poor correlation to the PDB structure. The number of residues averaged (recommended 5) and the minimum proportion of agreement for a consensus definition over the window (recommended 0.51) are both defined by the user.

2.3. RESULTS AND DISCUSSION

2.3.1. Verification of peptide random coil character

The established random coil peptide series of Wishart *et al.* (Wishart et al. 1995a), with sequence Ac-GGXAGG-NH_2, was used herein. In order to ensure that random coil character was maintained in organic solvent conditions, far-UV CD spectropolarimetry was performed on a sample set of 5 peptides containing a variety of X amino acids (Figure 10). Because DMSO is not a suitable solvent for CD due to absorbance in the far-UV region and because of incomplete solubility of all 21 peptides in the trisolvent system, the α-helix inducing solvent trifluoroethanol (TFE) (Merutka et al. 1995) was used. All five peptides, including some with X residues having high helical propensity in aqueous (Blaber et al. 1993) or membrane-mimetic environments (Li and Deber 1994), show a strong negative band at ~198 nm characteristic of a random coil (Greenfield and Fasman 1969) *vs.* the positive band expected of a helix or sheet structure. Weak negative ellipticity in the ~210-230 nm region is also observed in most of these peptides, with varying strengths. The CD banding in this region is not indicative of sheet or helix, particularly in the absence of a positive band at ~200 nm. Also informatively, the GGPAGG spectrum (most likely of all of the peptides to have polyproline-II character) contains no band structure indicative of polyproline-II character (Chellgren and Creamer 2004; Rath et al. 2005). We cannot explicitly rule out the weak negative ellipticity at 210-230 nm being caused by increased favourability of intramolecular or intermolecular interactions within or between peptides in TFE, relative to aqueous conditions. However, since all peptides examined had predominantly random coil CD characteristics even in an α-helical structure-inducing medium, we feel that the Ac-GGXAGG-NH_2 sequence, which was random coil in aqueous medium (Wishart et al. 1995a), should serve as a valid random coil model system in the non-structure inducing DMSO and trisolvent system environments.

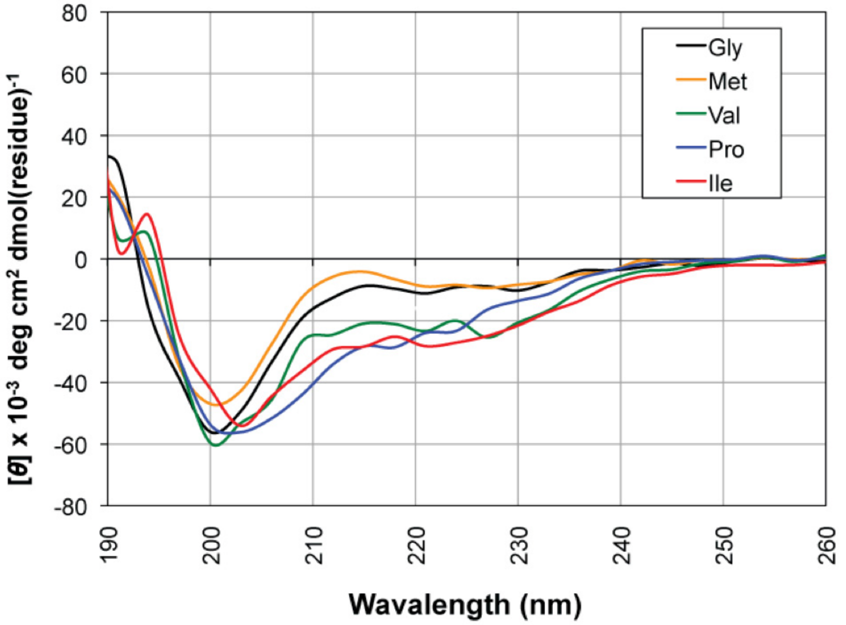


Figure 10 | Far-ultraviolet circular dichroism spectra of five random coil peptides (sequence Acetyl-GGXAGG-NH_2, with X given in legend) in 2,2,2-trifluoroethanol. The shown spectra are averages of three blank subtracted trials, with a weighted 3 nm sliding-window average applied.

2.3.2. Random Coil Chemical Shifts

The complete set of RCCS measured in DMSO is reported in Table 4 and Table 5 and in the trisolvent system in Table 6 and Table 7. Only 10 of the 21 random coil peptides dissolved in the trisolvent system with no discernable pattern in solubility. For comparison, the chemical shifts for Hyp measured using identical conditions to Wishart *et al.* (Wishart et al. 1995a) are as followed: H^{α} 4.53; H^{β} 2.35 and 2.06; H^{γ} 4.62; H^{δ} 3.79 and 3.57; C^{α} 61.56; C^{β} 39.76; C^{γ} 72.5; and, C^{δ} 57.3. Although DSS was incorporated as an internal standard in the DMSO samples, DSS and DMSO interact. Chemical shifts in these samples were therefore indirectly referenced to DSS in water using published reference values (Wishart et al. 1995b; Hoffman 2003; Hoffman 2006) to provide direct comparability for experimental data acquired in aqueous solution using the accepted biomolecular chemical shift standard of DSS (Wishart et al. 1995a). In contrast, the shifts reported in the trisolvent system were internally referenced to DSS since there is no straightforward way to indirectly reference in a ternary solvent system. Phase separation is noticeable in this solvent mixture (Slepkov et al. 2005), so it is likely that

DSS referencing is very similar to that of DSS in water since DSS would be most soluble in water-rich components of the mixture. However, this uncertainty in referencing should be taken into account when employing the trisolvent-derived shifts. Ideally, RCCS would also be determined in a solvent of much lower ε for direct comparability to phospholipid tail-group regions or in the core of a globular protein. However, examination of a variety of solvents with lower ε (~4-20) demonstrated uniformly extremely poor solubilization of the random coil peptides. Designing a more hydrophobic, but still random coil, peptide would likely be required in order to obtain RCCS in a low ε medium.

The Ac-GGXAGG-NH_2 peptide series allows determination of the effect of solvent environment in terms of decreased ε and the other perturbation factors discussed in the introduction to this chapter upon RCCS. These RCCS are most directly comparable to the experimental aqueous RCCS dataset of Wishart *et al.* (Wishart et al. 1995a), since the same peptide series was employed. Comparison was also performed to the experimental aqueous RCCS dataset of Schwarzinger et *al.* (Schwarzinger et al. 2000) and to the statistically derived amino acid chemical shift dataset of Wang and Jardetzky (Wang and Jardetzky 2002b). RCCS perturbations are evident to different degrees for different amino acids and, in some instances, for different RCCS comparison sets (Figure 11). Perturbations relative to aqueous or statistically derived RCCS may be presumed to be arising from decreased ε, from changes to favoured dihedral angle ranges within a given random coil peptide from other non-ε derived properties of the solvent and, in the case of ionizable residues or H-bonding donor/acceptor side-chains, from protonation or H-bonding state. The aliphatic residues, for example, tend to be strongly perturbed in DMSO. Ionizable or H-bonding donor/acceptor side-chains, such as Asn and Gln, are also strongly perturbed in DMSO. As would be expected (Quirt et al. 1974), C^{α} is also strongly affected in the acidic residues, which are clearly protonated on the side-chain carboxylic acid in DMSO (Table 4) but typically deprotonated in aqueous conditions. H^{α} and C^{α} in DMSO generally (but not uniformly) experience opposite trends relative to aqueous RCCS, with H^{α} being deshielded while C^{α} is shielded. For both nuclei, more than half of the amino acids' random coil values change by more than their respective structural thresholds defined by Wishart *et al.* (Wishart et al. 1995a), which should therefore be defined as a significant change.

In comparison, the chemical shifts determined in the trisolvent system show different trends from those observed with DMSO (Figure 11). Although only 10 amino acids can be compared, a change in solvent from aqueous conditions to the trisolvent system does cause perturbations in both H$^\alpha$ and C$^\alpha$ chemical shifts. In comparison to the DMSO environment, a lower proportion of chemical shifts are perturbed by more than the standard thresholds in the trisolvent system and, as a set, they do not show any trend in shielding and deshielding although the theoretical ε is lower then that of DMSO. This implies that effects of ε alone are not entirely responsible for the generally larger chemical shift perturbations observed in DMSO. Since some degree of phase separation is obvious in the trisolvent system (Slepkov et al. 2005), it is possible that preferential peptide solvation in aqueous-rich phases is giving rise to a decreased difference from RCCS derived in aqueous conditions. This hypothesis, however, is at odds with the insolubility of 11/21 of the random coil peptides in the trisolvent system.

Table 4 | ^{1}H random coil chemical shifts for peptides of sequence GGXAGG measured in dimethyl sulfoxide.

X residue	NH	H$^\alpha$	H$^\beta$	Others
Ala	8.21	4.47	1.43	
Arg	8.32	4.48	1.91, 1.74	γCH$_2$ 1.51, 1.51 δCH$_2$ 3.1
Asn	8.40	4.74	2.78, 2.65	γNH$_2$ 7.23, 7.67
Asp	8.41	4.78	2.93, 2.72	γOH 12.6
Cys (red)	8.44	4.60	3.00, 2.92	SH 2.32
Gln	8.30	4.42	2.09, 1.94	γCH$_2$: 2.30 δNH$_2$: 6.97, 7.45
Glu	8.28	4.46	2.12, 1.95	γCH$_2$ 2.45 δOH: 12.3
Gly	8.25	3.90		
His	8.43	4.80	3.31, 3.17	2CH 8.96 4CH 7.38
Hyp (*cis*)		4.74	2.39, 2.21	γCH 4.45 δCH$_2$ 3.87, 3.57
Hyp (*trans*)		4.54	2.23, 2.09	γCH 4.54 δCH$_2$ 4.15, 3.68
Ile	8.06	4.37	1.92	γCH$_2$ 1.62, 1.27 γCH$_3$ 0.85 δCH$_2$ 0.82
Leu	8.15	4.49	1.67, 1.67	γCH$_2$ 1.80 δCH$_3$ 1.09, 1.04
Lys	8.34	4.46	1.89, 1.73	γCH$_2$ 1.51 δCH$_2$ 1.72 εCH$_2$ 2.96
Met	8.29	4.54	2.13. 2.00	γCH$_2$ 2.66, 2.62 δCH$_3$: 2.24
Phe	8.41	4.68	3.24, 2.96	2,6CH 7.38, 3,5CH 7.43 4CH 7.37
Pro (*cis*)		4.68	2.39, 2.21	γCH$_2$ 1.97 δCH$_2$ 3.63, 3.60
Pro (*trans*)		4.50	2.25, 2.07	γCH$_2$ 2.09 δCH$_2$ 3.75, 3.69
Ser	8.37	4.50	3.81, 3.78	
Thr	8.00	4.41	4.21	γCH$_3$ 1.45
Trp	8.37	4.69	3.35, 3.13	2CH 7.35 4H 7.79, 5H 7.24, 6CH 7.16 7H 7.50
Tyr	8.38	4.60	3.12, 2.86	2,6CH 6.84 3,5CH 7.23
Val	8.37	4.36	2.18	γCH$_2$ 1.06, 1.02

Table 5 | ^{13}C random coil chemical shifts for peptides of sequence GGXAGG measured in dimethyl sulfoxide.

X residue	C^{α}	C^{β}	Others
Ala	51.4	20.9	
Arg	55.2	32.1	γCH$_2$ 28.0 δCH$_2$ 43.6 εC 159.8
Asn	52.8	40.2	
Asp	52.6	39.1	
Cys (red)	58.1	29.3	
Gln	55.5	29.6	γC: 33.4
Glu	53.9	30.2	γCH$_2$ 32.1
Gly	45.1		
His	54.5	30.1	1C 132.3 2CH 136.4 4CH 120.2
Hyp (*cis*)	60.8	40.7	γCH 70.1 δCH$_2$ 58.4
Hyp (*trans*)	62.0	43.3	γCH 70.8 δCH$_2$ 57.9
Ile	59.9	39.7	γCH$_2$ 26.6 γCH$_3$ 18.4 δCH$_2$ 14.2
Leu	54.1	43.6	γCH$_2$ 27.2 δCH$_3$ 26.2, 24.6
Lys	55.3	34.3	γCH$_2$ 25.2 δCH$_2$ 29.7 εCH$_2$ 41.9
Met	54.9	34.8	γCH$_2$ 32.6 δCH$_3$ 17.7
Phe	57.1	40.4	1C 140.9 2,6CH 132.3 3,5CH 131.2 4CH 129.4
Pro (*cis*)	61.7	34.8	γCH$_2$ 25.2 δCH$_2$ 49.9
Pro (*trans*)	62.9	32.2	γCH$_2$ 27.4 δCH$_2$ 49.1
Ser	58.2	64.7	
Thr	61.3	69.6	γCH$_3$ 22.8
Trp	58.4	30.6	2CH 126.9 3C 113.0 4CH 121.5 5CH 124.0 6CH 121.3 7CH 114.4 8C 139.2 9C 130.4
Tyr	57.5	39.6	1C 130.9 3,5CH 118.0 2,6CH 133.2 4C 158.9
Val	60.7	33.5	γCH$_3$ 21.2, 20.7

Table 6 | ^{1}H random coil chemical shifts for peptides of sequence GGXAGG measured in methanol:chloroform:water (4:4:1 by volume).

X residue	NH	H^{α}	H^{β}	Others
Arg	8.30	4.3412	1.89, 1.76	γCH$_2$ 1.51, 1.51 δCH$_2$ 3.1
Gly	8.26	3.86		
His	8.46	4.70	3.31, 3.16	2CH 8.56 4CH 7.30
Hyp (*trans*)		4.50	2.34, 2.05	γCH 3.33 δCH$_2$ 3.80, 3.56
Leu	8.13	4.30	1.64, 1.65	γCH$_2$ 1.68 δCH$_3$ 0.96, 0.91
Phe	8.13	4.5579	3.19, 3.01	2,6CH 7.26 3,5CH 7.30 4CH 7.23
Pro (*trans*)		4.386	2.27, 2.01	γCH$_2$ 2.04 δCH$_2$ 3.73, 3.59
Thr	8.24	4.3228	4.27	γCH$_3$ 1.27
Tyr	8.13	4.4892	3.06, 2.93	2,6CH 6.76 3,5CH 7.09
Val	8.24	4.10	2.14	γCH$_2$ 0.98, 0.97

Table 7 |^{13}C random coil chemical shifts for peptides of sequence GGXAGG measured in methanol:chloroform:water (4:4:1 by volume)

X residue	C^{α}	C^{β}	Others
Arg	53.6	30.6	γCH_2 26.4 δCH_2 42.0 εC 158.2
Gly	45.0		
His	54.6	29.0	1C 131.3 2CH 135.9 4CH 120.2
Hyp (*trans*)	62.2	39.9	γCH 72.0 δCH_2 57.1
Leu	52.6	42.1	γCH2 25.6 δCH_3 24.7, 23.0
Phe	57.9	39.3	1C 138.9 2,6CH 131.4 3,5CH 130.8 4CH 129.2
Pro (*trans*)	63.5	31.7	γCH_2 24.2 δCH_2 49.2
Thr	59.8	68.1	γCH_3 21.2
Tyr	56.0	38.1	1C 129.4 3,5CH 116.5 2,6CH 131.7 4C 157.3
Val	57.3	32.4	γCH_3 20.1

2.3.3. Statistical Analysis Of NMR Structure And Chemical Shift Datasets

The accuracy of $\Delta\delta$-based secondary structure prediction, derived using RCCS in aqueous vs. DMSO and trisolvent solutions, in both AQ and TM proteins datasets was assessed using the python program CS-CHEMeleon. A $\Delta\delta$ based prediction was assigned either using a single nucleus type or as the consensus (*i.e.,* agreement by at least 2) of the H^{α}, C^{α}, and C^{β} $\Delta\delta$ values considered relative to the threshold in question. Since RCCS differ with solvent (Figure 11), and since both DMSO and the trisolvent system are established membrane-mimetics, it was expected that the secondary structure of proteins in the TM dataset would have a higher agreement to the PBD structure when $\Delta\delta$ was calculated with RCCS in DMSO or the trisolvent system vs. in aqueous conditions (and *vice versa* for the AQ dataset). Using the original thresholds for defining helix, coil or sheet defined by Wishart *et al.* (Wishart et al. 1995a), the most obvious observation is the incredible accuracy of $\Delta\delta$ in detecting helices compared to sheets and coils. This was true for both the AQ and TM protein datasets and with all five random coil tables (Table 8 and Figure 12).

In an attempt to increase predictive accuracy of $\Delta\delta$ for non-helical structure, we performed iterative optimization of the thresholds for each RCCS table and each nucleus (Table 9; Figure 13). Threshold optimization was performed on the AQ dataset by comparing the number of correctly to incorrectly assessed amino acids over a range of thresholds for each nucleus. The total number of "correct−incorrect" was normalized over the total number of predicted amino acids and the highest value was deemed the optimized value. These optimized thresholds are very different, in some cases, from the values proposed by Wishart *et al.* (Wishart et al. 1995a) and they vary by nucleus and RCCS dataset. This suggests that the previous thresholds may be too general. As expected, the trend of chemical shift shielding and deshielding relative to secondary structure is the same regardless of the actual threshold value on the nucleus type: H^{α} and C^{β} experience shielding in helices and deshielding in sheets, while the opposite trend is true for C^{α}. Cursory examination of Table 8 implies overall higher accuracy of $\Delta\delta$ for predicting secondary structure in TM proteins (~87-88% accuracy) vs. AQ (~62-64% accuracy) proteins. However, the reason for this is actually the predominance of helical structure in the H^{α} containing portion of the TM dataset (Table 10), rather than an inherently better ability to predict TM protein structure from chemical shifts.

Addition of localized averaging for both the PDB file based dihedral angle test and $\Delta\delta$ analysis with both the optimized and original thresholds increased the overall accuracy of secondary structure prediction. For helix and sheet regions, accuracy generally increased modestly (~2-10%) for both the AQ and TM datasets, but by 20-40% for coils. This implies that structural averaging is important for use of $\Delta\delta$ values, particularly in identification of regions lacking defined secondary structure where individual residues may have characteristics of a secondary structure but where the segment as a whole is a coil. Only subtle differences in accuracy were observed between the original thresholds and our iteratively determined thresholds. This could be attributed to the fact that $\Delta\delta$ values are rather large in comparison to the thresholds for most instances of secondary structuring. Clearly, structural thresholds would be significantly more important for small values of $\Delta\delta$, since these would lie closest to the boundary for prediction of coil vs. structured. In these cases, the structural thresholds make a drastic difference. The predictions of sheets in both data sets were relatively poor even when using structural averaging. In comparison to helical structure, sheets are able to assume a much greater variety of backbone dihedral angles (Lovell et al. 2003), which may be a major factor in the relative inaccuracy in prediction of sheets since this should lead to increased variability in chemical shift perturbation from residue to residue. Furthermore, consideration and preferential weighting of different nuclei, such as those identified by Wang and Jardetzky (Wang and Jardetzky 2002b), may improve differentiation between sheet and coil. The addition of the capability to allow differential use of various chemical shift types to distinguish helix, coil and sheet to CS-CHEMeleon is likely in future iterations, but was not the focus of the present work.

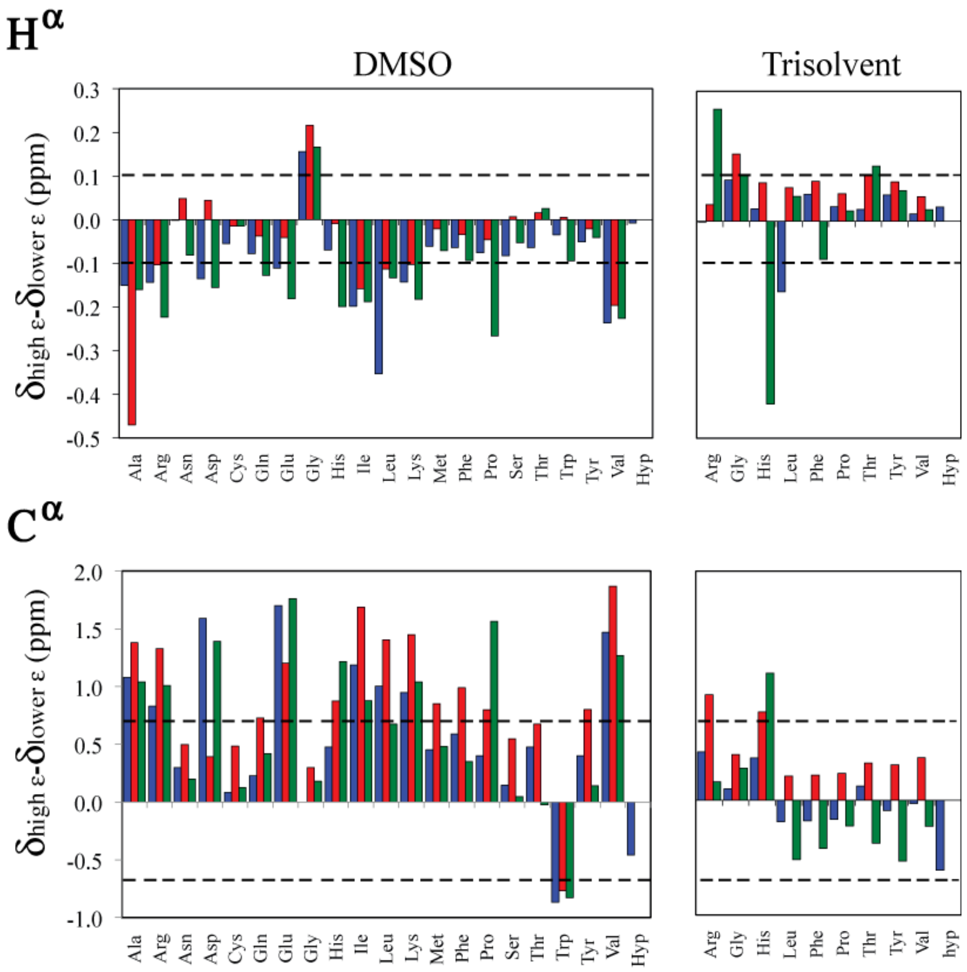


Figure 11 | Chemical shift differences between random coil H^{α} and C^{α} nuclei determined in high (aqueous) and intermediate (DMSO or trisolvent) dielectric environments. Aqueous random coil chemical shifts are from Wishart *et al.* (Wishart et al. 1995a) (blue), Schwarzinger *et al.* (Schwarzinger et al. 2000) (red), and Wang and Jardetzky (Wang and Jardetzky 2002b)(2002) (green) except for the modified amino acid 4-hydroxyproline (Hyp), which is tabulated herein in aqueous conditions duplicating those of Wishart *et al* (Wishart et al. 1995a).

Interestingly, there was no major improvement in accuracy when using RCCS acquired in different conditions when predicting protein secondary structure in the TM dataset vs. the AQ dataset. In other words, there is no obvious correlation between the solvent environment used for RCCS determination and protein secondary structure. The RCCS providing the best overall predictive accuracy were those derived in DMSO and those determined in 8M aqueous urea by Schwarzinger *et al*. (Schwarzinger et al. 2000) (Table 8), while the others fell close behind. A study by Mielke and Krishnan (Mielke and Krishnan 2004), evaluating only non-membrane proteins, came to the same conclusion about the applicability of the chemical shifts of Schwarzinger *et al.* (Schwarzinger et al. 2000) vs. other aqueous RCCS tables.

Although it would be most logical to correlate RCCS environment to the environment of the studied protein (*i.e.,* DMSO or trisolvent with TM proteins and the table of Schwarzinger *et al.* (Schwarzinger et al. 2000) for AQ proteins), our statistical analysis suggests that this is not actually the case. It is possible that RCCS derived in solvents with much lower ε or with complete lack of H-bonding capability, given an appropriately engineered peptide series, would provide significant improvement in secondary structure prediction for membrane proteins or the hydrophobic core region of globular proteins. However, the apparent insensitivity of Δδ-based prediction to fairly dramatic changes in RCCS does not provide a strong incentive to perform these studies. Furthermore, helical regions are already very well predicted, and there may be no major improvement in ability to predict sheet or coil regions, even with a set of chemical shifts derived in such an environment. Based upon our findings, the optimal method for any type of protein being studied in an aqueous or non-aqueous environment is to perform a comparison of the Δδ values derived using both the DMSO-based table and the table of Schwarzinger *et al.* (Schwarzinger et al. 2000) in order to determine a consensus for secondary structure prediction. The comparison should also be made with more than one nucleus if possible. This type of comparative analysis is readily possible using CS-CHEMeleon.

2.4. SUMMARY

I have determined the ^{1}H and ^{13}C chemical shifts of 21 hexapeptides with sequence Ac-GGXAGG-NH$_2$, where X is one of either the 20 naturally occurring amino acids or the modified amino acid 4-hydroxyproline, demonstrating significant differences in RCCS between aqueous environments and intermediate ε environments. Structural studies of TM proteins that use Δδ values for fast assessment of secondary structure have used RCCS measured in aqueous environments, but the predictive accuracy of Δδ utilization has never been evaluated over a range of solvent environments. In this paper, I provide evidence that, although Δδ values themselves are affected, the solvent in which the RCCS were measured does not significantly affect the prediction of secondary structure. Rather, the type of secondary structure is a major factor in the agreement between Δδ-based prediction and experimental secondary structures. For a thorough assessment, all three major nuclei (H^{α}, C^{α}, C^{β}) should be considered as well as the type of secondary structure being evaluated. Although Δδ use is well suited to an overall estimate of protein structure, the evidence presented herein of a bias towards helices with all RCCS datasets provides a strong incentive to employ alternative restraints during structure calculation and to ensure that Δδ-based restraints are not overrepresented in the energy expression being used for structure calculation. Choice of the Δδ threshold for helix vs. coil or sheet also may slightly improve overall secondary structure prediction accuracy. Furthermore, optimal accuracy in secondary structure prediction is probably obtainable by comparison of Δδ obtained with the table presented herein in DMSO and the table determined in aqueous conditions by Schwarzinger *et al.* (Schwarzinger et al. 2000). The web-based software CS-CHEMeleon, introduced herein, provides a rapid and versatile method to allow such a comparative analysis.

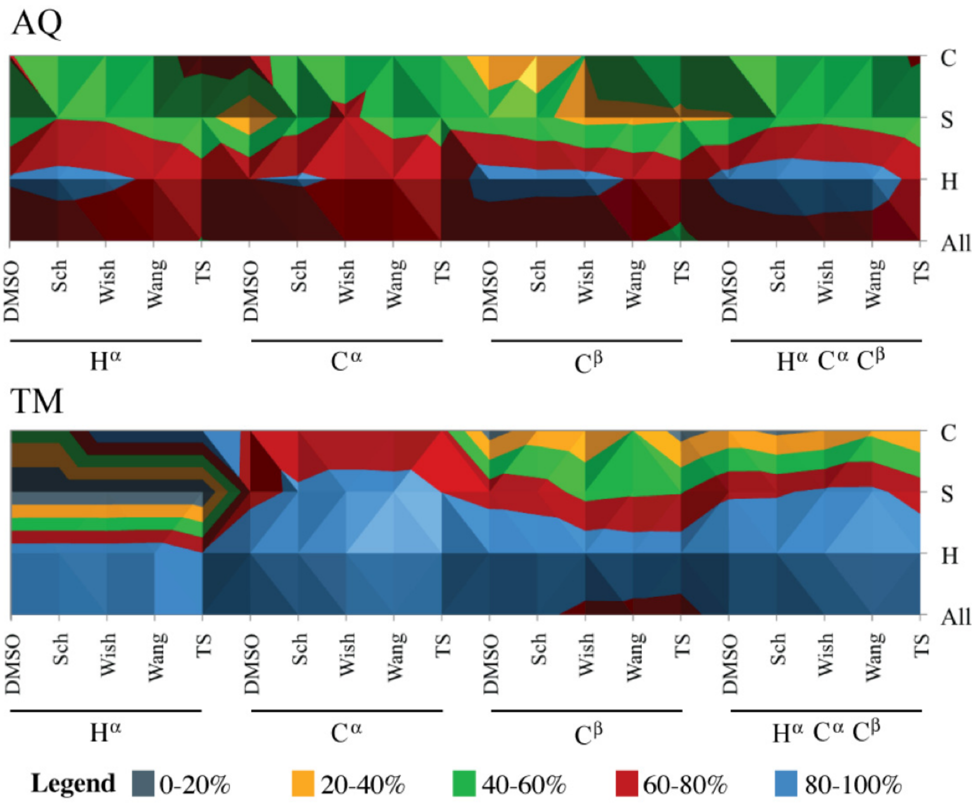


Figure 12 | Relative predictive accuracy of secondary structure by secondary chemical shift for the indicated nucleus type for the non-membrane (AQ) and membrane (TM) protein datasets (dataset details in Table 10). The percentages of correctly predicted structure per amino acid using iteratively optimized secondary chemical shift thresholds (Table 9) were calculated as a function of nucleus (H^{α}, C^{α}, C^{β}, and consensus ≥ 2 of the three atoms), random coil chemical shift, and secondary structure type (helix (H), sheet (S), and coil (C)). Localized averaging over 5 residues was applied for both the secondary structure derived by ϕ and ψ dihedrals and for secondary chemical shift predictions.

Table 8 | Agreement between secondary chemical shifts (requiring consensus of 2 of 3 of H^{α}, C^{α}, and C^{β}) based prediction of helix, sheet and coiled secondary structure in comparison to the consensus structure determined by average ϕ and ψ dihedral angles at that position in the ensemble of NMR structures. Secondary chemical shift based predictions are compared using the "original" threshold values and the optimized values determined herein (Table 9) and using localized averaging ("Local avg.", the 5-residue consensus both of secondary chemical shift predictions and dihedral angle based structure) vs. consideration of secondary chemical shifts and dihedral angles at each residue in isolation ("No avg."). Results are presented for datasets consisting of 109 non-membrane proteins and 19 membrane proteins, split into overall agreement (All) and agreement by secondary structure type of helix (H), sheet (S) and coil (C) The total number of amino acids compared are listed in Table 10.

		Non-membrane Proteins				**Membrane Proteins**			
		No local avg.		Local avg.		No local avg.		Local avg.	
Solvent	Type of 2° structure	Wishart	Iteration	Wishart	Iteration	Wishart	Iteration	Wishart	Iteration
DMSO	All	62.8	65.4	68.3	67.8	85.7	86.3	86.3	87.5
	H	86.0	84.2	92.7	87.6	97.3	95.7	98.6	98.3
	S	45.7	47.2	41.6	38.1	75.0*	80.9*	73.3*	77.4
	C	27.8	33.0	44.3	58.8	21.7*	8.7*	30.0*	25.0*
Schwarzinger	All	63.4	68.0	69.7	71.0	87.5	87.9	91.2	88.1
	H	75.6	84.7	81.0	92.1	93.0	98.7	96.6	99.7
	S	57.6	56.1	57.2	54.1	87.2	80.3	87.2	77.9
	C	36.2	28.4	58.6	42.0	17.4*	8.7*	50.0*	15.0*
Wishart	All	63.3	67.7	69.5	71.9	87.5	86.1	92.4	89.8
	H	72.8	83.2	77.9	89.7	93.0	96.3	96.6	99.3
	S	57.6	56.1	57.2	54.1	86.2	81.4	88.2	83.6
	C	40.8	31.1	61.3	49.4	4.3*	4.3*	40.0*	20.0*
Wang	All	62.3	66.9	67.6	71.4	87.5	86.1	92.4	89.8
	H	69.9	81.8	73.3	88.8	92.3	94.7	95.9	97.3
	S	61.1	54.6	62.4	52.0	88.8	81.9	91.8	84.1
	C	38.2	36.2	60.3	54.8	13.0*	8.7*	45.0*	35.0*
Trisolvent	All	63.7	66.0	61.6	62.7	87.9	76.7	87.1	83.0
	H	71.1	78.2	66.6	74.6	93.9	89.7	95.2	95.2
	S	64.2	59.5	56.8	45.9	86.0	68.6	78.0	71.6
	C	37.7	36.8	57.2	61.9	20.0*	13.5*	66.7*	22.2*

* Sample size with less than 30 amino acids

Table 9 | Iteratively optimized thresholds for secondary chemical shift based prediction of helix, sheet, and coiled structures for the RCCS presented herein (Table 4Table 5Table 6Table 7) and for three published random coil sets: Scharzinger *et al.* (Schwarzinger et al. 2000) (Schwar.), Wishart *et al.* (Wishart et al. 1995a) (Wishart), Wang and Jardetzky (Wang and Jardetzky 2002b) (Wang). Coils are defined as secondary chemical shifts between the helix and sheet threshold values. The 'original' structural thresholds were those defined by Wishart *et al.* (Wishart et al. 1995a) and included for comparison. A set of 109 proteins with published structure and chemical shifts (the AQ dataset in Table 10) were used for optimization. Full optimization results are presented at the end of Chapter 2.

	Original	DMSO	Schwar.	Wishart	Wang	Trisolvent
H^{α}-Helix	<-0.1	<-0.16	<-0.10	<-0.06	<-0.02	<-0.06
H^{α}-Sheet	>0.1	>0.22	>0.08	>0.16	>0.16	>0.24
C^{α}-Helix	>0.7	>1.30	>0.20	>0.50	>0.60	>0.80
C^{α}-Sheet	<-0.7	<-0.10	<-0.90	<-0.40	<-0.60	<-0.40
C^{β}-Helix	<-0.7	<-0.30	<0.20	<0.20	<0.20	<0.20
C^{β}-sheet	>0.7	>0.20	>1.00	>1.30	>1.60	>1.20

Table 10 | Number of residues in each dataset categorized by chemical shift availability per residue and classified by secondary structure (helix (H), sheet (S), coil (C)), with secondary structure determined using the consensus structure according to ϕ and ψ dihedral angles over a 5-residue local average around the residue in question.

	TM				**AQ**			
H^{α}	Total	H	S	C	Total	H	S	C
# of files	25				130			
#of Proline	3	3	0	0	418	171	198	49
# of glycines	33	33	0	0	698	242	121	335
# total amino acids	343	335	2	6	10014	5176	3308	1530
C^{α}								
# of files	25				99			
#of Proline	19	6	10	3	402	165	189	48
# of glycines	80	37	39	4	677	228	121	328
# total amino acids	606	338	241	27	9526	4954	3084	1488
C^{β}								
# of files	25				130			
#of Proline	17	5	9	3	403	166	188	49
# of glycines	0	0	0	0	0	0	0	0
# total amino acids	500	274	203	23	8777	4702	2939	1136
≥ 2 of H^{α}, C^{α}, C^{β}								
# of files	19				109			
#of Proline	19	6	10	3	294	104	151	39
# of glycines	80	37	39	4	453	130	94	229
# total amino acids	500	274	203	23	6127	2855	2269	1003

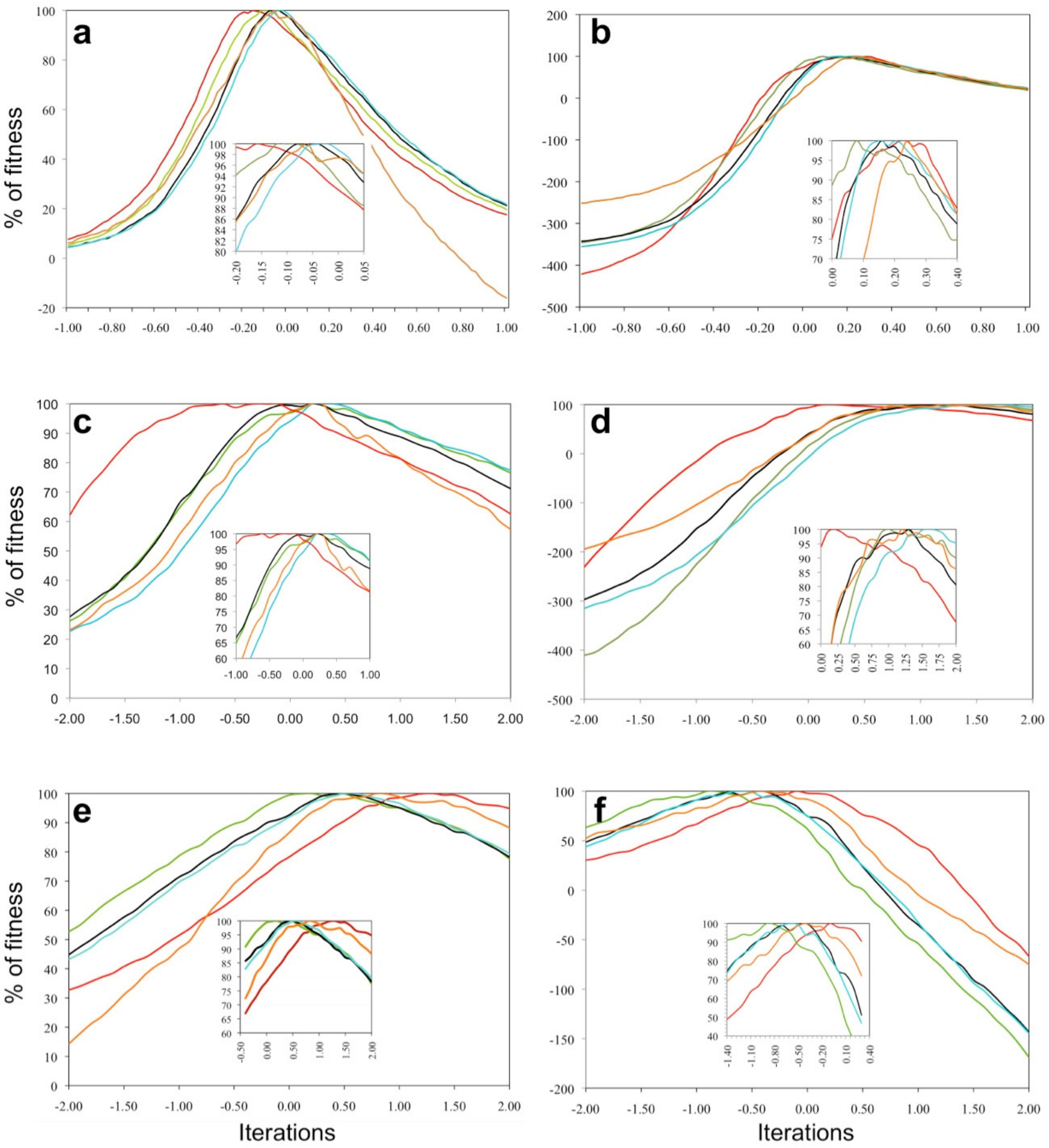


Figure 13 | Graphical representations of the normalized iterative threshold data for each nuclei, secondary structure, and random chemical shift table. The number of correctly predicted amino acids was subtracted from the number of incorrectly assessed amino acids. Each iteration was normalized and turned into a percentage. A close-up of the maximum (at 100%) is showed to provide better resolution at that region. a. Helix iteration for H$^{\alpha}$ b. Sheet iteration for H$^{\alpha}$ c. Helix iteration for C$^{\alpha}$ d. Sheet iteration for C$^{\alpha}$ e. Helix iteration for C$^{\beta}$ f Sheet iteration for C$^{\beta}$. Legend: DMSO (red), Schwarzinger *et al*. (Schwarzinger et al. 2000) (green), Wishart *et al.* (Wishart et al. 1995a) (black), Wang and Jardetzky (Wang and Jardetzky 2002b) (blue), trisolvent (orange).

CHAPTER 3. THE STRUCTURE OF W$_1$ SPIDER WRAPPING SILK (INCL. XU ET AL. (2012B))

This chapter focuses on the assignment, structure calculation, and characterization of the 200 amino acids repeat domain of AcSp1 that we called W$_1$ (W for *W*rapping silk and 1 for 1 repeat domain). The basic concepts and algorithms discussed in chapter two will be revisited. 2D and 3D NMR experiments were acquired, processed and assigned. This was followed by simulated annealing calculations to determine a final structural ensemble consistent with the NMR data. In depth characterization of the W$_1$ structure follows.

3.1. INTRODUCTION TO PROTEIN STRUCTURE DETERMINATION

To obtain a high-resolution structure, the great majority of the (or, ideally, all) NMR active nuclei within a protein have to be specifically assigned. Figure 14 outlines a workflow allowing the determination of an NMR structure. Each step will be described in detail.

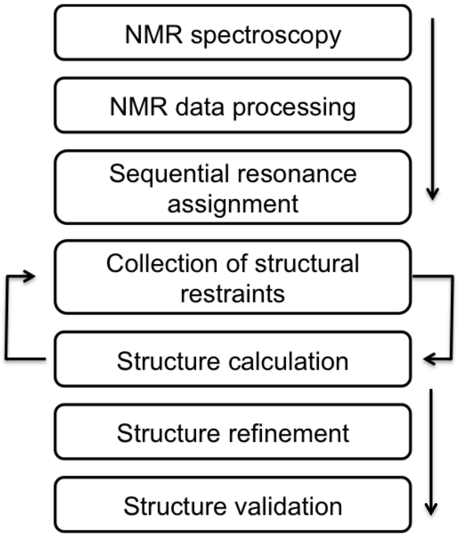


Figure 14 | Outline of a typical procedure for protein structure determination by NMR

3.1.1. Sequential Resonance Assignment With 3D NMR Data

I often describe NMR resonance assignments to new users as a gigantic 3D puzzle. All NMR peaks are discrete, differing from each other by chemical shift variations that allow one to distinguish one atom's resonance frequency from another. As discussed in chapter 2, chemical shifts are highly dependent on their environment. If one is lucky enough, none of the peaks overlap with one another, but most of the time that isn't the case. In the end, all the peaks in the NMR puzzle have to fit together because the NMR spectrometer doesn't lie; each visible peak exists for a reason.

Using high-resolution NMR spectroscopy, the resonance frequency arising from each nucleus in a protein sample can be assigned individually to the corresponding atom expected on the basis of the known covalent protein structure. The fingerprint and root spectrum for the process of triple-resonance biomolecular NMR sequential assignment is the ^{1}H-^{15}N HSQC (Farrow et al. 1994), which is a 2D spectrum with cross-peaks that correlate an amide proton to a backbone or side-chain nitrogen (^{1}H-^{15}N). Note that the term "correlation" implies that nuclear spins are coupled to each other through bonds, an effect called scarlar coupling (or *J*-coupling). Nuclear spins are able to interact with each other through chemical bonds, an effect mediated by the polarization of electrons, and give rise to constants highly exploited in protein NMR for nuclei selection during experiments (Vuister et al. 2002). Since there is only one amide bond per amino acid (with exception of side chains, which have different chemical shifts), the combined pattern of peaks arising from ^{1}H-^{15}N correlations (*J*-coupling = 92 Hz) provides a structural fingerprint for the protein. Changes in chemical shifts indicate a change in environment and/or protein structure.

The HSQC does not provide enough information to carry out sequential assignment of an entire protein; therefore, more experiments are required to correlate the backbone amide nuclei to the other nuclei present in the protein. Typically, it is common to acquire a series of 3D experiments for larger proteins (>50 aa). 3D backbone-walk experiments employ a ^{1}H-^{15}N HSQC as a basis of the experiment (think of the x and y axis of a cube), with the peaks observed in the 2D HSQC spectrum spread over a third dimension (z axis) that separates these on the basis of the chemical shift of a nucleus or nuclei that are not part of the ^{1}H-^{15}N pair.

The first step of the assignment process for 3D data "walk" along the backbone of the protein using the HNCA/HNcoCA and the HNCO/HNcaCO spectral pairs (Kay et al.

1990; Bax and Ikura 1991) via what is called a 'backbone walk' (Figure 15). The experiment names are based upon the nuclear correlations that are employed - "H" and "N" signify the ^{1}H and ^{15}N of the amide bond, while CA and CO refer to the backbone C$^{\alpha}$ and C', respectively. Nuclei denoted in lower case are part of the magnetization transfer pathway, but are not observed; those in upper case represent the chemical shifts observed. When assigning non-trivial backbone data, both pairs of experiment are typically used together, both to reinforce confidence in the assignments and to improve ability to deal with overlap.

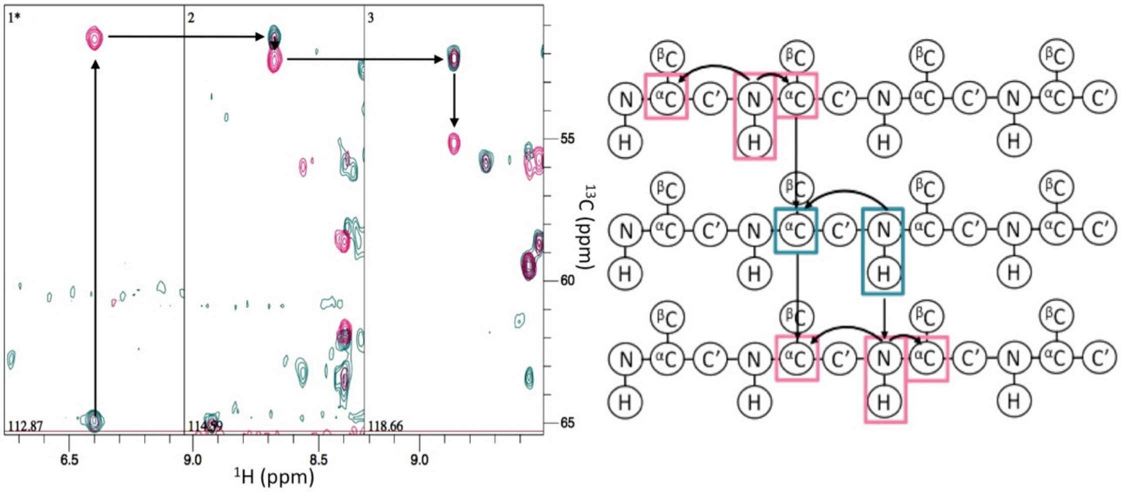


Figure 15 | Illustration of the principle underlying the 3D NMR protein backbone walk. In pink is the HNCA, with 2 cross-peaks per ^{1}H-^{15}N strip plot corresponding to the C$^{\alpha}$ nuclei of residues i-1 and i in teal is the HNcoCA, which only contains a peak for C$^{\alpha}$ i-1. The ^{1}H dimension is at the bottom, the ^{13}C$^{\alpha}$ dimension on the vertical, and the chosen ^{15}N chemical shift at the bottom corner of each window. Combined together, for each amino acid you see a peak for the i amino acid and the i-1 amino acid. You know which one is which by overlaying the HNcoCA spectra on top o the HNCA, which only has one peak for the i-1. In this example, the travel direction is towards the C-terminal. Marking the chemical shifts for ^{13}C$^{\alpha}$ (pink, first strip), you travel through the nitrogen planes to find an HNcoCA peak (teal, 2nd strip) that aligns with the ^{13}C$^{\alpha}$ in strip 1. This 2nd H-N strip is the plane for the i+1 amino acid in the sequence. On that same plane, a pink peak not overlayed by a teal peak should be visible - this will be the ^{13}C$^{\alpha}$ chemical shift for the i+1. That shift is marked, and the process repeated. To the right is a cartoon illustration of how the backbone walk proceeds by using both spectra and what atoms are correlated in the experiment. The first row is the pink HNCA peak seen in strip 1 and the next amino acid can be found by matching the ^{13}C chemical shifts (teal, 2nd row). The 2nd and 3rd row represent the 2 peaks found in strip 2 and not 3 individual protein chains

The HNCO and the HNcoCA contain only one peak per amide of residue *i*, correlated to the named carbon of residue *i*-1. Conversely, the HNCA and HNcaCO have two correlated carbon peaks per ^{1}H-^{15}N pair of residue *i*: one for residue *i* and one for the *i*-1. Figure 2 illustrates the utility of these correlations in the HNCA/HNcoCA pair of experiments for the sequential assignment process. It should be noted that prolines do not have amide protons and hence do not show up as residue *i* in any of these experiments, resulting in “breaks” in the sequential assignment. The proline C^{α} and C’ chemical shifts can be assigned using the i-1 peak correlated to the NH of the residue C-terminal to the proline. The CBCANH experiment (or, in the lower molecular weight range, HNCACB experiment) adds the ability to correlate between the amide N-H and the C^{β} nucleus. Specifically, this particular experiment correlates the backbone amide to both the C^{α} and C^{β} of residues *i* and *i*-1.

The ^{15}N-edited HSQC-total correlation spectroscopy (^{15}N-edited HSQC-TOCSY (Marion et al. 2002)) experiment can be used to assign and identify amino acid types by relating the backbone amide ^{15}N and ^{1}H chemical shifts to the H^{α} and side chain protons. In this experiment, a spin-lock pulse train induced *J*-coupling mediated transfer of magnetization within an amino acid, with transfer back to the amide proton for detection. It should be noted that α-helical structure leads to poor TOCSY transfer from the amide, meaning that this experiment is of limited use in helical regions. Also, breaks in the side chain ^{1}H-^{1}H *J*-coupling network are caused by quaternary carbons and lead to incomplete TOCSY transfer (Figure 16). Carbon and proton resonances that can’t be assigned with a ^{15}N-edited HSQC-TOCSY are typically assigned via ^{13}C-edited HSQC-TOCSY (entirely analogously to the ^{15}N-edited version) and/or HCCH-TOCSY (Bax et al. 1990; Olejniczak et al. 1992) experiments. Magnetization in the HCCH-TOCSY experiment is transferred from the side-chain protons or H^{α} to an attached ^{13}C nucleus. This is followed by an isotropic ^{13}C mixing step and transfer of magnetization back to a covalently coupled proton for detection.

Although not used in this thesis, it should be noted that there are other 3D experiments, such as the H(CCO)TOCSY-NH (Gardner et al. 1996) that can be used for the assignment of protonated methyl groups in an otherwise deuterated protein. The HA(CA)NH and HA(CACO)NH experiments (Reddy and Hosur 2013) can also be used in a pair along with C^{α} chemical shifts determined using, e.g., HNCA/HNcaCO to provide unambiguous assignments of the H^{α} chemical shift, allowing full assignment of H^{α} and C^{α} shifts and providing a starting point for remaining assignments within the side chain using the HCCH-COSY and HCCH-TOCSY experiments. Ultimately, the HNHA and the HNHB can serve the same purpose.

3.1.2. Scalar Couplings

Scalar coupling are not only used for experimental setup in the selection of nuclei required during experiments, but also define bond geometry and provide the possibility of assigning atoms stereo-specifically. For small molecules, the *J*-couplings can be directly extracted from the splitting of the resonance signal. However, overlap of signals in larger biomolecules complicates the measurement of *J*-coupling constants. The 3D HNHA (Vuister and Bax 1993; Vuister et al. 2002) and HNHB (Archer et al. 1991; Düx et al. 1997; Barnwal et al. 2007) experiments allow determination of coupling constants between the indicated pairs of protons. This, in turn, allows dihedral angle constraints to be determined (ϕ for HNHA and χ_1 for HNHB). The signal intensity is compared between a cross-peak (S_{cross}) and its corresponding diagonal (S_{diag}):

$$\frac{S_{cross}}{S_{diag}} = -\tan^2(2\pi J_{HH}\zeta) \tag{3.1}$$

where J_{HH} is the *J*-coupling term, and ζ is the transfer period. The three bond *J*-couplings (^3J) obtained are then related to the dihedral angles θ, through the Karplus relationship (Karplus 1963):

$$^3J = A\cos^2\theta + B\cos\theta + C \tag{3.2}$$

The HNHA (Vuister and Bax 1993) will result in a coupling value of 5-10 Hz and the HNHB (Archer et al. 1991) will have smaller coupling values of 1-3 Hz. In the HNHB, the H^{β} have to be stereospecifically differentiated.

3.1.3. The Nuclear Overhauser Effect (NOE) And Associated Experiments

The NOE is the most important experimental parameter for the determination of a biomolecular structure in solution (Wüthrich 1986) and *the* most tedious assignment task. The assignment of NOESY cross-peaks requires assignment of the protein's chemical shifts. The NOE is a transfer of nuclear spin magnetization through space from one spin to another via dipole-dipole cross-relaxation (Figure 16). In other words, when two protons are sufficiently close in space, their magnetic dipoles interact to give rise to a peak with intensity *I* that is inversely proportional to the sixth power of the distance r between the pair of spins in accordance to the relation $I = f(\tau_c)<r^{-6}>$, in which $f(\tau_c)$ is a function of the rotational correlation time that varies by molecular size, solution temperature, and solvent conditions (Teng 2005). A cross peak arises when $r \leq$ 6 Å. For a quantitative NOE to occur, the mixing time must be properly determined. An overly short mixing time might lead to small NOEs that contain important tertiary information being unobservable. Overly long mixing times, on the other hand, will allow for spin diffusion, which means that NOEs will be indirectly generated at inter-proton distances of > 6 Å.

In one 3D version of the NOESY experiments, the peaks in the 2D ^{1}H-^{15}N HSQC are further separated in a 3rd dimension as a ^{1}H-^{1}H NOESY, resulting in a spectrum of the distances between the amide proton and all other protons in the protein. The ^{13}C based 3D NOESY-HSQC(Marion et al. 1989; Marion et al. 2002) follows the same principle, but starts with a ^{13}C-^{1}H HSQC, which contains all side chain and backbone ^{13}C-^{1}H, separated in the 3rd dimension by the ^{1}H-^{1}H NOESY.

Another type of NOESY spectra is the 3D ^{15}N-HSQC-NOESY-HSQC, more commonly known as the CNH-NOESY (referred to as the ChNH herein) or NCH-NOESY depending on the magnetization transfer order (Diercks et al. 1999). The chemical shift of the carbon is recorded, and the magnetization is then passed on to the attached proton and then to a nearby amide proton by dipolar coupling, creating the NOE. Generally, the ^{1}H-^{1}H diagonal present in the other mentioned 2D and 3D spectra, is eliminated in the ChNH by an orthogonal heteronuclear filter, which substantially reduces overlap and improves spectral quality. Another advantage of the ChNH is the possibility to obtain most side chain carbon assignments without the need for a 3D HCCH-TOCSY while still obtaining ^{1}H-^{1}H distances from the amide proton to other nearby protons. Alternatively, assignment of the intraresidue side-chain carbons is facilitated by the parallel use of both the ChNH and the HCCH-TOCSY (Bax et al. 1990; Olejniczak et al. 1992).

If of interest, more information on assignments is very nicely presented in the CcpNMR wiki tutorial website [http://www.ccpn.ac.uk/software/tutorials] and the descriptions of the experiments I mentioned (and more) can be found on the CcpNMR Protein NMR practical guide [http://www.protein-nmr.org.uk/ccpnmr-analysis/].

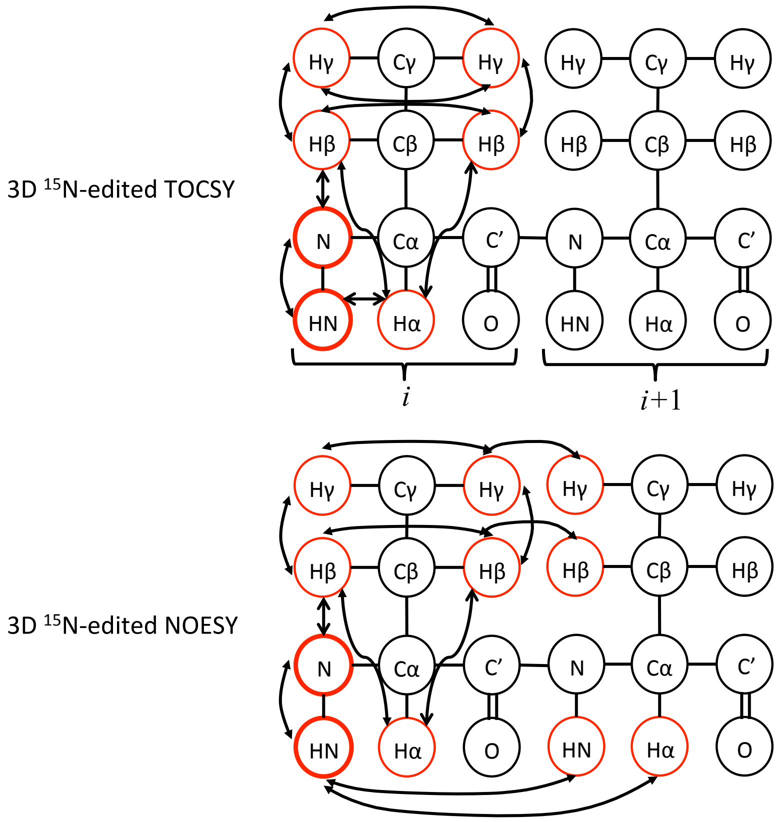


Figure 16 | Transfer of magnetization of the 3D ^{15}N-edited TOCSY-HSQC compared to the ^{15}N-edited NOESY-HSQC. The ^{1}H-^{15}N HSQC correlations are bolded in red. In the ^{15}N-edited TOCSY-HSQC , the magnetization travels from the H-N bond (in bold) to the other protons within the amino acid only. In the ^{15}N-edited NOESY-HSQC, the magnetization also starts on the H-N bond (in bold) but travels through space instead of through bonds, giving rise to signals from the *i*+1 protons in this figure. In sequential assignment, the peaks from the ^{15}N-edited TOCSY-HSQC spectra overlap with the ^{15}N-edited NOESY-HSQC with the NOESY containing more peaks.

3.1.4. Types Of Restraints Use In Structure Calculations

3.1.4.1. Distance Restraints

Distance restraints are hugely important, ultimately determining the overall 3D fold of the protein. These are generated between pairs of protons from the corresponding assigned NOE cross-peak intensity or volume. Based on peak intensity, NOE cross-peaks are separated into bins with different upper and lower distance boundaries. Medium (*i* ± 2-4) and long-range (*i* ≥ 5) distances are often the least intense cross-peaks, but are by far the most crucial because they provide important secondary and tertiary structural information.

Ideally, an observed NOESY cross peak will arise due to only one ^{1}H-^{1}H distance pair. Unfortunately, such ideal cases are often infrequent in larger proteins, especially in W_1 with its 41 serines, 31 glycines and 29 alanines! When an NOE peak can correspond to more than one assignment due to overlap of assigned resonance frequencies, we produce an ambiguous distance restraint (ADR) (Nilges et al. 1997). An ADR peak intensity can be considered as a cumulative sum of possible inter-nuclei distances:

$$\bar{d} = \left(\sum_{k=1}^{N_k} d_k^{-6} \right)^{-\frac{1}{6}} \tag{3.3}$$

where $\bar{d}$ is the effective distance, N_{k} is the number of assignment possibilities, and d_{k} is the interatomic distance between two protons corresponding to the k^{th} contributions (Nilges et al. 1997). Some, or all, of the ambiguity can often be dealt with during iterative structure calculations since some ambiguously assigned proton pairs will frequently be inconsistent with the other experimental data.

3.1.4.2. Dihedral Angles

Restraints for protein backbone ϕ and ψ dihedral angles and, in some instances, side-chain χ dihedral angles can be predicted on the basis of assigned chemical shifts *in silico* by freely available software. DANGLE (Cheung et al. 2010), which is integrated directly into the assignment program CcpNMR Analysis (Vranken et al. 2005), uses Bayesian inference to estimate backbone ϕ and ψ dihedral angles for each residue based upon chemical shifts and amino acid conformational preference. TALOS+ (Shen et al. 2009a), a very frequently employed program to predict dihedral angles, is a hybrid algorithm that combines empirical data from published structures and the chemical shifts from 6 nuclei (H^{N}, HA, CA, CB, C’, N) to predict the ϕ and ψ dihedral angles.

As introduced above, experimental dihedral angle restraints can also be obtained from *J*-couplings extracted from the 3D HNHA (ϕ) (Vuister and Bax 1993) and HNHB (χ) (Archer et al. 1991). These restraints are usually employed in structure calculation protocols directly as *J*-coupling restraints, but there is always the possibility to convert them into dihedrals if need be.

3.1.4.3. Other Restraints

Residual dipolar couplings provide information on the relative orientations of bond vectors with respect to the direction of the external magnetic field B_o. Unlike NOEs, these restraints may be long- or short-range, making them extremely useful in the determination or validation of a 3D fold. Much more detailed information is available in Appendix D.

Radius of gyration (R_g) restraints (Ortega et al. 2011) may also be employed. These can be obtained through scattering experiments, hydrodynamics measurements, or predicted *in silico* by software such as HYROPRO (García de la Torre et al. 2000) if the structure is known.

H-bonds restraints can also be included when evidence suggests that H-bonds are located at specific site (Wüthrich and Wagner 1979). These are also often incorporated in the absence of such evidence on the basis of evidence for a particular secondary structure.

3.1.5. Simulated Annealing

Simulated annealing is a term for an algorithm that uses a probabilistic method for finding a global minimum of a function that may possess several local minima, such the folding of an extended polypeptide chain into a 3D structure (Granville et al. 1994). These computer algorithms use both empirical and experimental data to fold a structure in three dimensional space (Bassolino-Klimas et al. 1996). Frequently employed simulated annealing software includes Crystallography & NMR System (CNS) (Brunger et al. 1998), DYANA/CYANA (Güntert 2004) (CYANA, version 1.0, www.guentert.com), Ambiguous Restraint for Iterative Assignment (ARIA2) (Rieping et al. 2007), and Xplor-NIH (Schwieters et al. 2006). There are several good reviews that compare and contrast on the methods available for automated protein structure calculation (Williamson and Craven 2009; Guerry and Herrmann 2011).

Embedded in any simulated annealing-based structure calculation programs is an empirical energy function (force fields) specifically defined for amino acids and biological molecules. The energy function (*E*) is defined as

$$E_{\text{total}} = E_{\text{empirical}} + E_{\text{effective}} \tag{3.4}$$

where $E_{\text{empirical}}$ involves the energy derived from bond lengths, angles, H-bonds, or VDW contact, etc (Teng 2005) and $E_{\text{effective}}$ represents the experimental user-defined restraints. The $E_{\text{empirical}}$ relies upon parameter and topology files that specifically define atom charges, masses, energy constants and other standard values. Typically, initial randomized structures are generated and energy minimized. Then the system is heated to high temperature (~3500 K), enough to cross local energy barriers, and slowly cooled at a user-defined rate in a series of steps (~9000 - 72000 steps for small to large proteins) (Fossi et al. 2005). During cooling, the weights of the experimental restraints are ramped up to allow for a balance between satisfaction of restraints from assignments, covalent character and folding. This process is iterated until all restraints and geometric considerations are satisfied (Figure 17). In NMR simulated annealing calculations, an ensemble of structures is typically calculated at each iterative step in the refinement. Following refinement, a subset of the final ensemble – the converged structural ensemble – is retained as being representative either on the basis of lowest energy or minimal restraint violations.

Xplor-NIH (Schwieters et al. 2006) is a rigorously tested and validated structure calculation program, based on the earlier X-PLOR and its older comrade CNS (Brunger et al. 1998). The initial X-PLOR package (Brünger et al. 1987) focused on 3D structure determination using either crystallographic diffraction or NMR data and it was the first program to integrate X-ray crystallographic data with the molecular dynamics and minimization program CHARMM (Brooks et al. 1983) for refinement. Unlike the newer structure calculation programs mentioned in the next paragraph, restraint filtering during refinement in Xplor-NIH is manual, requiring much human input at each step. This allows for greater control and understanding of the refinement process. Although not an optimal program for the treatment of large numbers of ambiguous distance restraints, it is often used for final refinement of structures in explicit solvents.

Larger proteins (>~50 residues) give rise to significantly more unique and ambiguous NOE restraints than small peptides. The implementation of automated NOE assignment, a movement partly initiated by Peter Güntert, with the addition of ADR treatment by Michael Nilges (Nilges et al. 1997) has revolutionized protein NOE assignment and accelerated the protein structure calculations process by automating

NOE peak assignment while also providing greater accuracy and precision in the ADR assignments(Guerry and Herrmann 2011). Many automated NOE assignment programs have been implemented: ASNO (Güntert et al. 1993), SANE (Duggan et al. 2001), NOAH (Mumenthaler and Braun 1995), ARIA (Nilges et al. 1997; Linge et al. 2001; Linge et al. 2004), the CANDID algorithm (Herrmann et al. 2002) used in CYANA (Güntert 2004), KNOWNOE (Gronwald et al. 2002), and AutoStructure (Greenfield et al. 2001). Nevertheless, input from an experienced NMR spectroscopist is still very valuable for making intuitive decisions in the face of NMR spectral data imperfections (artefacts, solvent signals, high noise level, etc) which is the case for most data collected on proteins, or filtering automated assignment errors arising from spectral data defects.

ARIA2 (Rieping et al. 2007) was selected for use herein for structure calculations because of 1) its direct incorporation into CcpNMR Analysis (Vranken et al. 2005), the software used for my chemical shift assignments, which facilitated the transition to structure calculation and validation; 2) because of the advantages of ADR filtering (Nilges et al. 1997); and, 3) because of the option to use network anchoring (described below in section *3.1.5.1*) and 4) because it's free, unlike CYANA.

Starting from peak lists and chemical shifts, ARIA2 (Rieping et al. 2007) proceeds in iterative cycles of NOE assignments, NOE disambiguation, and structure calculation with a final refinement in water if one so desires. Additional restraint sets do not require scripting - restraints such as dihedral angle constraints, HNHA (Vuister and Bax 1993) *J*-couplings, RDCs, hydrogen bonds, and disulfide bridges are instead easily incorporated using a GUI. In each cycle, ARIA2 calibrates and assigns NOESY spectra, merges the constraint lists for different spectra, and calculates a user-defined number of structures. In iteration 0, ARIA2 produces a template structure. At the end of each iteration, the best structures are retained (usually sorted by lowest energy) and used as a template for the next iteration. At each iteration, the parameters that define violations are tightened after each cycle to finally generate a converged structure in very little computational time (Figure 17). The structures and peak lists can be imported into CcpNMR Analysis (Vranken et al. 2005) for visualization and inspection of violated NOE assignments.

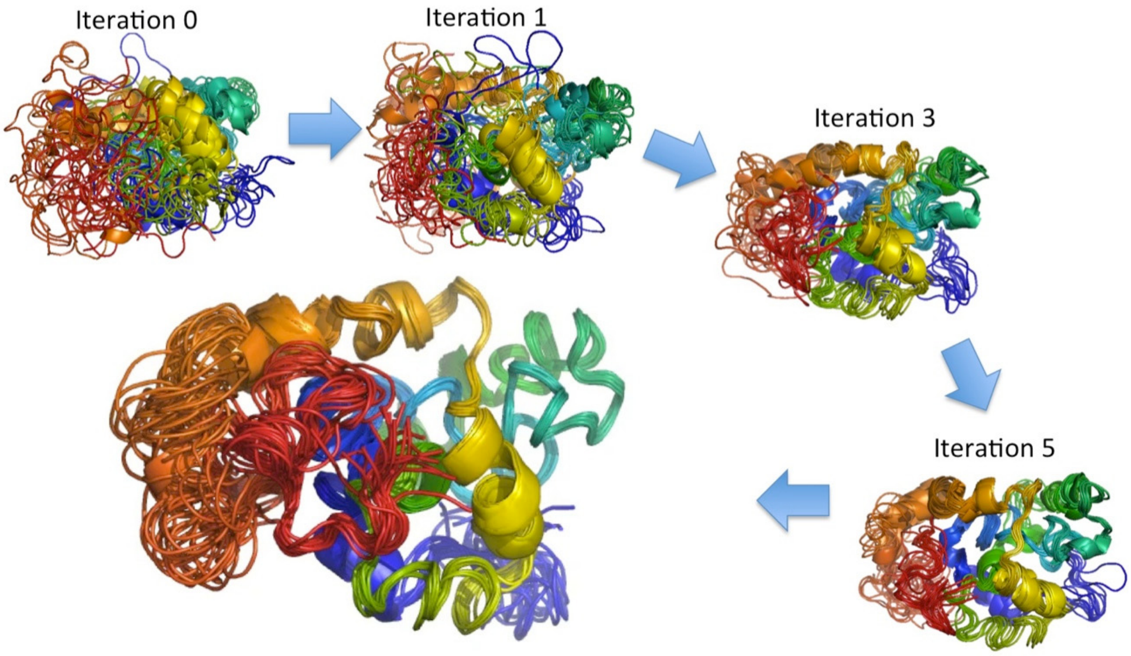


Figure 17 | An example of structural ensembles produced during selected iterations using ARIA2.1. Following each iteration, restraints and ambiguity are both refined and removed, as appropriate. SA is then resumed and a new set of structures will be produced and evaluated, with repetition of the cycle until agreement with experimental restraints is deemed satisfactory.

3.1.5.1. Network Anchoring

Network anchoring (NA) is an algorithm that is integrated into both ARIA2 (Rieping et al. 2007) and CYANA (Güntert 2004) for assistance in filtering correct from erroneous NOEs. Network-anchoring evaluates the self-consistency of NOE assignments independent of any previous knowledge of the protein 3D structure (Herrmann et al. 2002; Güntert 2004). A network can be formed to determine protein 3D structures if a set of sufficiently dense, self-consistent restraints are present during structure calculations. Network anchoring has proven efficient for finding well defined, essentially correct, structures during the first cycle of the structure calculations and is robust for automatic NOE assignments as well as filtering correct ADR assignments from more erroneous possibilities (Zech et al. 2005; Guerry and Herrmann 2011). More importantly, network anchoring serves to detect erroneous, 'lonely' restraints that could cause convergence into a non-native conformation.

3.1.6. Structure Validation

The quality of a structural ensemble must be assessed in terms of its agreement with the experimental restraints (NOE, dihedral and other restraints), and also on the basis of covalent geometry (Nabuurs et al. 2004; Spronk et al. 2004). The backbone or heavy atom root mean square deviation (RMSD) of the retained ensemble of lowest energy structures is also evaluated as a measure of convergence. PROCHECK-NMR (Laskowski et al. 1996) is a commonly used validation program that tests Ramachandran plot statists, side-chain torsion angles, bond geometry, and electrostatic interactions. Programs such as WHATIF (Vriend 1990) and MolProbity (Davis et al. 2007) are also currently used by the wwpdb.org (Berman et al. 2000) to validate structures before they are published online.

3.2. MATERIALS AND METHODS

3.2.1. Sample Preparation

Dr. Lingling Xu performed protein expression, labeling of W_1 with ^{13}C and/or ^{15}N NMR active isotopes, and purification. The 199 amino acid W_1 sequence (199 and not 200 due to cloning purposes with S200 missing) is outlined in Figure 4.

Expression and purification of W_1 were performed by Dr. Lingling Xu and are detailed herein for the reader's reference.

A synthetic gene was produced (Integrated DNA Technologies, Coralville, Iowa) to encode a 199-aa protein (W_1) that corresponding to the 200-aa consensus repeat sequence of the *Argiope trifasciata* AcSp1 protein, with the N-terminal serine of the 200-aa repeat omitted. 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 into a pET plasmid vector (Novagen, Darmstadt, Germany). *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 1g/L $^{15}NH_4SO_4$ (Cambridge Isotope Laboratories, Andover, MA) and 2g/L ^{13}C-D-glucose (Cambridge Isotope Laboratories, Andover, MA) to provide uniform ^{15}N- and ^{13}C- enrichment. Expression was induced with isopropyl-β-Dthiogalactoside (IPTG) and cells were incubated overnight at room temperature. The resulting overexpressed His$_6$-SUMO-W$_1$ fusion protein was purified using affinity chromatography on Ni-NTA Sepharose (Qiagen, 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 (Xu et al. 2012b).

Lyophilized uniformly ^{13}C /^{15}N-enriched W$_1$ (~3mg) was suspended in NMR buffer (90% deionized water, 10% D_2O containing 20 mM CD_3COO^-, 1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) and 1mM NaN_3 at pH 5.02 ± 0.05. The protein suspension was vigorously mixed for 1 min and centrifuged at 2000 g for 5 min to precipitate any remaining protein aggregates. The supernatant containing the soluble protein was decanted, filtered through a 0.45 mm 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 as measured by ultraviolet absorption spectroscopy at 210 nm (estimated ε_{210}: 270858 M^{-1} cm^{-1}) in a 0.5 cm path length quartz cuvette (Hellma, Müllheim, Germany).

3.2.2. Stability Studies By Circular Dichroism (CD)

W$_1$ with uniform ^{15}N-enrichment was weighed (2.2 mg) and dissolved into 500 µL of deionized water and filtered through a 0.2 um filter to get rid of irreversible aggregates formed as a result of protein lyophilization. The sample was placed in a 0.1 mm quartz cuvette (Hellma, Müllheim, Germany) and the CD spectrum was collected between 260 to 180 nm at 20 nm/min in 0.1 nm intervals using a J-810 spectropolarimeter (Jasco, Easton, MD, USA) with a temperature control water-jacketed cuvette holder. The spectra were collected from 5˚C to 30˚C in 5˚ increments and 10˚ increments from 30˚C to 50˚C. Three spectra were collected at each temperature and exported into Excel. The average spectrum at each temperature was blank corrected and converted to mean residue ellipticity using a concentration (~0.01 mM) determined by ultraviolet absorbance at 210 nm described above.

3.2.3. NMR Data Collection For Structural Studies, Processing, And Assignment (Adapted From Manuscript (Xu et al. 2012b))

All NMR experiments (with exceptions detailed below) were collected on a 16.4 T Avance III spectrometer (Bruker Canada, Milton, ON) equipped with a 5 mm indirect detection TCI cryoprobe. Additional ^{1}H-^{15}N HSQC and HNCO experiments (the exception) were acquired at 11.7 T on an INOVA spectrometer (Varian, Palo Alto, CA) equipped with a 5 mm HCN cold probe. The HNCO was acquired with a larger spectral width in the ^{13}C dimension to alias some of the folded CO peaks. See Table 11 for all experimental parameters, with exception of the series of ^{1}H-^{15}N HSQC experiments used for stability studies. Initial temperature screening and stability studies were performed for AcSp1 W_1 with ^{1}H-^{15}N HSQCs at 5˚, 10˚, 15˚, 20˚, 25˚, 30˚, 35˚, 40˚ C with 64 scans and 2048×32 points in the ^{1}H and ^{15}N dimensions.

W_1 NMR experiments used for structural elucidation were acquired at 303.15 K. Backbone resonances were assigned using the following 3D experiments: HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCANH, (Kay et al. 1990; Bax and Ikura 1991), HNHA(Vuister and Bax 1993), HNHB (Archer et al. 1991), and ^{15}N-edited NOESY-HSQC (mixing time: 85 msec) (Marion et al. 1989; Marion et al. 2002). In addition to the latter four experiments, side chain chemical shifts were determined with the aid of 3D HCCH-TOCSY (Bax et al. 1990; Olejniczak et al. 1992) (mixing time: 12 msec), ^{13}C-edited NOESY-HSQC (mixing time: 85 msec), aromatic optimized ^{13}C-edited NOESY-HSQC (mixing time: 85ms) (Marion et al. 1989; Marion et al. 2002), ^{13}C-edited TOCSY-HSQC (mlev spinlock for 60 msec), and ^{1}H-^{13}C HSQC-NOESY-^{1}H-^{15}N HSQC (Diercks et al. 1999)(ChNH for short) (mixing time: 85 msec) experiments. Arg, Asn and Gln side chain assignments were determined using both 2D ^{1}H-^{15}N HSQC and the ^{15}N-edited NOESY-HSQC experiments. 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).

3.2.4. Experimental Restraint Generation For Structural Calculation

3.2.4.1. Distance Restraints From NOE Experiments

Distance restraints were generated from the ^{13}C-edited NOESY-HSQC, ^{13}C-edited NOESY-HSQC, ^{1}H-^{13}C HSQC-NOESY-^{1}H-^{15}N HSQC, and ^{15}N-edited NOESY-HSQC spectra within Analysis (Vranken et al. 2005) using the "Make Distance Restraint" command, dividing the distances into bins. The ^{15}N-edited NOESY-HSQC peak assignments led to 1285 distance restraints, the ChNH peak assignments generated 638 distance restraints, the ^{13}C-edited NOESY-HSQC peak assignments generated 2267 restraints, and the aromatic ^{13}C-edited NOESY-HSQC peak assignments generated 55 distances restraints. The restraints were exported from Analysis into CNS format. Using a python script written by David Langelaan, a previous PhD student from our group, the CNS distance restraints were converted to the Xplor-NIH (Schwieters et al. 2006) format. All restraint lists were concatenated into a single restraint file.

3.2.4.2. Generating H-Bonds Restraints

Uniformly labeled ^{15}N-enriched W_1 was produced by Lingling Xu and transferred into a 5 mm NMR tube with a protein concentration ~0.2-0.3 mM in NMR buffer. ^{1}H-^{15}N HSQC experiments at 16.4 T were used to monitor H/D exchange at 0h (control at 90%H_2O, 10%D_2O), 6h and 40h time points with 24 scans, 2048x192 points in the ^{1}H and ^{15}N dimensions respectively, and a recovery delay of 1.5 s. H/D exchange was performed by dialyzing out the original NMR buffer containing 90% H_2O 10%D_2O and replacing it with 100%D_2O NMR buffer (20 mM acetate buffer, 1mM DSS, and 1mM NaN_3, pH 5) using centrifugal 50mL spin dialysis filters (EMD Millipore). Backbone amide peaks remaining at 6h and 40h were assigned and used to infer H-bond restraints within Analysis' 'Make H-bond' module, with distances of 2.2 Å between the O···H in N-H···O=C, and based on the premise that the non-exchangeable H^N are H-bonded to the i-4 'O' (Wüthrich and Wagner 1979; Sticke et al. 1992).

3.2.4.3. J-Couplings Restraints

HNHA (Vuister and Bax 1993) and HNHB (Archer et al. 1991; Düx et al. 1997) experiments were acquired with the parameters detailed in Table 11. The less overlapped cross-peaks were assigned. In the case of the HNHA spectrum (59 cross-peaks), the peaks were converted using equation 3.5 into J_{HH} through the *J*-coupling module within Analysis (Vranken et al. 2005). The Karplus relation used was A = 6.51, B = -1.76, and C = 1.60 with a phase of -60 for the angle and the transfer period (d23 = $1/4J_{(\mathrm{HNH}\alpha)}$) was 13.05 msec.

The HNHB peaks were also assigned but Analysis (Vranken et al. 2005) does not contain a module that allows the conversion of the peak intensities into *J*-coupling values. Therefore, the peak heights and assignments (not volume, due to overlap) were imported into Microsoft Excel, converted to *J*-couplings manually, and converted to Xplor-NIH format (Schwieters et al. 2006). The delay period for the HNHB was 37.8 msec. The Karplus relation used for the HNHB is as follows (Demarco et al. 1978; Vuister et al. 2002):

$$\begin{aligned} {}^{3}J_{NH^{\beta 2}} &= -4.4\cos^{2}(\chi+120)+1.2\cos(\chi+120)+0.1 \\ {}^{3}J_{NH^{\beta 3}} &= 4.4\cos^{2}(\chi-120)+1.2\cos(\chi-120)+0.1 \end{aligned} \tag{3.5}$$

3.2.5. NOE Filtering With ARIA2

ARIA2 (Rieping et al. 2007) was used to filter the large degree of ambiguity in the initial set of W_1 NOE assignments (sometimes ~20-30 ambiguous options for a peak). Distance restraints imported from CcpNMR Analysis (Vranken et al. 2005), dihedral angles predicted using DANGLE (Cheung et al. 2010), and HNHA-based *J*-couplings (with constants A = 6.51, B = -1.76, and C = 1.60 with a phase of -60) were used for restraints with ARIA2.

The general protocol I followed for simulated annealing was as follows. Eight iterations were performed in torsional space, with 20 or 40 structures calculated in each step, using the best 5-10 as the template for the next round, sorted by total energy, and cooling over 40000 and 32000 steps for the cool1 and cool2 steps. The first 3 iterations employed network anchoring, with high and minimal network anchoring score per residue thresholds of 6.0 and 1.0, respectively, for the first 2 runs and 4.0 and 1.0, respectively, for the 3rd iteration. The minimal network anchoring score per atom threshold was set to a value of 0.75, 0.5, and 0.25, respectively, for each of the first 3 iterations. The violation tolerance was initially 1000 Å and decreased to 5.0 Å at iteration one, then to 1.0 Å at the last iteration. For peak filtering, lower and upper bound corrections were used with values of 1.8 and 6.0, respectively. Force constants for the cool1 and cool2 steps were set at 50 and 400 for dihedrals, 50 and 50 for ambiguous and unambiguous restraints, 50 and 50 for H-bonds, and 0.2 and 1.0 for scalar couplings, respectively. These force constants were the default in ARIA2, with the exception of the dihedral and H-bond terms. In the last iteration, 100 structures were calculated and the best 20, sorted by energy, were used to assess final violations and were refined in water (Linge et al. 2003).

An initial ensemble of 20 structures (of 100) along with the new peak list and violation list were imported back into my Analysis project. Using the new peak list generated by ARIA, each NOE assignment was checked individually and ambiguity was re-introduced based on a 20 Å cut-off according to the structure. This process was iterated one more time with an ambiguity cut-off of 10 Å. The ARIA2 peak list generated from that last run was used for the final Xplor-NIH refinement (Schwieters et al. 2006).

Table 11 | NMR experiments and parameters pertaining to the structure calculation of W_1.

<table>
<tr><th>Experiment</th><th>Bruker pulse sequence</th><th>Relaxation delay</th><th># of scans</th><th>Increments</th><th>Spectral width (ppm)</th><th>Center position (ppm)</th><th>^{1}H frequency</th><th>Facility</th></tr>
<tr><th></th><th></th><th></th><th></th><th>F1|F2|F3</th><th>F1|F2|F3</th><th>F1|F2|F3</th><th></th><th></th></tr>
<tr><td>^{15}N-HSQC</td><td>hsqcetf3gpsi</td><td>1.5</td><td>8</td><td>4096|64</td><td>16|31</td><td>4.4724|117</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>Good ^{15}N-HSQC</td><td>hsqcetf3gpsi</td><td>1.5</td><td>16</td><td>2048|256</td><td>16|31</td><td>4.723|115.0</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>^{13}C-HSQC</td><td>hsqcetgpsisp2</td><td>1.5</td><td>8</td><td>4096|160</td><td>16|85</td><td>4.724|40.191</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNCA</td><td>hncagp3d</td><td>1.5</td><td>8</td><td>2048|32|42</td><td>16|23|24</td><td>4.724|115.5|50.5</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNcoCA</td><td>hncocagp3d</td><td>1.5</td><td>8</td><td>2048|32|42</td><td>16|23|24</td><td>4.724|115.5|50.5</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNCO</td><td>hncogp3d</td><td>1.5</td><td>8</td><td>2048|32|32</td><td>14|23|9.5</td><td>4.724|115.5|174.15</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNCO</td><td>[Varian]</td><td>?</td><td>?</td><td>1024|64|64</td><td>16|30.5|18</td><td>4.724|115.5|119.3</td><td>500</td><td>QUANUC</td></tr>
<tr><td>HNcaCO</td><td>hncacogp3d</td><td>1.5</td><td>24</td><td>2048|32|32</td><td>14|23|9.5</td><td>4.724|115.5|174.15</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>CBCANH</td><td>cbcanhgp3d</td><td>1.5</td><td>8</td><td>2048|32|108</td><td>14|23|60</td><td>4.724|115.5|39</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNHA</td><td>hnhagp3d</td><td>1.5</td><td>16</td><td>1024|96|32</td><td>14|23|8.8</td><td>4.724|115.5|50.5</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNHB</td><td>hnhbgp3d</td><td>1.5</td><td>16</td><td>1024|92|32</td><td>14|23|8.8</td><td>4.724|115.5|54</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>^{15}N-TOCSY</td><td>mlevhsqcetgp3d</td><td>1.5</td><td>8</td><td>2048|32|102</td><td>14|23|8.8</td><td>4.723|115.5|4.723</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>^{13}C-TOCSY</td><td>mlevhsqcetgp3d</td><td>1.5</td><td>8</td><td>2048|80|112</td><td>14|65|10</td><td>4.723|40|4.724</td><td>701</td><td>NRC-IMB</td></tr>
<tr><td>HcCH-TOCSY</td><td>hcchdigp3d</td><td>1.5</td><td>8</td><td>2048|104|60</td><td>14|8.8|35.5</td><td>4.724|4.724|56</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>^{15}N-NOESY</td><td>noesyhsqcetf3gp3d</td><td>1.5</td><td>16</td><td>1024|32|104</td><td>14|23|8.8</td><td>4.724|115.5|4.724</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>^{13}C-NOESY</td><td>noesyhsqcetf3gp3d</td><td>1.5</td><td>8</td><td>2048|80|112</td><td>14|65|10</td><td>4.724|115.5|37.5</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>Ar ^{13}C-NOESY</td><td>noesyhsqcetf3gp3d</td><td>1.5</td><td>8</td><td>2048|40|108</td><td>14|22|8.8</td><td>4.724|125.671|4.724</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HSQC-NOESY-HS</td><td>noesycngp3d</td><td>1.5</td><td>8</td><td>2048|80|112</td><td>14|65|10</td><td>4.699|115.5|37.5</td><td>700</td><td>NRC-IMB</td></tr>
</table>

3.2.6. Final Structure Calculations In Xplor-NIH

Xplor-NIH version 2.19 (Schwieters et al. 2006) was used for cycles 1-8, and only included NOE restraints. At each cycle, 96 structures were calculated and violations were assessed for those structures using an in-house tcl/tk script (as initially detailed in (Ding et al. 2006; Rainey et al. 2006)). 50-100 distance bin changes were made at each cycle if they violated by more than 0.5 Å for the 50% lowest energy structures.

Xplor-NIH was upgraded to version 2.32 in cycle 9. In the new version, 72 H-bonds were incorporated using the HBDA potential (Lipsitz et al. 2002; Schwieters et al. 2006), 256 TALOS+ dihedral angles were incorporated, and structures were calculated with the RAMA multidimensional torsion angle database potential term (Kuszewski et al. 1996b; Kuszewski et al. 1997). The NOE violation threshold was tightened to 0.3 Å at cycle 14, with NOEs that were violated in >35% of the structures being moved into longer distance bins and/or checked manually, and to 0.2 Å at cycle 18. The dihedral term (CDIH) was ramped up from 10 to 20 and 100 to 200 for the 'highTempParams' and 'rampedParams' terms, respectively; the scale was increased from 5 to 20 and the H-bond scale was increased from 1 to 5.

3.2.7. Assessing Violations

3.2.7.1. Distance Restraints

Distance restraints were assessed for violations using Dr. Rainey's in-house tcl/tk script that outputs the violations at the distance required. The backbone and all atom RMSD values relative to the lowest energy structure were obtained using VMD (Humphrey et al. 1996) between residues 15-140, 15-77, and 84-140. PROCHECK-NMR (Laskowski et al. 1996) was used to backbone dihedral angles of each structure of the ensemble. J-coupling violations were assessed by the violations output generated after structure calculations by Xplor-NIH.

3.2.7.2. H-Bond Violations

Since Xplor-NIH (Schwieters et al. 2006) does not output the violations for the H-bond restraints I devised a way to determine H-bond restraint violations. Since I initially started incorporating H-bonds with Xplor-NIH version 2.19, I had made a distance restraint file for H-bonds at the 6h time point. Using this file and Dr. Rainey's tcl/tk in-house script for obtaining distance violations, I determined which bonds had to be lengthened and which bonds were completely violating and should be taken out. Using

H-bonds as distance restraints, I was therefore able to determine H-bond restraint violations.

3.2.7.3. Dihedral Violations

Xplor-NIH (Schwieters et al. 2006) outputs the dihedral angle violations. In the first rounds, violating restraints were only commented out if they violated in 100% of the generated structures. After a few rounds, I started getting a little more strict and commented out dihedrals that violated 90% of the structures, then 80%, etc., until there were no more violations in more than 40 to 50% of the ensemble. At this point, it became more important to know by how many degrees the dihedral angles were violating and the number of violations per structure rather than the ensemble average given by Xplor-NIH. For this task, I made Microsoft Excel spreadsheets to compare the DSSP dihedral angle output (Kabsch and Sander 1983) dihedral angle output for each of the 20 lowest energy structures to the TALOS+ restraints. Comparing each DSSP angle to the range allowed by the TALOS+ angle and error, I was able to determine which angles of which ensemble members were being violated by how much per structure or on average.

3.2.8. Accessible Surface Area Calculations

The accessible surface area (ASA) was obtained using DSSP output (Kabsch and Sander 1983) for each of the 20 final ensemble members calculated from Xplor-NIH (Schwieters et al. 2006) and converted to an average for each amino acid in the sequence. The area in $Å^2$ was converted to a normalized % exposure to based on side chain area using a published table of ASA for each of the 20 amino acids in a GXG random coil (Yuan et al. 2006).

3.2.9. Molecular Dynamic Simulations Of W_1

The lowest energy structure of W_1 in the final Xplor-NIH (Schwieters et al. 2006) ensemble was used to simulate W_1 stability in water (spce water model) using GROMACS version 4.6.5 (Van Der Spoel et al. 2005; Pronk et al. 2013), and with the gromos54a8 force field (Reif et al. 2012). W_1 was solvated using a cubic box. 4 Na^+ and 6 Cl^- ions were added randomly at a concentration of 0.02 M (matching the acetic acid concentration in the original NMR buffer). The simulation was energy minimized over 500 steps, temperature, pressure, and the system's density were equilibrated for 100 ps each, and finally calculated over 500 ns, using 2 GPUs equipped with CUDA. The box contained 10169 water molecules. The protocol was adapted from the Bevan Lab GROMACS tutorials of the lyzozymes in water example. [http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/index.html].

3.3. PEAK PICKING AND STRUCTURE CALCULATIONS

3.3.1. The Thermostability Of W_1

Species of the *Araneae* family are cold-blooded, live in all regions of the world, and build webs outdoors (preferably) and indoors (only acceptable during fruit fly season). With this in mind, we had a range of workable temperatures for gathering NMR data. The purpose of the stability studies was to determine the best conditions for NMR and to establish the length of time a sample would survive under the temperature that would be used for gathering 3D NMR data. Acquisition of 3D NMR data is lengthy, requiring up to a week or more depending on the experimental time. The perfect temperature requires establishing a fine balance between a temperature that will not cause the protein to destabilize, aggregate etc. within 2-3 weeks but also provide enough kinetic energy through heat to decrease the tumbling time of the protein and correspondingly increase signal intensity through decreased peak width.

The initial temperature screening was performed using CD between 10˚ - 50˚C (Figure 18). The W_1 CD spectrum does not change as the temperature is increased and the protein does not unfold. The same conclusions hold using NMR (Figure 19): the ^{1}H-^{15}N HSQC peak pattern remains extremely similar from 20˚C to 40˚C. Importantly, W_1 is stable for a few hours at elevated temperature (>35˚C) but not for a week (Xu et al. 2013), which is the time required to gather most of the NMR experiments. Briefly, W_1 samples were incubated for lengthy periods of time (~2 weeks to a month) and analyzed by gel electrophoresis to look for evidence of degradation. W_1 is not stable in solution for extended times above 35˚C and tends to aggregate during longer periods of time, so we chose 30˚C for triple-resonance NMR experiments. All AcSp1 NMR experiments described from here onwards were carried out at 30˚C.

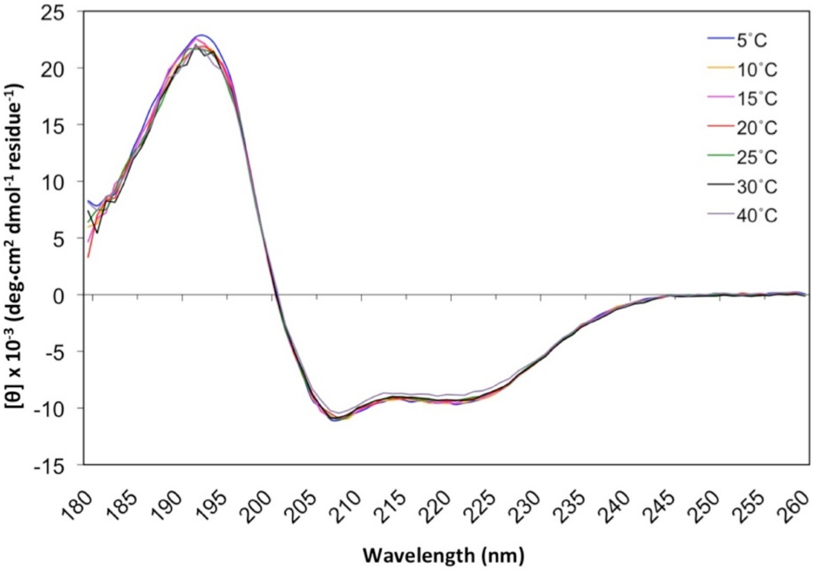


Figure 18 | CD spectroscopy of W_1 at various temperatures.

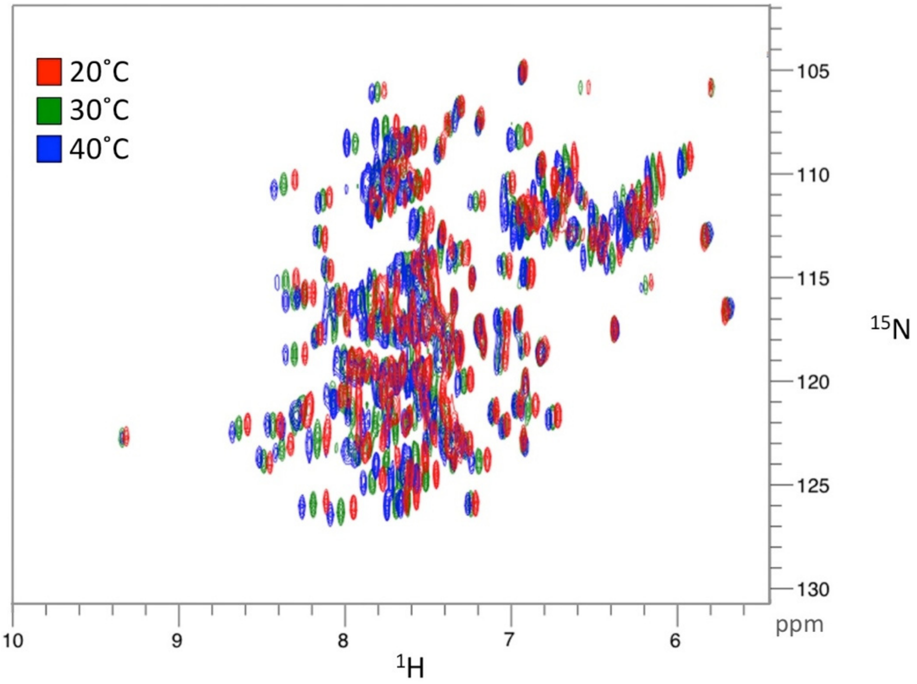


Figure 19 | Three representative ^{1}H-^{15}N HSQCs. The observed peak pattern was unchanged from 5˚C to 40˚C.

3.3.2. Assignment And Data Deposition (Adapted From Manuscript (Xu et al. 2012b))

Backbone ^{1}H, ^{13}C and ^{15}N resonances were 97.5% unambiguously assigned (Table 12). Most backbone amide peaks in W_1 are remarkably well dispersed, given the abundance of Ser (41), Gly (31), and Ala (29), which indicates that the protein is mostly structured and folded. The only unassignable residues were S152 and S153, located in the last polyserine section. Both S152 and S153 remained unassigned due to spectral overlap in the ^{1}H ^{13}C, and ^{15}N in the backbone spectra. T30, located in a region where some H^{N} chemical shifts are significantly perturbed, was the most difficult peak to link in backbone walk experiments due to its highly deshielded H^{N} (^{1}H: 10.5 ppm) and extremely low intensity, suggesting that this residue might be undergoing chemical exchange leading to line broadening and weak signals 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. According to BioMagResBank (BMRB) (Ulrich et al. 2008) statistics, chemical shifts in these ranges have been reported for <3% of the submitted assignments for those residues. Correspondingly, L31 and V34 also exhibit significantly perturbed H^{N} chemical shifts relative to their random coil chemical shifts (Δ0.76 ppm and Δ0.64 ppm respectively). Two other amino acids also have perturbed H^{N} chemical shifts: L100 and R76 with Δ1.71ppm and Δ1.28 ppm respectively, which are all well above their ^{1}H structural threshold of Δ0.1 ppm (Wishart et al. 1992; Tremblay et al. 2010).

There are ~27 unassigned ^{1}H-^{15}N cross-peaks in the ^{1}H-^{15}N HSQC that are connected to spin systems in 3D experiments. Some have intensities comparable to the assigned protein residues, while others are weaker. Based on chemical shifts, these appear to be spin systems corresponding to 7 glycines, 8 serines, 3 alanines, and 9 “other” amino acids. None of these spin systems appear to be sequentially connected, none have any NOEs and all remain unassigned. These spin systems may arise from an as yet unassigned minor secondary conformation of the protein.

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’

(Figure 21). The secondary structure was also predicted using the algorithm DANGLE, as implemented in CcpNmr Analysis (Vranken et al. 2005; Cheung et al. 2010). The consensus of secondary structure assignments of DANGLE and CSI indicates that W_1 is composed of 6 major helical regions (Figure 21). The last quarter of the protein (residues 160-199) appears unstructured on the basis of chemical shifts, but this is not surprising given that 7/8 Pro and 14/31 Gly are located in that region. The ^1H, ^{13}C, and ^{15}N backbone and side chain resonances were deposited in the BMRB with accession number 17899.

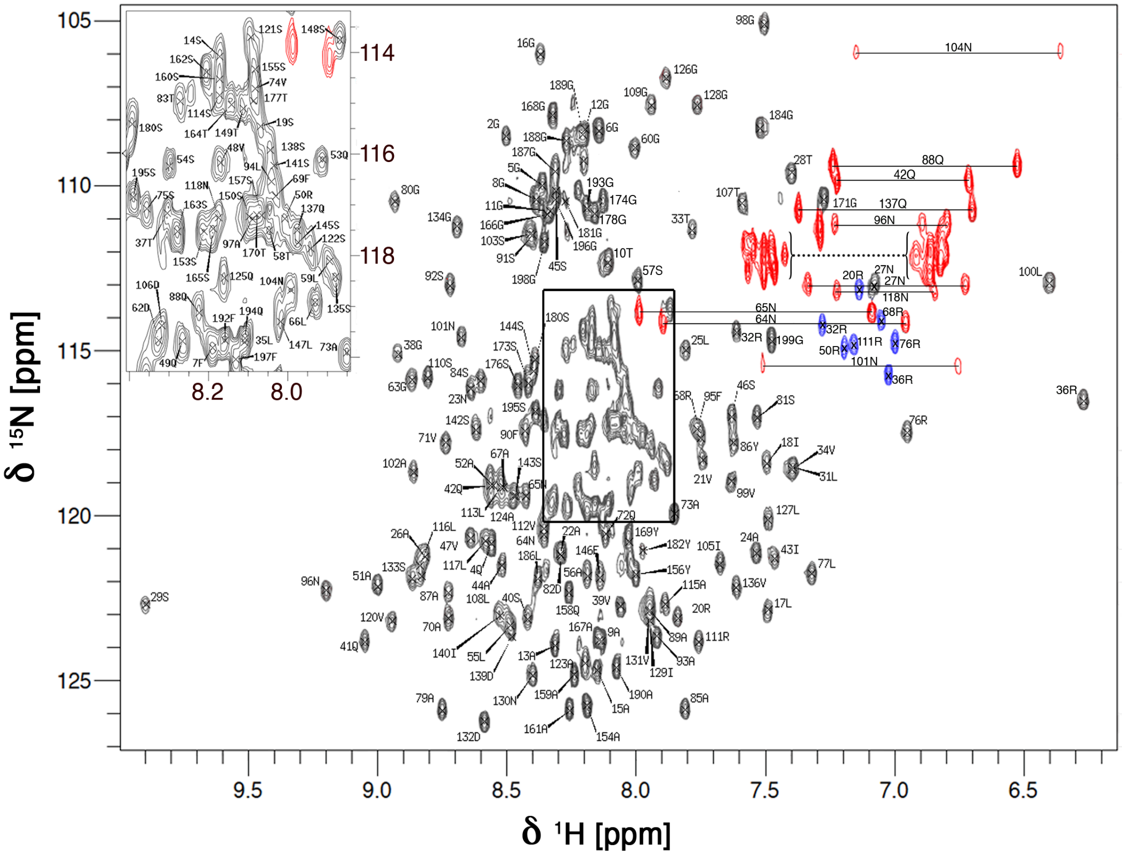


Figure 20 | ^1H-^{15}N HSQC of W_1 with assignments (figure from (Xu et al. 2012b)). The Asn H-N^{δ} and Gln H-N^{ε} side chain cross-peaks are in red and the Arg H-N^{ε} side chain cross-peaks are in blue.

Table 12 | W_1 resonance assignment statistics.

Category	Available	Assigned	%
Carbon	787	710	90.2
Proton	964	928	96.3
Nitrogen	240	209	87.1
Amide	389	376	96.6
Backbone	787	766	97.3
Backbone non H	597	578	96.8
Side-chain H	774	740	95.6
Side-chain non-H	430	341	79.3

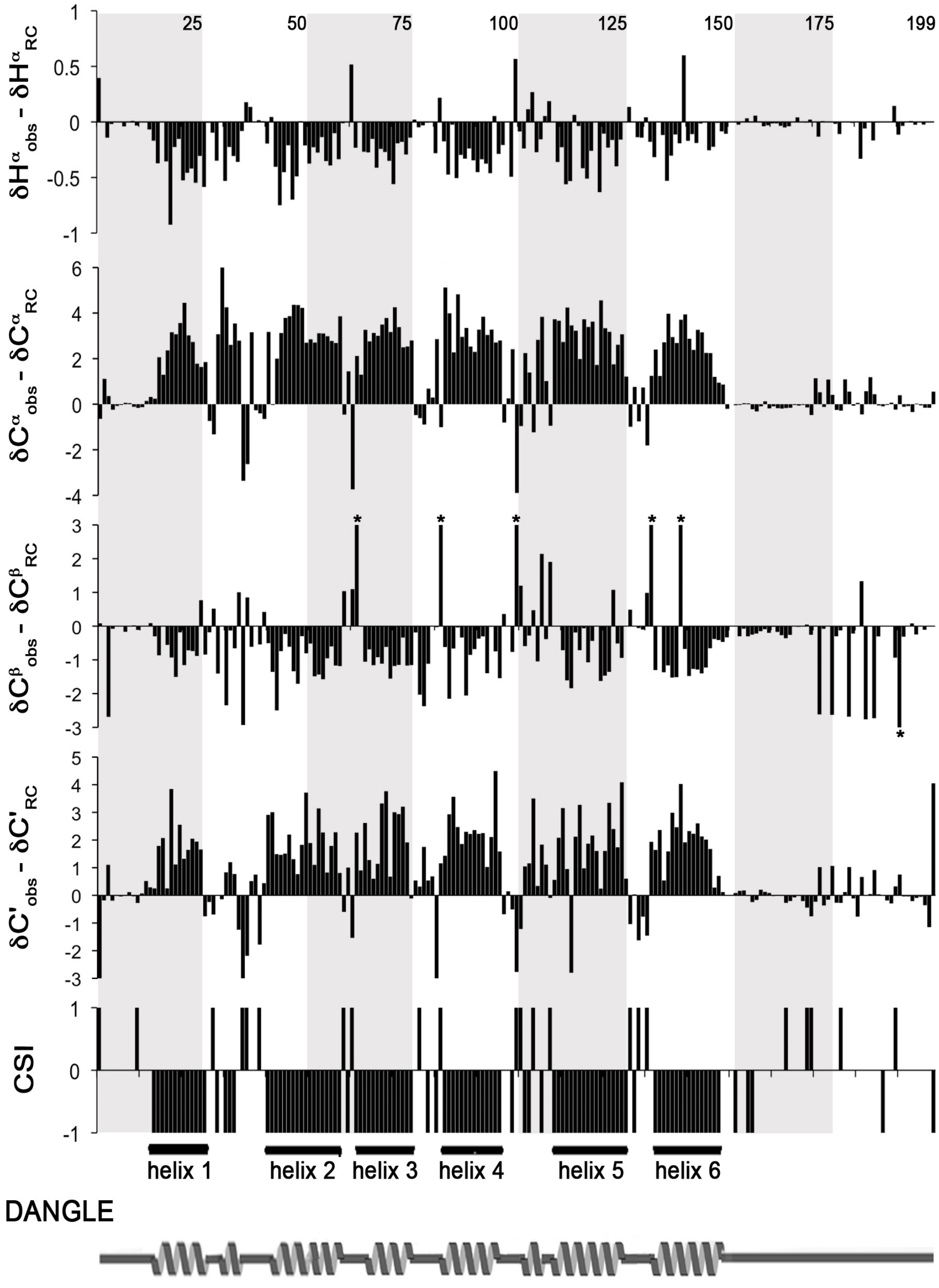


Figure 21 | Secondary chemical shifts for H^{α}, C^{α}, C^{β}, and C', chemical shift index (CSI) and DANGLE secondary structure prediction as a function of amino acid number. The CSI plot and DANGLE-based secondary structure prediction were generated by Analysis (Vranken et al. 2005; Cheung et al. 2010). Stars above bars indicate deviations greater than the range shown.

3.3.3. HNHA/HNHB To Dihedral Angles By J-Couplings

Peaks from *J*-coupling measurement with minimal overlap were carefully chosen from the HNHA and the HNHB (Barnwal et al. 2007) (Figure 22). 59 $^3J_{NH\alpha}$ couplings and 42 $^3J_{NH\beta}$ couplings were assigned given the overlap on the diagonal and in some cases between geminal H^β resonances. $^3J_{NH\beta}$ couplings can't be generated from the peak intensities if geminal H^βs chemical shifts overlap. Combining this factor with the overlap in the diagonal, the number of $^3J_{NH\beta}$ couplings produced was very low compared to the number of amino acids in W_1.

Unfortunately, the H^βs $^3J_{NH\beta}$ couplings were not very useful during either structure calculations by ARIA2 (Rieping et al. 2007) nor Xplor-NIH (Schwieters et al. 2006). Nevertheless, the couplings were exported and the list was transformed into a '.tbl' file to import into analysis and use with Xplor-NIH. A great advantage in having the 3D HNHA and HNHB data lay in the initial resonance assignments, since initially, no spectra containing side chain information were acquired.

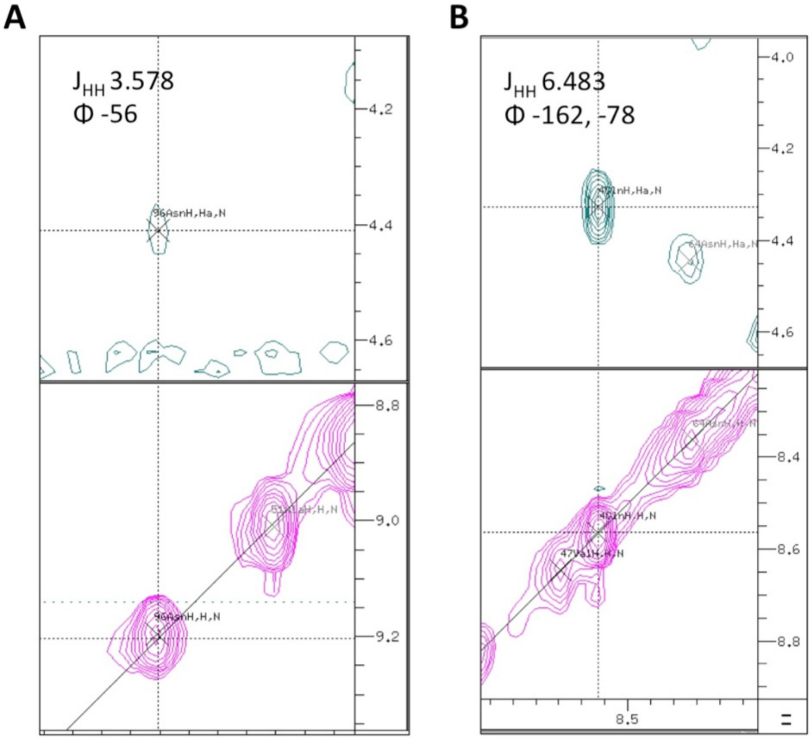


Figure 22 | HNHA example from my data demonstrating the manner in which the ratio between peak heights/volumes is analyzed in CcpNMR Analysis in order to infer a ϕ angle. A shows residue 96N and B shows 4Q.

3.3.4. H_2O/D_2O Exchange: inferring H-Bonds In W_1

Two time points were examined using ^1H-^{15}N HSQCs acquired at 700 MHz for the exchange of W_1 from an H_2O to D_2O environment: one 6h and one 40h post-exchange (Figure 23a). At the 6h time-point, 72 peaks remained, while at the 40h time point, only 40 peaks remained. More than half of the amides in W_1 demonstrate exchange with the solvent, including the side chains since no side chain correlations remained at the 6h time-point. The peaks remaining after the 6h and 40h time points were assigned, with intensities and volumes imported to Microsoft Excel for analysis. Ratios of remaining intensity or volume were calculated between each post-H/D exchange time point and the control 0h time point and are represented as cumulative ratios in Figure 23c.

The residues where peak intensity remained after 6h correlated very well with the α-helical regions, with exception of residues 135-150 where exchange was observed before acquisition but despite prediction of α-helical character by DANGLE (Cheung et al. 2010) and the CSI Figure 21). Thus, H-bonds restraints were inferred based on the peaks remaining after 6h and an assumption of α-helical structuring as implied by chemical shifts (Wüthrich and Wagner 1979).

Retrospectively, the H-bonds should have been based not only on amides protected from exchange but also on the basis of the NOE connection pattern (Figure 25). The typical NOE connection pattern for α-helices demonstrates a contact between protons H^{α} and H^{N} of residue *i* and *i*+4, respectively. Although there are some residues displaying that connection pattern, many proton connections were found between backbone residues *i* and *i*+3, implying 3_{10}-helices, as discussed in section 3.5 of this chapter. Had this been implemented, perhaps a larger number of H-bond restraints would have been satisfied during structural refinement.

More details on restraint violations during structural refinement are presented in the next section.

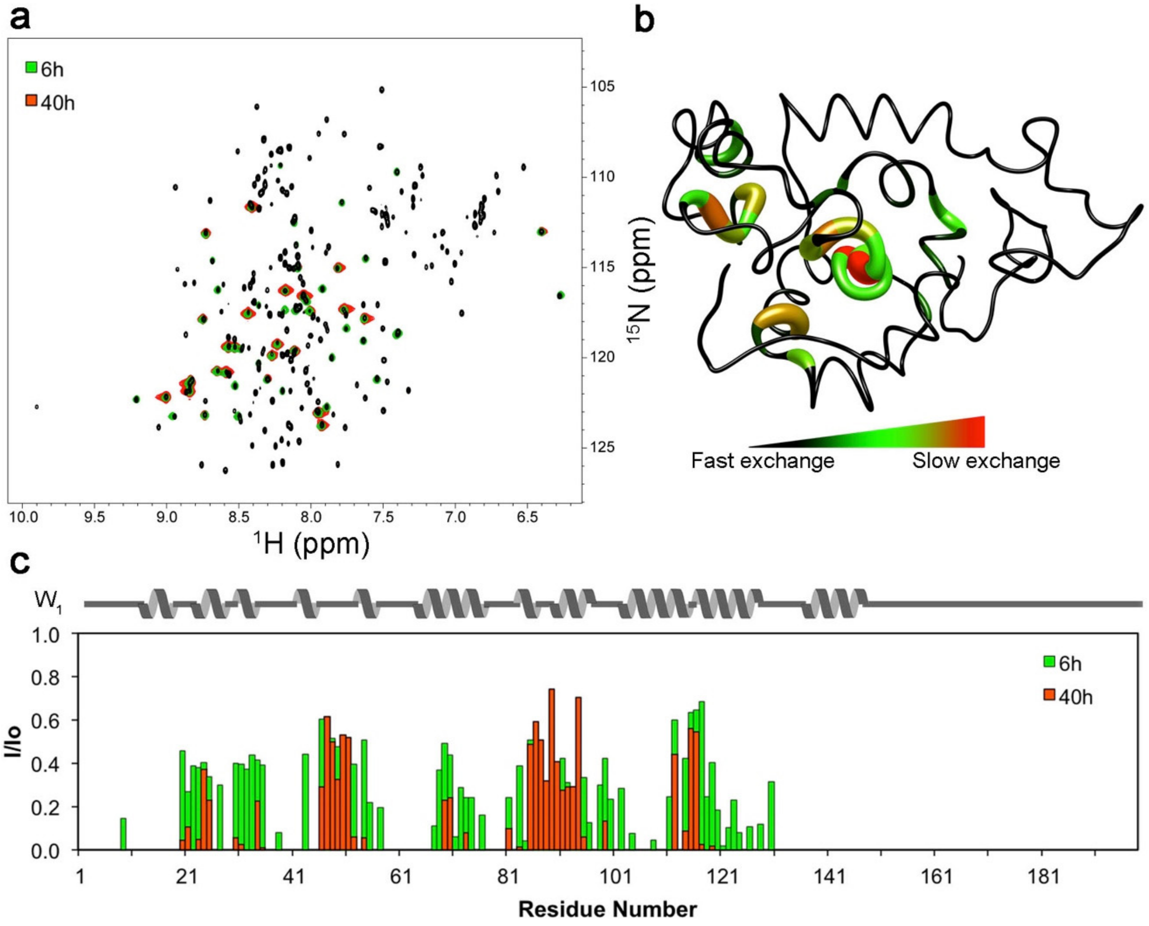


Figure 23 | W$_1$ H_2O/D_2O exchange. (a) ^{1}H-^{15}N HSQCs as a function of time following exchange of H_2O for D_2O. In black is 0h, green is 6h and red 40h post-exchange. (b) Peak intensities at the 40h time point represented on the lowest energy structure of W$_1$. (c) The bars represent the cumulative signal decrease ratio of ^{1}H-^{15}N HSQC peak heights between either the 6h or the 40h post-H/D exchange and the control (0h) sample. Above the graph is the representative linear cartoon of W$_1$.

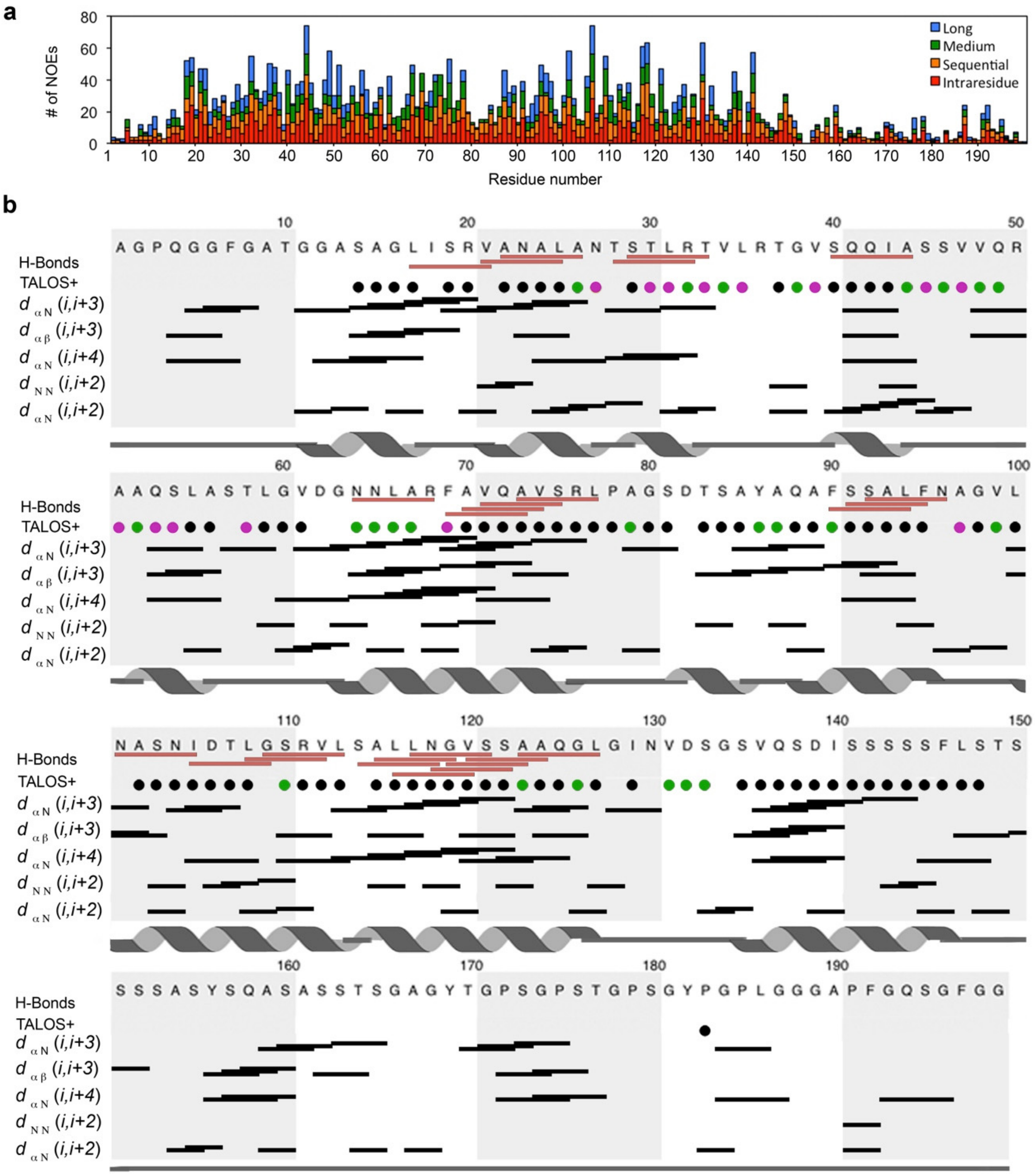


Figure 24 | Graphical summary of NOE, TALOS+ and H-bond restraints. (a) Graphical representation of the distribution of unambiguous NOE restraints as a function of residue following restraint refinement (*i.e.*, those employed in the final round of structure calculation using Xplor-NIH (Schwieters et al. 2006) with water refinement). (Legend: sequential (*i* ± 1), medium (*i* ± 2, 3, or 4), long (*i* ± > 4)). (b) Summary of H-bond, TALOS+ dihedral angle restraints (Shen et al. 2009a) (circles coloured black for both phi/psi for the given residue, magenta are psi angle only and green are phi only) and canonical (Wüthrich 1986) helical *i*+2, 3 and 4 NOE restraints following restraint refinement. The linear cartoon structure of W_1 is displayed at the bottom, on the basis of DSSP (Kabsch and Sander 1983) analysis of the final ensemble of 20 NMR structures.

3.3.5. The Long Journey Of Filtrating NOE Through ARIA2 And Xplor-NIH Refinement

The first distance restraint file fed to Xplor-NIH (Schwieters et al. 2006) resulted in an inability to produce a converged ensemble. We had a concern that XPLOR-NIH version 2.19 was not able to cope with the high amount of ambiguity in the distance restraint file (many restraints had $10\times{}^{1}H \rightarrow 10\times{}^{1}H$ = 100s of possible combinations). Because most NOE peaks had many assignment possibilities, thorough filtering methods had to be used, hence the title "the long journey".

ARIA2 (Rieping et al. 2007), a program frequently employed for structure calculations on larger proteins, was used on W_1 for NOE filtering. One advantage of ARIA2 over Xplor-NIH is ARIA2's ability to integrate the ambiguity efficiently into simulated annealing using the ADR algorithm and, with the help of network anchoring, eliminating the most improbable NOE connections. Much of the restraint ambiguity disappeared over the first 8 iterations of ARIA2, and the structure converged into a single well-converged structure. This convergence was very promising at first glance, but the results from the ARIA2 run were doubtful; many restraints were completely removed, even if they corresponded to intense NOESY peaks. Sceptical about the rejections that ARIA2 made, I used the updated CcpNMR Analysis project to re-introduce some but not all distance restraints to the restraint list with a 20Å cut-off. Guidance from the final 20-member ensemble of structures (out of 100) from the last iteration was used as a starting point for distance restraint re-introduction. Despite this filtering, ARIA2 still rejected many peaks that were valid on the second round of structure calculations.

The process of calculating structures and re-introducing ambiguity was repeated once more using another 8 iterations of ARIA2, but the ambiguity was reintroduced with a 10Å cut-off this time. Following the last ARIA2 structure calculation with the 10 Å cut-off, the generated NOE peak lists were satisfactory with regard to peak assignments and the level of ambiguity. But, this time, most of the noise mistaken for peaks had been filtered out. The distance restraint list was trimmed from 1313 down to 1285 H-H^{N} NOE peaks from the ^{15}N-filtered NOESY; 2963 to 2267 and 63 to 55 peaks from the aliphatic and aromatic ^{13}C-filtered NOESY spectra, respectively; and, 657 to 638 peaks from ^{15}N-^{13}C HSQC-NOESY-HSQC (ChNH). Ambiguity in assignments was also greatly reduced, going from ~20-30 options from the initial peak assignments to about 2-3 options for most peaks. This level of ambiguity is entirely within the Xplor-NIH capabilities, and restraint lists were therefore reformatted for additional refinement using Xplor-NIH (Schwieters et al. 2006). The data used to generate H-bonds were acquired at a much later stage in peak filtering; therefore, these restraints were only introduced during this final Xplor-NIH refinement stage.

3.3.6. *The J-Coupling Issue*

Quite a large effort was spent trying to use the HNHA- and HNHB-derived *J*-coupling values for W$_1$ structure calculation. When used in both ARIA2 (Rieping et al. 2007) and Xplor-NIH (Schwieters et al. 2006), it was clear that the *J*-couplings were persistently being violated in large numbers, hence not particularly useful. The $^3J_{NH\beta}$ restraints especially increased the energy of the ensemble by a factor of 2-3 compared to calculations without them in ARIA2 (Rieping et al. 2007). *J*-couplings result in a 4-fold degenerate angle when converted to dihedral angles via the Karplus equation (Karplus 1996; Schmidt et al. 1999). The Karplus coefficients that are reported in the literature (Vuister et al. 2002) also contain uncertainties that contribute to errors. It is also possible that the *J*-couplings are actually an average value over a variety of ensemble-averaged conformations, contributing to inability to use *J*-coupling values to determine a single conformer (Schmidt et al. 1999; Wang et al. 2013). Since the HNHA *J*-coupling and dihedral angle restraints are redundant, the *J*-couplings were omitted for structure calculation at the final stage refinement.

3.4. THE FINAL STRUCTURE OF W$_1$

3.4.1. *W$_1$ Structural Features*

The final structural ensemble for W$_1$ was calculated using Xplor-NIH (Schwieters et al. 2006) with a water refinement step (PDB deposition ID 2MU3). The average RMSD was calculated over the folded domain, spanning residues 12-140. Although the last helical segment (135-150) (Figure 25b) is well folded, it is not well converged over the 20 retained lowest energy ensemble members (out of an ensemble of 96), which is likely due to the small amount of long-range NOE contacts between this segment and the remainder of the folded domain (Figure 24). There are only 11 NOE violations over the 20 retained ensemble members above 0.5 Å and 215 from 0.2-0.5 Å, which is minimal considering that a total of 5241 NOE restraints were employed (Table 13). The error associated with dihedral angle restraint satisfaction is also very small (4.38˚ total, 1.22˚ per structure on average), considering the initial margin of error used in simulated annealing was up to ~20˚ depending upon the angle.

Table 13 | Full NMR statistics on W_1.

NMR distance and dihedral contraints	
Distance restraints	
Total NOEs	5241
Intra-residue	1470
Sequential (i±1)	1368
Medium (i±2,3,4)	908
Long-range (i±>4)	846
Ambiguous	649
Hydrogen bonds	25
Dihedral angles	
ϕ	107
ψ	94
Energies	
E_{Total}	-4136±51
E_{NOE}	91.1±14
$E_{Improper}$	93.7±7.8
E_{Bonds}	6.9±1.6
E_{CDIH}	-509±28.6
E_{HBDA}	0.015 ±0.008
Structure statistics	
NOE Violations	
Number of violations >0.2 (Å)	215
Number of violations >0.5 (Å)	11
Avg distance violation (Å)	0.07 ± 0.06
Dihedral angle violations	
Avg violation per model in ensemble	1.22±0.13
Avg maximum violation in ensemble	4.38±1.2
Average RMSD (a.a.15-140)	
Backbone	0.88±0.18
Heavy atoms	1.32±0.15
Deviations from idealized geometry	
Bond lengths	1.039±0.003
Bond angles	0.307±0.008
Ramachandran plot statistics (%)	
Residues in most favoured region	71.0
Residues in additionally allowed regions	25.6
Residues in disallowed regions	3.4

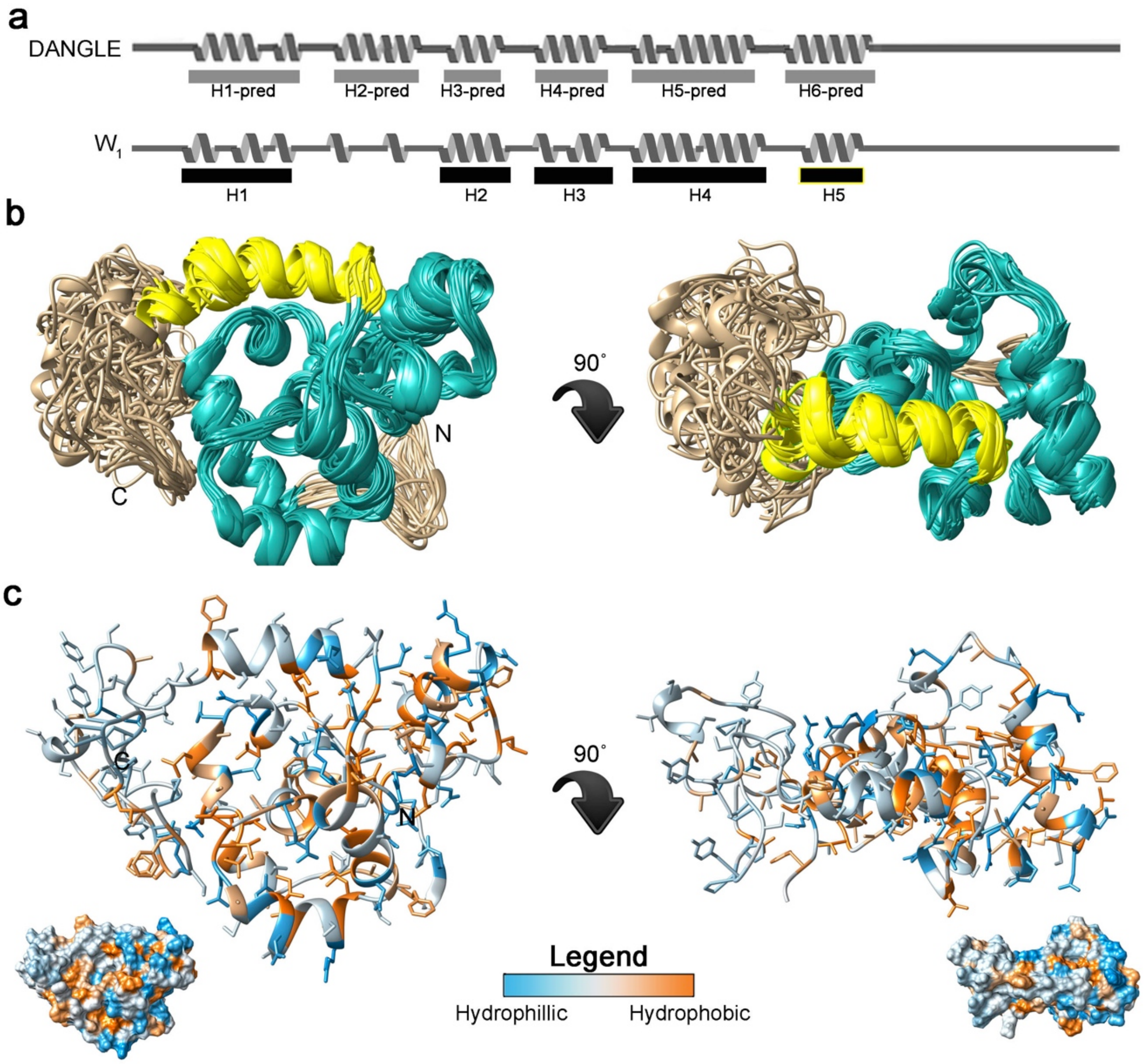


Figure 25 | Solution NMR structure of W_1. (a) Helical assessment of W_1 using DANGLE (Cheung et al. 2010) and linear structure inferred from the final W_1 structural ensemble. Helix numbers are annotated below each linear structure. (b) Overlay of 20 lowest energy members of the NMR ensemble. The helical domain over which the RMSD is calculated is in teal (helical segments 1-4; see figure Figure 32) and yellow (H5, discussed further with respect to Figure 23; the intrinsically disordered portions of W_1 are in tan. (c) The lowest energy structure coloured according to the Kyte-Doolittle hydrophobicity scale (Kyte and Doolittle 1982) shown in ribbon/atom and surface (inset) representations.

W_1 is a flat ellipsoidally-shaped protein, composed of a helical folded domain from residues 12-150 flanked by unstructured tails (Figure 25b; tan). The residues in a helical conformation total 40% ± 3%, which is lower than the 50% α-helical structure predicted for soluble full-length *Nephila clavipes* AcSp1 in the gland by deconvolution of Raman spectroscopy (Lefèvre et al. 2011). (No experimental data exists for full-length *A. trifasciata* AcSp1.). There are 5 defined helical regions made of 3_{10}-helices,

α-helices, and π-helices in the folded domain (Figure 25 and Figure 26), in contrast to the 6 predicted by DANGLE (Cheung et al. 2010). The DANGLE 'H2-pred' (see Figure 25) is not present in W_1; within this region, however, there are two 4-amino acid α-helical turns (Figure 26). Although not completely helical, 7 of the 20 residues have dihedral angles corresponding to α-helices (Figure 27), supporting the secondary structure assessment observation by CSI (Figure 21) and DANGLE (Figure 25). This region, located under the polyserine H5, was a challenging region to assign. Residue T30 is located within this region, which contains an extremely deshielded H^N chemical shift (10.3 ppm).

H5 is a relatively unstable helix, as disclosed by H/D exchange (Figure 23) and also by an elevated RMSD. Notably, the loop located before H5 from residues 128-134 exchanged with D_2O more slowly than H5, with peak intensity still visible at 6h but not 40h, implying that H5 is more dynamic than the rest of the folded core of W_1. Another flexible, non-converged segment is a loop that spans residues from ~77-82 that contribute to an elevated overall RMSD (Backbone RMSD residues 11-77, 83-140 = ~0.7 Å). P78 is located in that loop, which is the only proline in the folded core of W_1.

The amide moieties that resisted H/D exchange the most correspond to residues 46-54 and 86-95 (Figure 23). The region from 86-95 is located in the central core of the protein in an α-helix. The region from 47-56, on the other hand, is not buried in the core nor contained in a helical segment, although its backbone is somewhat pointed towards the interior of the protein and opposite to the flexible C-terminal face. Also worth mentioning, with regard to the H/D exchange results, is the long bent (at residue S114) amphipathic α-helix from residue 102-128, which is not exchanged at 6h but completely exchanged by 40h. These results show that W_1 has a densely packed folded core with varying degrees of accessibility to H/D exchange. A single stable conformation is apparent through structural determination, but slow time-scale dynamics from the NMR perspective are, therefore, apparent.

W_1 contains 4 phenylalanines and 7 tyrosines. Of these 11 aromatic residues, 8 of them are oriented towards the solvent, 2 (90F and 95F in H3) are located in the central core, and 86Y is located in H2 on the opposite face to the C-terminal tail (Figure 28).

The solvent exposed aromatics fall in the top 50% of solvent exposed residues, which is unusual for aromatic amino acids (Moelbert et al. 2004). H3 spans the central core of W_1. The side chains of F90 and F95 are directed towards the solvent on either face of the flattened part of W_1 (Figure 28; right structure). In the C-terminal tail, there are clearly two positions not exposed to solvent at 179-180 and ~191. These residues are prolines and are buried. H4 is an amphipathic α-helix (Figure 28C) with an ~133° bend at residue S144 according to the online program MC-HELAN (Langelaan et al. 2010). One face of the helix is solvent exposed and the other side is buried.

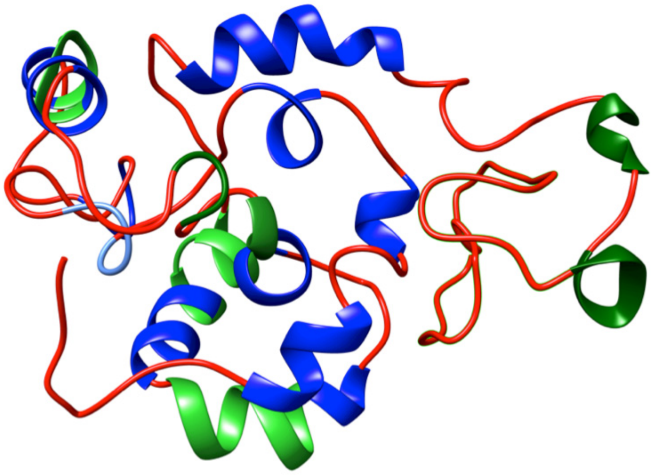


Figure 26 | Cartoon representation of W_1 coloured by secondary structure type. In blue are α-helical regions, light green are π-helices (I_5) helices, and in dark green are 3_{10} helices as defined by DSSP (Kabsch and Sander 1983). In red are random coil regions.

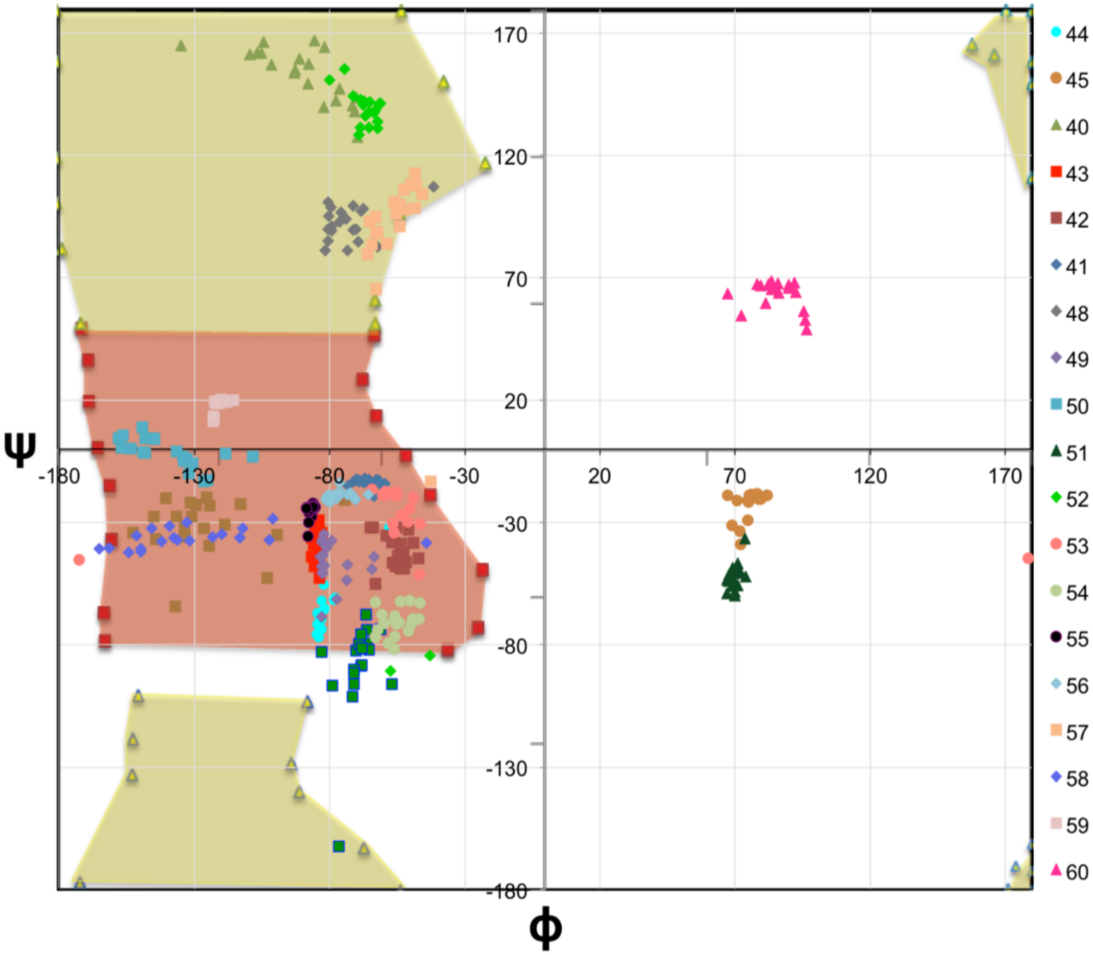


Figure 27 | Ramachandran plot of region between residues 40-60 (H2pred). In red is outlined the region for α-helices and in yellow is the region for β-sheets. Residues 45S 51A and 60G are in the glycine region or the turn II region and residues 40Q, 48V, 52A, 57S are in the β-sheet region. The Ramachandran plot boundaries illustrated are based on (Lovell et al. 2003).

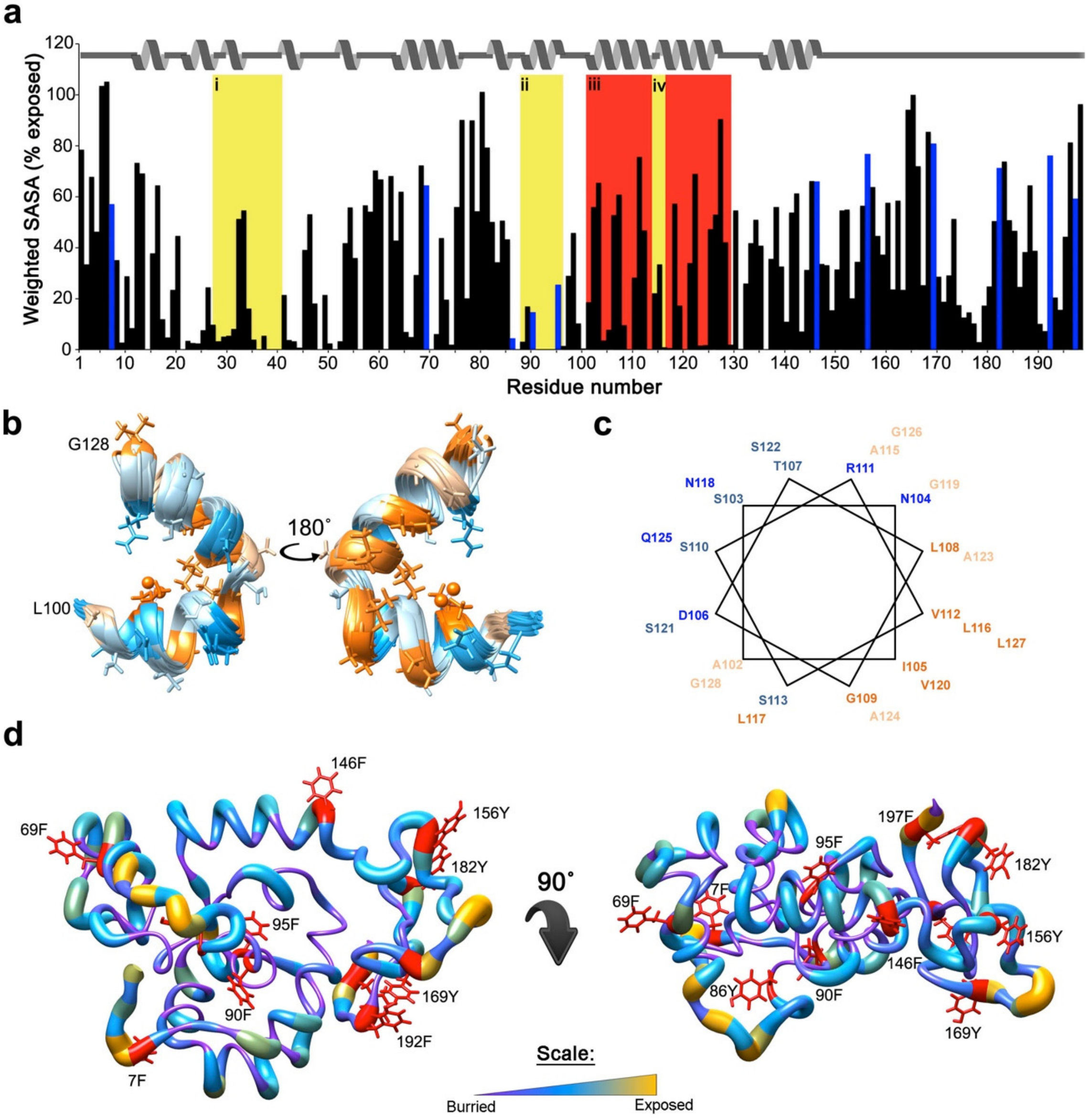


Figure 28 | Accessible surface area as a function of position in W_1. The accessible surface area was calculated by DSSP, weighted using values from Yang *et al.*(Yuan et al. 2006) and converted to a percent exposed area. (a) Graphical representation of ASA per residue. In blue are the aromatic residues that comprise the top 50 most exposed residues. Highlighted yellow block "i" represents H2-pred, "ii" is the central helical region where 90F and 95F are pointing out in opposing direction. The amphipathic H4 is represented by the red block (iii) with the yellow block (iv) displaying the kink in the helix. (b) H4 ribbon and atom representation colour coded by the Kyte-Doolittle hydrophobicity scale (Kyte and Doolittle 1982). (c) H4 helical wheel representation (residues 102-113, 115-128) displaying its the amphipathic properties where one side is hydrophilic and the other hydrophilic. Blue amino acids are polar and orange are non-polar as approximately coloured in (b). (d) W_1 coloured according to weighted DSSP ASA values. In red are the aromatic residues.

3.4.2. The Stability Of The W_1 Structure

The stability of a protein results from features such as salt-bridges, H-bonds, van der Waals interactions, disulfide bonds, and the hydrophobic effect (Voet and Voet 2010). Individually, salt-bridges, disulfide bonds, and H-bonds heavily contribute to protein stability. van der Waals and hydrophobic interactions are much weaker in isolation, but summed together significantly contribute to stability. There are no salt bridges within W_1. Most of the H-bond restraints between the amide of residue *i*+4 and carbonyl of residue *i* were rejected. Therefore, the remaining forces conferring the W_1 stability are van der Waals and hydrophobic forces.

The NOEs can also help define interactions that confer to stability. For instance, there are ~80 NOE contacts found between residues I18-A22 in H1 to F90-L100-A102, I105, and L108 found in H3 and H4, which form a hydrophobic cluster close to the C-terminal tail (Figure 30b). Residues I18-A22 also hold the C-terminal tail close to the core by hydrophobic interactions, with ~30 NOE contacts to 171-191 (Figure 30a). Residues 70-74 in H2 have ~35 NOE contacts to residues A44 and V48 located in the H2-pred region. Another region with a large number of long-range NOE restraints is from residues 26-37 to V136 and I140 in H5. These restraints hold H5 close to the remainder of the core. T30 and L35 also have contacts to A87 and F90 located in the central H3 helix. In addition, T10, located in the flexible N-terminal portion, has NOE contacts from its backbone $H^{\gamma 2}$ to the side chain of Q42, restricting movement of the N-terminus.

The hydrophobicity of W_1 represented using the Kyte-Doolittle scale (Kyte and Doolittle 1982) is shown in Figure 25c. The surface of W_1 is very evenly distributed between hydrophobic and hydrophilic residues, while the C-terminal tail is more neutral in character. H4 is amphipathic with the hydrophobic residues pointing towards the interior of W_1. L100, the residue having the most perturbed chemical shifts throughout the amino acid relative to statistical ranges for all residues reported in the BMRB (Ulrich et al. 2008), is located at the beginning of H4 and has strong NOE connections to L17, I18, and V21. L100 is part of a hydrophobic buried cluster located near the C-terminal and, along with L17, I18, and V21, I105, L108, A13, and A93, contributes to the hydrophobic core of W_1 (Figure 30a-b). The reason for its chemical shift deviation may be due to solvent exposure, as the side chain is very slightly solvent exposed in 7 of the 20 ensemble members.

One of the 7 arginines in W_1, R36, is buried within the center of the protein and is predicted to have the lowest pKa of all arginines in the protein according the software PROPKA (Rostkowski et al. 2011) (Figure 29). Only one H^{β} and part of the side chain moiety of R36 is surface accessible through a 'hole' located on the surface of W_1, ensuring the stability of a charged residue in the interior of the protein (Figure 31). But R36 is only ever so slightly solvent accessible: the exposed H^{β} is responsible for most of the solvent accessibility. S40 may help stabilize R36 by H-bonding of its Oγ with a distance of ~1.75Å with an angle of ~175˚ to the proton in the guanidinium moiety of R36. The oxygen in the -OH group of the T37 side chain is ~3 Å away from the R36 guanidinium moiety, G38 and V39 carbonyls are ~2.6 Å and 2.7 Å oriented at an angle of ~90˚, which doesn't allow strong H-bond interactions but still would help stabilize the charge.

To test the effect of burying this arginine upon protein stability, the lowest energy structure of W_1 was simulated over 500 ns in a solvent box containing water molecules with 20 mM of NaCl ions. The structure itself did not unfold over time, meaning that the calculated structures are stable and likely natively folded. This was further validation for the natively folded experimentally NMR W_1 ensemble presented within this thesis.

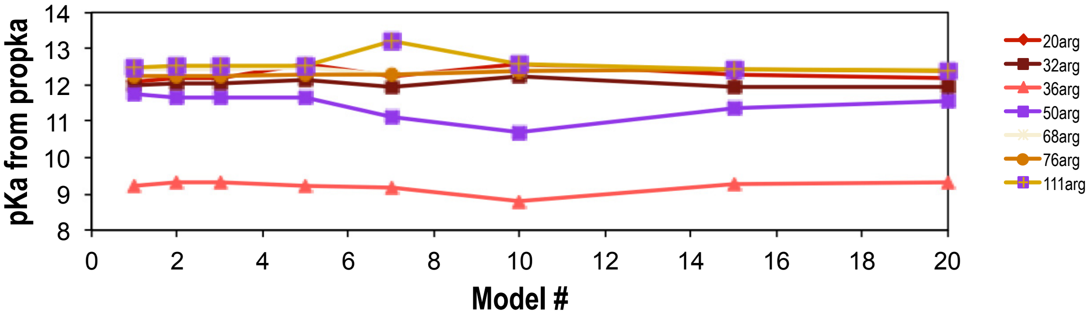


Figure 29 | Arginine pKa values predicted from the software PROPKA (Li et al. 2005). Here we see that the pKa of the buried Arg36 is much lower than the rest of the Arg in W_1.

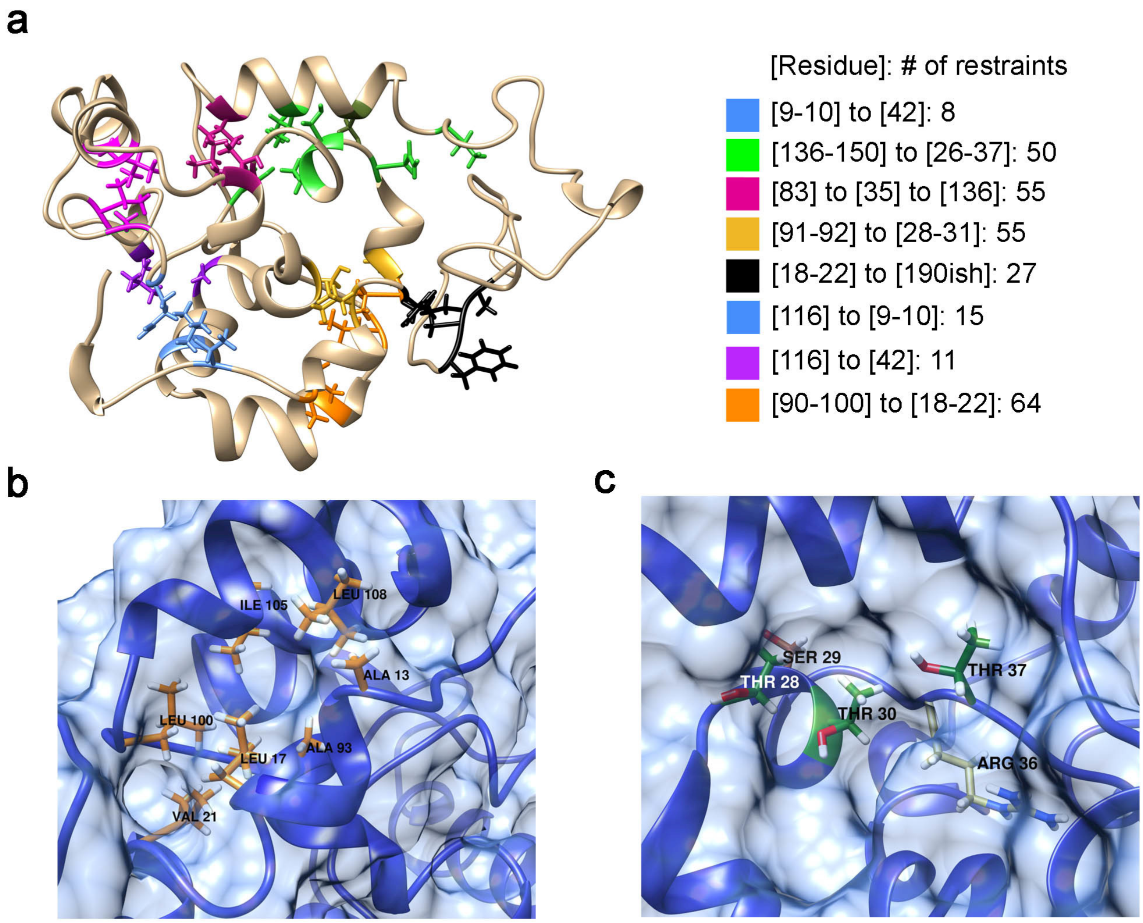


Figure 30| Representation of long-range NOE connections. (a) Colour-coded NOE groupings in W_1. (b) Hydrophobic patch incorporating L100. (c) The hard to assign region with low signal in the backbone 3D NMR data

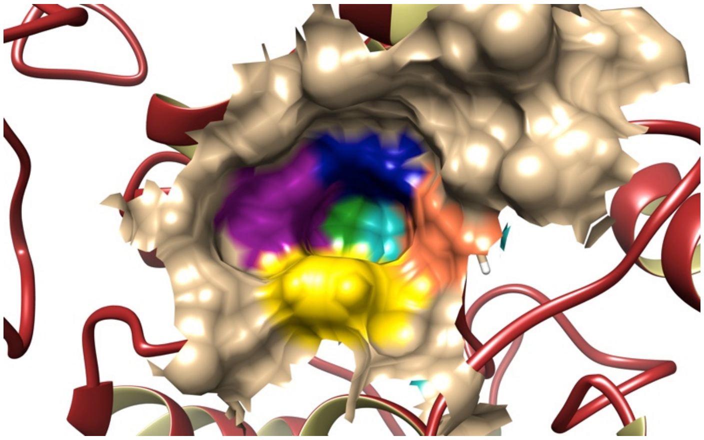


Figure 31 | The solvent accessible pocket in the center of the protein. Colouring is as follows: yellow: 95F, magenta: 23N, navy, 30T, orange: 37T, green: 91S, teal: 36R.

3.4.3. Bioinformatics: Assessing Disorder And Secondary Structure From Chemical Shifts And Sequence.

Four chemical shift based algorithms (TALOS+ (Shen et al. 2009a), the random coil index (RCI) (Berjanskii and Wishart 2007), the secondary structure propensity (SSP) (Marsh et al. 2006) and CSI (Wishart et al. 1992) (Figure 21)) and two structure prediction programs based solely on sequence (PONDR (Xue et al. 2010) and MetaDisorder (Kozlowski and Bujnicki 2012)) were used to asses the structure and the predicted amount of disorder in W_1 (Figure 32). In these algorithms, whether based on sequence databases and/or on chemical shifts, the first 11 N-terminal amino acids and the last 50 C-terminal amino acids are defined as disordered regions of W_1, with none of the programs implying a hint of structure or order in those regions Figure 21, Figure 32).

The core of W_1 is defined to be helical by CSI (Wishart et al. 1992), SSP (Marsh et al. 2006) and TALOS+ (Shen et al. 2009a), all of which agree with the amount of order/disorder predicted by MetaDisorder (Kozlowski and Bujnicki 2012), RCI (Berjanskii and Wishart 2007), and PONDR (Xue et al. 2010). There are three regions in the RCI (Berjanskii and Wishart 2007) where the order decreases in the core, somewhat reflected in the primary structure analysis by PONDR (Xue et al. 2010), and correlated with significant decreases in the helical propensity predicted by SSP (to a value of 0) (Marsh et al. 2006): ~38-40, 77-81, 129-131. Residue 78 is a proline, a residue known to break helices (Langelaan et al. 2010). Residues 77-84 constitute a disordered loop located opposite to the C-terminal tail in the folded structure. When not included, that loop that decreases the overall RMSD of W_1 from 0.89 Å (residues 12-140) to 0.71 Å (residues 12-77) and 0.72 Å (residues 84-140). The other two regions that show an increased disorder in the core do not contain prolines. Residues 38-40 are located under the flexible, H/D exchangeable H5, and also exchange readily with the solvent (Figure 23). The loop before H5 (residues 125-133) contains many long-range NOE contacts (Figure 24 and Figure 30a).

Overall, CSI (Wishart et al. 1992), RCI (Berjanskii and Wishart 2007), SSP (Marsh et al. 2006), MetaDisorder (Kozlowski and Bujnicki 2012), PONDR (Xue et al. 2010), DANGLE (Cheung et al. 2010) and TALOS+ (Shen et al. 2009a) agree (Figure 21 and Figure 32) with the structure I calculated through simulated annealing (Figure

25) with the exception of residues 40-60. The algorithms SSP (Marsh et al. 2006) and TALOS+ (Shen et al. 2009a) identically predict with utmost certainty the locations of all putative helical segments. The lack of helicity in W_1 H2-pred is certainly not caused by a scarcity of medium to long-range NOEs, since typical residues in this region have 25-40 *i* or *i*+1 NOE contacts; furthermore, the lack of helical character is supported through the absence of typical helical contacts that display step-like patterns in Figure 24b.

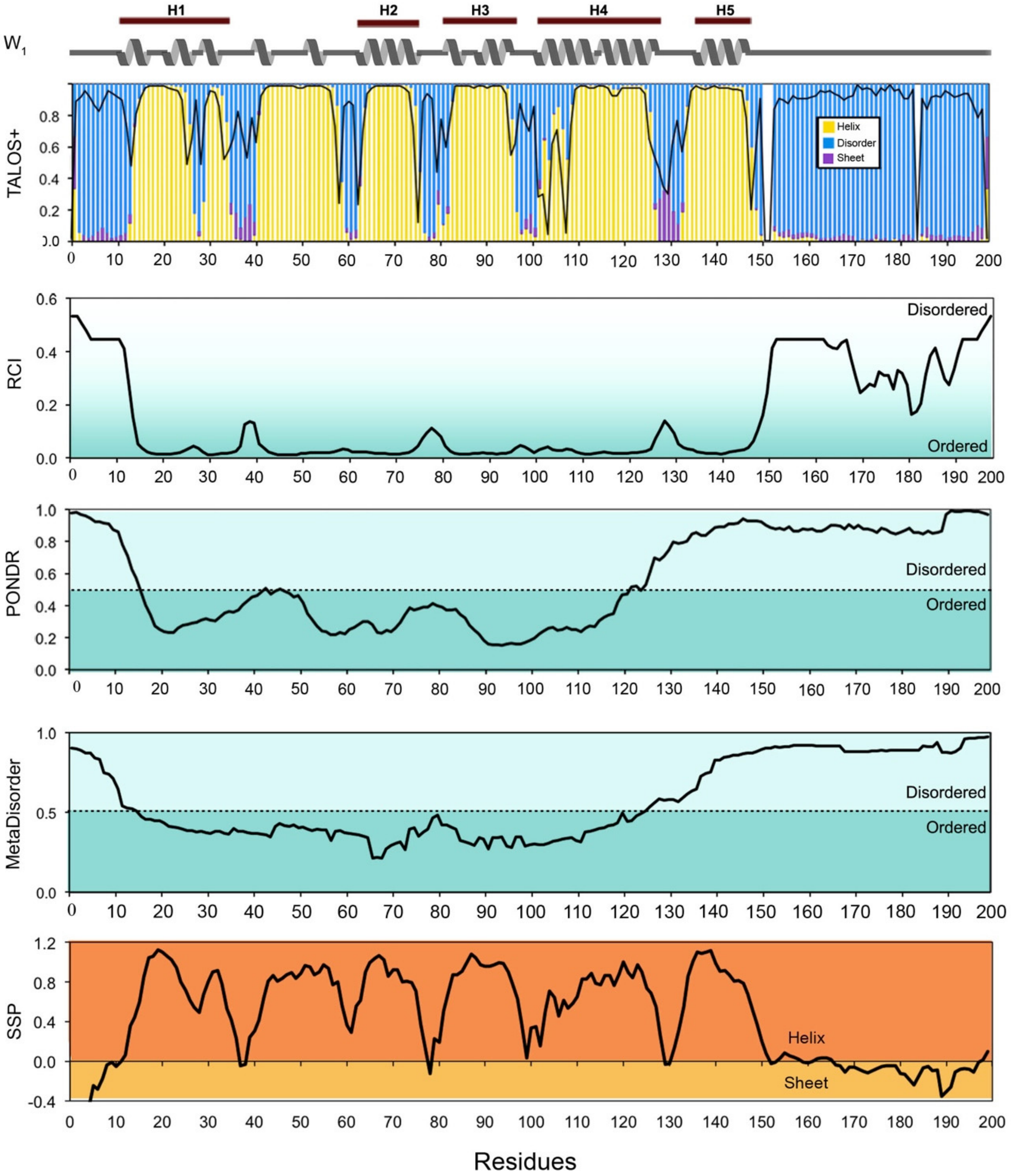


Figure 32 | Chemical shift and sequence-based prediction of structure and intrinsic disorder in W_1. Predictions of structured vs. disordered segments in W_1 are illustrated based upon chemical shift (TALOS+ (Shen et al. 2009a), the random coil index (RCI) (Berjanskii and Wishart 2007), and the secondary structure propensity (SSP) (Marsh et al. 2006)) and amino acid sequence (PONDR (Xue et al. 2010), and MetaDisorder (Kozlowski and Bujnicki 2012)). An extended cartoon representation of the converged features of the calculated W_1 structural ensemble is overlaid, with extended helical segments denoted. The black line in the TALOS+ (Shen et al. 2009a) panel indicates the confidence with which the secondary structure is assessed. The amount of disorder/order predicted by the RCI (Berjanskii and Wishart 2007), PONDR (Xue et al. 2010), and MetaDisorder (Kozlowski and Bujnicki 2012) is presented as a gradient (RCI) or a structural threshold (PONDR and MetaDisorder). The amount of sheets/helix is presented, in a different colour, as a threshold for SSP (Marsh et al. 2006).

3.4.4. Comparison of W_1 To The Other Structure Of AcSp1

At present, the only available structural information for an AcSp1 protein is from *Nephila antipodiana*. The solution-state NMR structure of a truncated (160 of 202 residues, RP_{tr}, PDB_ID: 2LYI; Figure 7), His_6-tagged repetitive domain of *N. antipodiana* AcSp1 was determined to be a compact, seven-helix bundle and a C-terminal disordered region (T147–G160) (Wang et al. 2012). Although the remainder of the repeat (161-202, RP_{ln}) was predicted to be unstructured based on our previously reported chemical shift assignments (Xu et al. 2012b), whether or not RP_{tr} remains the same structure in solution when residues 161-202 are present remains to be seen. Also notably, this repetitive domain only shares 27% sequence identity with the *A. trifasciata* sequence we used. Therefore, the structure I presented in this chapter is the first full-length structure of any wrapping silk.

The dissimilarity between the previous wrapping silk structure and my W_1 structure may be attributed to many factors, including the difference in pH and buffers used for NMR. The disparity in sequence between the two species may also be a large factor, with 37% sequence similarity and ~27% sequence identity. Notably, the RP_{tr} structure presented by Wang *et al* (Wang et al. 2012) also shows very high structural similarity to the *N. antipodiana* TuSp1 (Lin et al. 2009) spider silk repeat domain, part of the eggcase silk, determined by the same group. The *N. antipodiana* AcSp1 protein shares 26% sequence identity with the *N. antipodiana* tubuliform silk repeat unit (TuSp1) – this is approximately the same level of identity as our protein. One possibility is that repetitive domains from all types of silks have to be structurally similar within a spider species, but not between species. Or, it is possible that the *N. antipodiana AcSp1* gene was misannotated and in fact encodes a TuSp-related protein (X-Q. Liu, personal communication).

3.5. SUMMARY

The work I described in this chapter was dedicated to solving the structure of the repetitive domain of the *Argiope trifasciata* wrapping silk AcSp1 protein (W_1: 199 amino acids, lacking the C-terminal serine of the 200 amino acid repeat for cloning purposes) using high-resolution solution-state NMR. Three dimensional NMR experiments were collected at 16.4 T, processed using NMRPipe (Delaglio et al. 1995), and analyzed in CcpNMR Analysis (Vranken et al. 2005). Due to the W_1 sequence composition with 100 of 199 amino acids consisting of 41 serines, 31 glycines, and 29 alanines, the ambiguity in the initial NOE assignments was very high. Ambiguous distance restraints were first filtered with 3 passes through iterative structure calculations using ARIA2 (Rieping et al. 2007) by reintroducing ambiguity to 20 Å or 10 Å each time, using the generated structures as templates, until the NOE assignments fit the cross-peaks and the structure. A final cycle of simulated annealing was performed with Xplor-NIH with a water refinement (Schwieters et al. 2006) resulting in the final W_1 structure.

The stable, soluble form of W_1 exhibits a well-folded, flattened ellipsoidally-shaped domain (residues 12-149) flanked by disordered residues at the N-terminus (1-11) and C-terminus (150-199). Although chemical shift patterns are indicative of ~6 α-helices (Xu et al. 2012b), my structural refinement using NOE-based distance restraints with >20 NOE contacts per residue over most of the globular region, coupled with TALOS+ chemical shift-derived dihedral angle, and H/D exchange-derived H-bond restraints (Wüthrich and Wagner 1979) demonstrates 5 α-helical regions. My structure is folded quite differently from the other published structure of a truncated AcSp1 protein from *N. antipodiana,* with a major factor for this difference potentially being a very disparate sequence.

My H/D exchange experiments revealed that the fifth helix (residues 135-150) is a locally structurally destabilized region due to the high exchange propensity despite a strong helical presence. Sequence and/or chemical shift-based structure prediction algorithms corroborate the W_1 NMR structure with regard to prediction of unstructured and helical regions, with the exception of residues ~40-60; my NMR structure, based on experimental evidence, concluded this region to be converged but not α-helical. Subsequent chapters will build upon these results to fully characterize soluble AcSp1.

CHAPTER 4. THE STRUCTURE OF W_2

4.1.1. Premise

The lack of structure in the C-terminus and the first 11 amino acids at the N-terminus of W_1, work presented in the previous chapter, was intriguing. Although all chemical shift-based predictions confirmed the disorder in those regions (See section *3.5.4*), I was interested to see if the presence of another repeat unit concatenated at the N- or C-terminus of W_1 would cause the disordered segments to adopt a specific conformation. Of course, uniform labeling with NMR active isotopes of two identical concatenated repeat units would give rise to highly complex and ambiguous spectra due to overlap. To address this issue, segmental labeling of individual W repeat units within W_2, a fibre forming AcSp1 variant, was carried out using split-intein mediated *trans*-splicing technology (Southworth et al. 1998; Wu et al. 1998).

4.2. SEGMENTAL ISOTOP LABELING WITH INTEIN TRANS-SPLICING

Since their discovery, inteins have been applied in many aspects of structural biology and biotechnology (Volkmann and Iwaï 2010; Vila-Perelló and Muir 2010). Previous studies have demonstrated their use for protein cyclization (Scott et al. 1999; Evans et al. 2000; Volkmann et al. 2010), segmental isotope labeling (Yamazaki et al. 1998; Otomo et al. 1999; Muona et al. 2008; Busche et al. 2009; Skrisovska et al. 2010; Buchinger et al. 2010), protein switches (Ozawa et al. 2001; Mootz et al. 2003; Buskirk et al. 2004; Kanwar et al. 2013) and *in vivo* protein engineering and probe attachment for biophysical studies (Giriat and Muir 2003; Muir 2008; De Rosa et al. 2013).

Inteins (derived from 'internal protein') are proteins that auto-catalyze their excision from a polypeptide. In the process, the surrounding extein (external protein) is spliced together through formation of a native regioselective peptide bond (Goto and Kay 2000; Busche et al. 2009). Split-inteins, like their name suggests, are inteins that have been separated into two polypeptide chains fused to a single extein component. The components of the split-intein are referred to as intein-N and intein-C; -N if fused to the N-terminal domain of the extein and -C if fused to the C-terminal domain of the extein. Split-inteins are catalytically inactive individually, but functional once the complimentary split intein halves associate (Figure 33). In addition, the intein or split-intein auto-catalytic process does not require the use of any co-factors and demonstrates an efficiency of up to ~98% (Iwaï et al. 2006). The actual efficiency is a function of the intein-extein combination employed (Appleby-Tagoe et al. 2011).

For splicing to occur, the N-extein must contain a cysteine, serine or threonine residue at the ligation junction of the protein segment upstream of the intein, with cystein being the most efficient for ligation and threonine the least. The non-covalent association of the split-inteins is the first step of the reaction (Figure 33a-b) (Muralidharan and Muir 2006; Liu et al. 2009). Then, the N-O or N-S of the Cys, Ser, or Thr nucleophilically attacks the peptide bond of the C-terminal residue of the N-extein to form a linear (thio)ester. To free the intein's N-terminus, the first residue of the C-extein attacks the newly formed (thio)ester to undergo a *trans*-esterification. In this intermediate, the N- and C-exteins are covalently linked, but not in a peptide bond. The last residue of the intein-N is a conserved asparagine whose role is to facilitate bond rearrangement while excising itself out of the final product. In the final step, the C-extein N-terminal backbone amide attacks the (thio)ester joining both exteins together forming a peptide bond.

4.2.1. Project Goal

The use of split intein-mediated *trans*-splicing for NMR segmental isotope labeling has not been thoroughly exploited, although several successful attempts have been made (Züger and Iwaï 2005; Busche et al. 2009; Buchinger et al. 2010; Lee et al. 2014; Nabeshima et al. 2014; Schubeis et al. 2015) which are reviewed in {Skrisovska:2010hz, (Volkmann and Iwaï 2010). Intein-mediated *trans*-splicing was a very demonstrated to be an useful tool in my project, allowing structural study of individual W domains by NMR in the context of the larger, fibre forming protein. Segmental labeling of W_2 was performed through separate expression of each domain; one being enriched with ^{13}C and/or ^{15}N NMR active isotopes, and the other produced at natural abundance rendering it effectively invisible in triple-resonance NMR experiments. By ligating the segments together, two protein constructs with differing labeling schemes arose: one with the first W domain labeled and the second at natural abundance (termed W_{2-1}), and the second construct being *vice versa* (W_{2-2}) (Figure 33c). The chemical shifts of W_2 were sequentially assigned, and, based on these results, simulated structures of W_2 and W_3 were produced and placed in context of the larger AcSp1 protein.

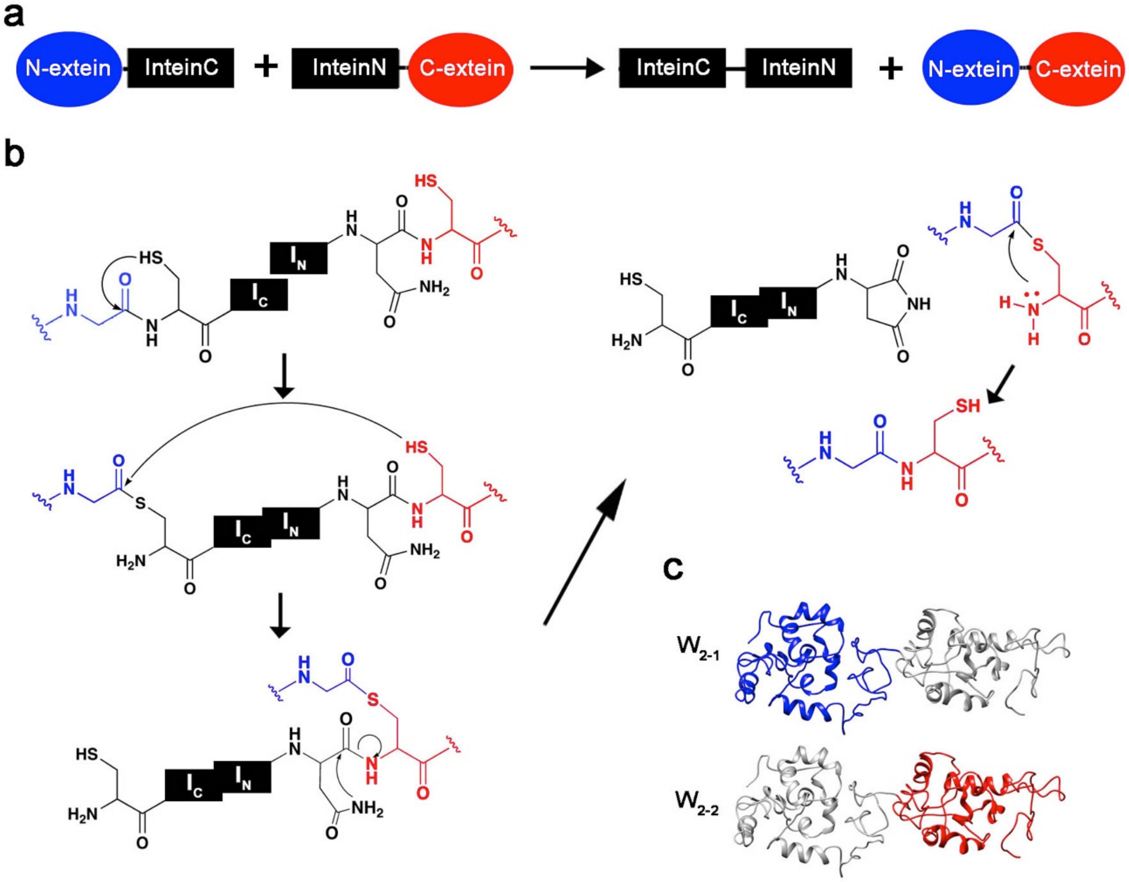


Figure 33 | Mechanism for intein-mediated *trans*-splicing. (a) General split intein *trans*-splicing where half of one intein (red) is attached to an extein (blue). The two inteins associate in solution and ligate the exteins through nucleophilic displacement reactions (b). The mechanism of protein ligation. In colour are the exteins that are the protein of interest and in black are the intein moieties. I_C is the Intein-C (for C-terminal) and I_N is Intein-N (N-terminal intein). (c) The resulting two protein constructs presented in this chapter. The ^{13}C-^{15}N labeled domains are coloured and the gray are at natural abundance.

4.3. METHODS

4.3.1. Protein Expression And Labeling

Expression and purification were performed by Dr. Lingling Xu. Details are provided for reference.

SUMO-W_1 and -W_2 fusion proteins were constructed in a modified pET32 plasmid, expressed, labeled with NMR active isotopes, cleaved using SUMO protease and reverse purified as previously described (Xu et al. 2012b; Xu et al. 2012a). For segmental labeling, two constructs containing the N-precursor (W_1In: W_1 + intein N (In) + His6x tag) and C-precursor (IcW_1: His6x tag + intein C (Ic) + W_1) were constructed for use with the split intein *Ssp* GyrB (Volkmann and Iwaï 2010). N- and C-precursors were purified by Ni–NTA affinity chromatography. To make segmentally labeled W_2 proteins through intein *trans*-splicing, an excess of unlabeled N- or C-precursor was mixed with labeled C- or N-precursor, respectively, to ensure that the labeled precursor was efficiently used. The splicing reaction was carried out in elution buffer (20 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0) with addition of 1 mM DTT at 4 °C for > 6 hours. The mixture was then dialyzed in 50 mM potassium phosphate, pH 7.5 at 4 °C for > 2 hours and reverse purified by passing through Ni–NTA Sepharose. The remaining precursors and intein fragments all have His6x tag and were trapped in the column, leaving the tag free W_2 protein flow through the column. Splicing and purification efficiencies were analyzed by SDS-PAGE and visualized by staining with Coomassie Brilliant Blue R-250. The amino acid sequence of W_2 and W_3 are presented in Figure 34 and Figure 35, respectively.

```
1   AGPQGGFGAT GGASAGLISR VANALANTST LRTVLRTGVS QQIASSVVQR
51  AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
101 NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSSSSFLSTS
151 SSSASYSQAS ASSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGGS
201 AGPQGGFGAT GGASAGLISR VANALANTST LRTVLRTGVS QQIASSVVQR
251 AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
301 NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSSSSFLSTS
351 SSSASYSQAS ASSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGG
```

Figure 34 | W_2 amino acid sequence (W_1+Ser+W_1).

```
1   AGPQGGFGAT GGASAGLISR VANALANTST LRTVLRTGVS QQIASSVVQR
51  AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
101 NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSSSSFLSTS
151 SSSASYSQAS ASSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGGS
201 AGPQGGFGAT GGASAGLISR VANALANTST LRTVLRTGVS QQIASSVVQR
251 AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
301 NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSSSSFLSTS
351 SSSASYSQAS ASSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGGS
401 AGPQGGFGAT GGASAGLISR VANALANTST LRTVLRTGVS QQIASSVVQR
451 AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
501 NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSSSSFLSTS
551 SSSASYSQAS ASSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGG
```

Figure 35 | W_3 amino acid sequence (W_1+Ser+W_1+Ser+W_1).

4.3.2. NMR Spectroscopy

Triple-resonance backbone walk NMR experiments (HNCA, HNcoCA, HNCO, HNcaCO, and CBCANH) were acquired for both segmentally-labeled W_2 proteins (~0.2 mM in NMR buffer [20 mM sodium acetate, 1 mM 2,2-dimethyl-2-sila-pentane-5-sulfonic acid (DSS), 1 mM NaN_3, at pH 5 in H_2O:D_2O, 9:1 (v:v)]) using a 16.4 T Avance III spectrometer equipped with a 5 mm TCI cryoprobe (Bruker, Milton, ON, Canada) at 303.15 K in the same manner as for W_1 (Table 14) In addition, 1D ^{1}H NMR experiments employing excitation sculpting and presaturation for water suppression were performed on natural abundance W_1, W_2, and W_3 (expression described in Xu *et al.* (Xu et al. 2012a) and carried out by Lingling Xu) at 11.7 T on an Avance NMR spectrometer equipped with a z-axis gradient and a BBFO SmartProbe (Bruker Canada). 2D and 3D experimental data were processed with NMRPipe (Delaglio et al. 1995) while 1D data were processed using Topspin 2.1. Spectra were assigned via CcpNmr Analysis (Vranken et al. 2005) and combined chemical shifts analysis was performed using the following equation for H^N, N, C^α, and C' nuclei (Schumann et al. 2007) as follows:

$$\Delta\delta_{comb} = \sqrt{\frac{1}{N_a}\sum_{i=1}^{N_a}(w_i\Delta\delta_{ij})^2} \tag{4.1}$$

where w_i is the weighing factor described as $|\gamma_i|/|\gamma_H|$, with γ_i being the gyromagnetic ratio of nucleus *i* and γ_H that of ^{1}H, N_a is the number of types of atoms, and $\Delta\delta_{ij}$ is the difference between chemical shifts *i* and *j* of the same nucleus between the spectra being compared.

Table 14 | List of NMR experiments for $W_{2\text{-}1}$ and $W_{2\text{-}2}$.

Experiment	Bruker pulse sequence	Relaxation delay (s)	Number of scans	Number of points F1|F2|F3	Spectral width (ppm) F1|F2|F3	Center position (ppm) F1|F2|F3	^{1}H frequency	Facility
HSQC	hsqcetf3gpsi	1.75	8	2048|128	14|25	4.7|116.5	700	NRC-IMB
HNCA	hncagp3d	1.75	8	2048|48|80	16|22.5|24	4.7|115.25|53.0	700	NRC-IMB
HNcoCA	hncocagp3d	1.75	8	2048|48|80	16|22.5|24	4.7|115.25|53.0	700	NRC-IMB
HNCO	hncogp3d	1.75	8	2048|48|64	14|22.5|10	4.7|115.25|176.0	700	NRC-IMB
HNcaCO	hncacogp3d	1.75	8	2048|48|64	14|22.5|10	4.7|115.25|176.0	700	NRC-IMB
CBCANH	cbcanhgp3d	1.75	8	2048|48|128	14|22.5|58	4.7|115.25|41.0	700	NRC-IMB

4.3.3. Calculation Of The Radius Of Gyration By Dr. Muzaddid Sarker (Included For Reference)

Translational diffusion coefficient (D_C) values for W_1, W_2 and W_3 proteins (303.15 K, 0.2 mM in NMR buffer, 1mM DSS and 1 mM NaN_3 at pH 7.5 (H_2O:D_2O = 9:1); 0.06% dioxane was added as an internal viscosity control (Jones et al. 1997)) were determined from ^{1}H diffusion ordered spectroscopy (DOSY) experiments employing pulsed field gradient (PFG) NMR (Morris and Johnson 1992) using an 11.7 T Avance NMR spectrometer equipped with a z-axis gradient and a BBFO SmartProbe (Bruker Canada). DOSY (64 scans, sweep width 12 ppm, relaxation delay of 2 s incorporating presaturation) employed stimulated echo and longitudinal eddy current delay (LED) with bipolar gradient pulses and two spoil gradients (Wu et al. 1995). The envelope of ^{1}H signals was attenuated by increasing the gradient strength from 2% to 95% in 16 steps. The observed signal intensity as a function of gradient strength was fit using a single component exponential fit of the signal decay and the D_C was determined from the fit using the Simfit program within the T1/T2 Relaxation module of Bruker Topspin 3.1 using the Stejskal-Tanner formula (Stejskal and Tanner 1965):

$$I = I(0)\exp[-D_C\times(2\pi\gamma g\delta)^2\times(\Delta-\delta/3)\times 1e4] \quad (4.2)$$

where I is the observed signal intensity, I(0) is the back-calculated unattenuated signal intensity, γ is the gyromagnetic ratio of ^{1}H (4257.7 Hz/G), g is the gradient strength (based on maximum amplitude 50 G/cm at 100%), δ is the gradient pulse length (8 ms), and Δ is the diffusion time (100 ms). The R_g (in Å) is calculated as (Tyn and Gusek 1990):

$$R_g = 5.78 \times 10^{-8} (T/\eta D_C) \quad (4.2)$$

where T is the temperature (in K), η is the viscosity (in cP) estimated using the experimentally observed D_C of a dioxane internal standard (Jones et al. 1997), and D_C is the observed diffusion coefficient (in cm^2/s).

4.3.1. Structure Calculations

The W_1 NOE, H-bond, and dihedral angle restraints retained in the final round of structure calculations (summarized in Table 13) were propagated over residues 200-400 (W_2 and W_3) and 400-600 (W_3) (Figure 34 and Figure 35) to generate restraint files for W_2 and W_3. 100-member structural ensembles were calculated in the same manner and using the same restraint weights as for W_1 without water refinement using Xplor-NIH 2.32 (Schwieters et al. 2006). Structural figures were produced in Chimera (Pettersen et al. 2004). Structural ensembles for W_{1-3} were calculated with or without the respective cylindrically modeled DOSY-derived R_g restraints (W_1 = 17.13 Å W_2 = 22.50 Å, W_3 = 28.65 Å). The scaling of R_g in the simulated annealing potential (Schwieters et al. 2006) was determined iteratively until a optimal value was reached, determined as the maximum weight at which no major increase in overall energy was observed. An in-house python script was used to test for NOE restraint violations, VMD (Humphrey et al. 1996) was used to obtain the RMSD, and PROCHECK-NMR (Laskowski et al. 1996) for the Ramachandran plot statistics. HYDROPRO (Ortega et al. 2011) was used to calculate a diffusion coefficient of each ensemble member based upon its coordinate file, allowing direct comparison of the calculated diffusion coefficient to the ensemble-average diffusion.

4.4. RESULTS AND DISCUSSION

4.4.1. Chemical Shift Comparison

Visualization of the W_{2-1} and W_{2-2} ^1H-^{15}N HSQCs overlaid upon W_1 for comparison demonstrates practically identical chemical shift distribution patterns for all three protein constructs (Figure 36). Although several peaks do clearly deviate in each construct, sequential assignment, as detailed below, demonstrates that these fall at the

W_2 linkage point. As would be expected from this result, uniformly ^{15}N-labelled W_2 also demonstrates the same peak dispersion as W_1 (Figure 37).

Backbone assignments for W_{2-1} and W_{2-2} were carried out independently of the previous W_1 assignments. Clearly, not only are most of the W unit backbone amide chemical shifts identical to those of W_1, but the C^{α}, C^{β} and C' shifts are as well (Figure 38, chemical shift data deposition BRMB 25197). It is only at the junction between the first and second W unit that significant perturbations in chemical shifts relative to those of W_1 are evident. These residues are F195, F198, and G199, G202 and Q204.

Since chemical shifts are highly dependent on environment (as detailed in chapter 2), the lack of chemical shift perturbation provides a strong indication that the domains are identically structured. Furthermore, stable domain-domain interactions would be expected to lead to chemical shift perturbation at least locally at sites of interaction; lack of such perturbation implies a lack of long-lived interactions. A postulated structure of AcSp1 would therefore be composed of independent, non-interacting structured domains (residues 12-150 in W_1) separated by intrinsically disordered linkers (residues 1-11, 151-200 in W_1), giving rise to a "beads-on-a-string" protein architecture.

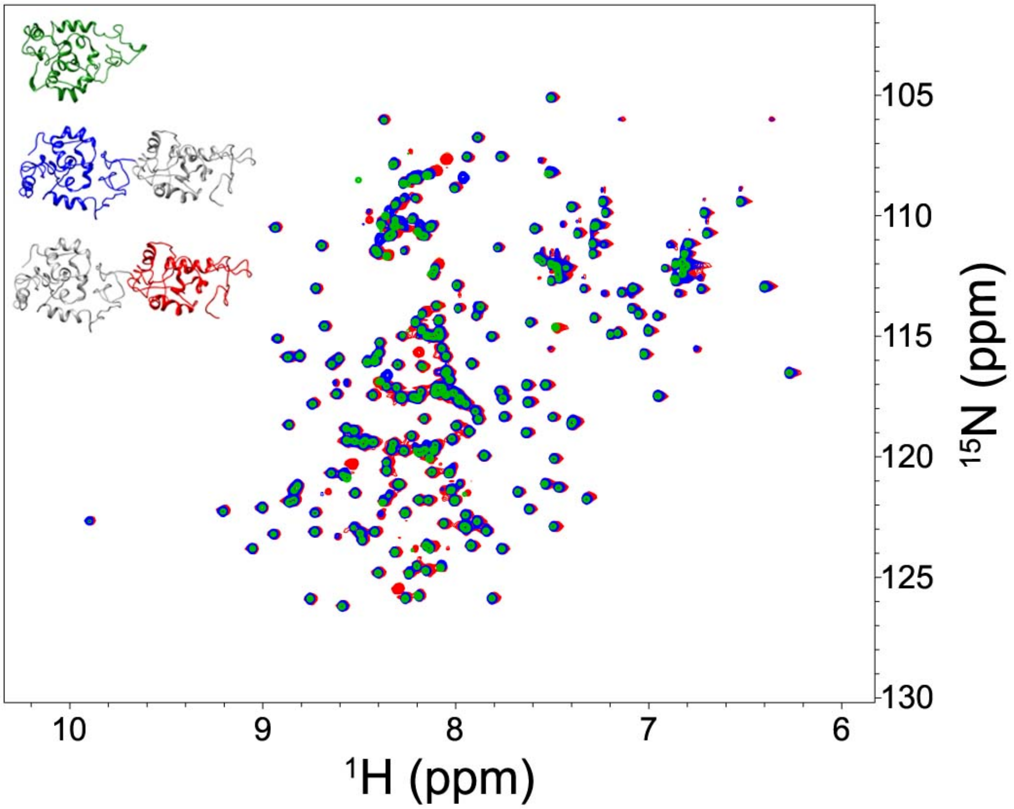


Figure 36 | ^{1}H-^{15}N HSQCs of W_1 (green), $W_{2\text{-}1}$ (blue), and $W_{2\text{-}2}$ (red) at pH 5 in deuterated acetate buffer.

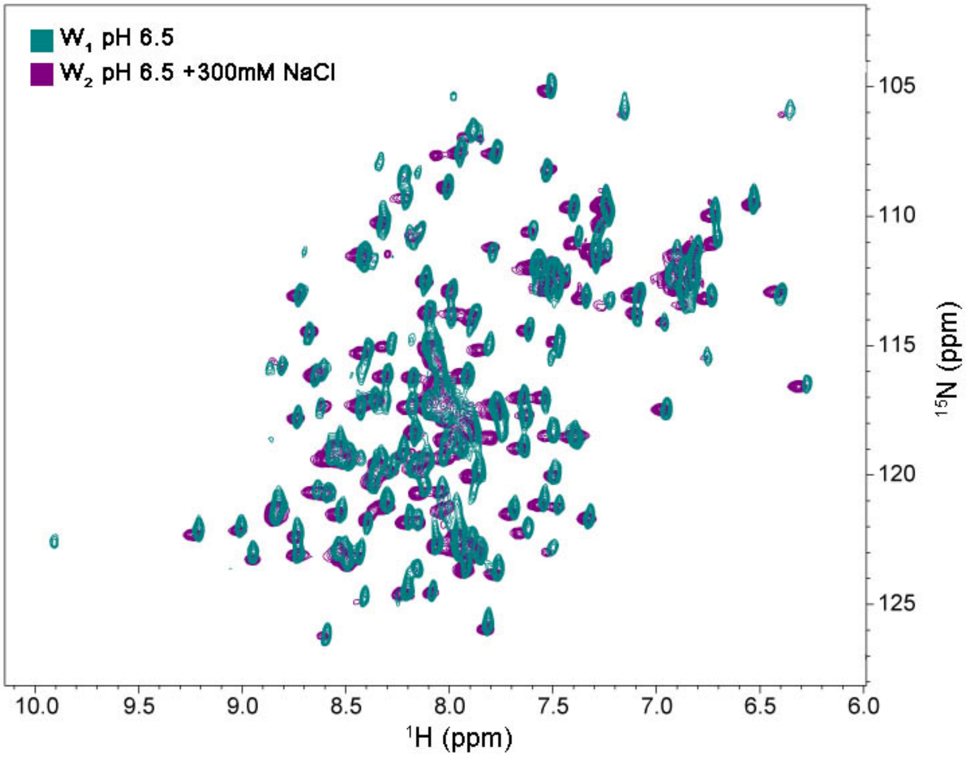


Figure 37 | Overlay of uniformly ^{15}N labeled W_1 and W_2 at pH 6.5 at 500 MHz.

4.4.2. The Structure Of W_2

Based on the outstanding chemical shift agreement between W_1 and W_2, I calculated a structural ensemble for W_2 using a concatenated set of W_1 NOE, dihedral angles, and H-bond restraints based upon the restraints used in the final iteration of Xplor-NIH water refinement (detailed in chapter 3). A second, parallel set of calculations was performed with the inclusion of a radius of gyration restraint (R_g) based upon experimental diffusion data, a restraint unique and specific to W_2. The statistics for the 20 lowest energy ensemble members (of 100) are reported in Table 15.

The radius of gyration provides a measure of compaction for a given protein (Lobanov et al. 2008) and is defined as the root mean squared distance of all atoms within a protein from the center of mass (Huang and Powers 2001). Given that the unstructured linker separating the folded domains of W_2 has very little NOE and dihedral angle information to restrict conformation, the calculated $W_{>1}$ structures are likely biased towards an extended conformation. The incorporation of an R_g restraint hence serves as a global, long-range restraint to limit protein dynamic expansion during simulated annealing and results in protein compaction. Incorporating an R_g as a structural restraint in simulated annealing has been shown to improve NMR-derived structural accuracy (Kuszewski et al. 1999).

Each individual W unit in W_2 converges to a structure indistinguishable from W_1. However, without the radius of gyration restraint (R_g), the positioning of the domains relative to each other is not restricted (Figure 39a). Although each folded domain is superimposable over all 20 lowest energy structures, the entire W_2 protein is not due to the flexible linker. Given that some residues in the linker have long-range NOE contacts to the globular core, the distance between the two folded cores is highly variable, but perhaps not as pronounced as one would expect.

With the addition of R_g restraint, the degree of domain-domain conformational variability and linker extensibility are restricted (Figure 39b and d). Notably, however, the orientation of one domain relative to the other is still rather loosely defined because, again, there are no restraints forcing a precise orientation of $W_{2\text{-}1}$ relative to $W_{2\text{-}2}$. The addition of R_g primarily restricts the extensibility of the linker and compresses the structure by bringing the domains close enough in space to satisfy the R_g but not so close as to violate NOEs.

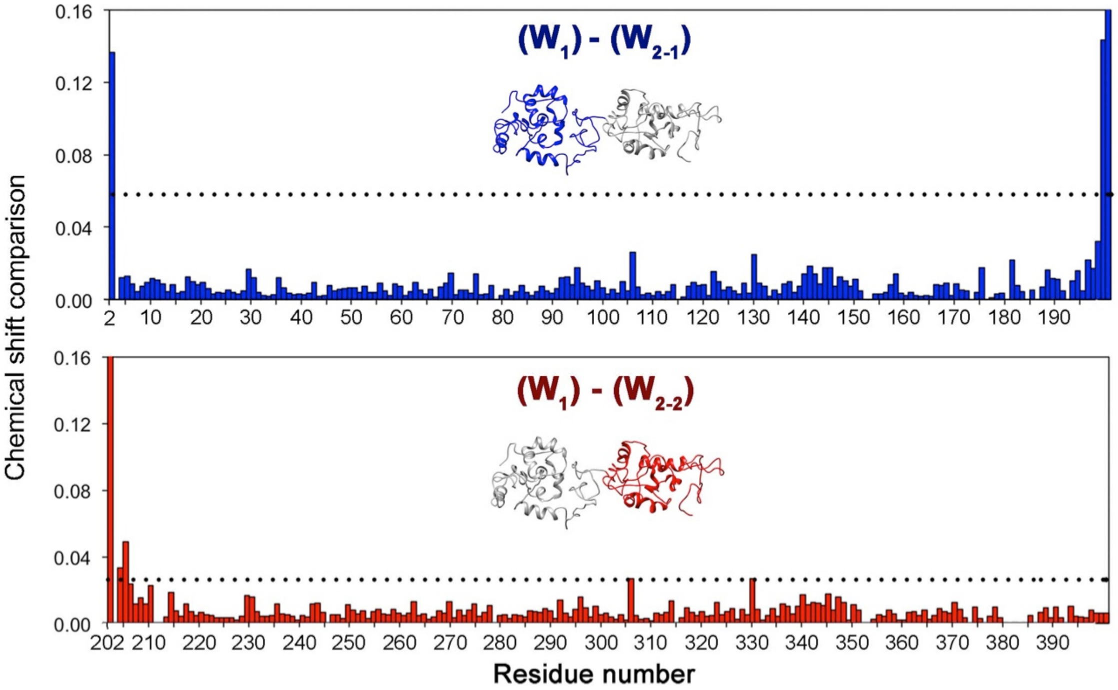


Figure 38 | Comparison of NMR chemical shifts for W_1, W_2 and W_3. Quantitative combined chemical shift (CCS) comparison (Schumann et al. 2007) between W_1 and W_2 (W_1-W_{2-1} in blue; W_1-W_{2-2} in red, where n in W_{2-n} refers to the W unit that is uniformly ^{15}N and ^{13}C-enriched in a given segmentally labeled W_2 concatemer). The CCS for C^{α}, H^N, N, and CO was calculated using the square root of the sum of squares weighted by the nuclei's gyromagnetic ratio. The mean plus one standard deviation (W_{2-1}: 0.054 + 0.011, W_{2-2}: 0.017 + 0.008) of both CCS is represented by the dotted line, providing an estimated significance in terms of chemical shift difference.

Table 15| NMR statistics for W_2 and W_3 with and without R_g.

Restraints	W_2		W_3	
Total NOE	10482		15723	
Intra-residue	2940		4410	
Sequential (|i-j| = 1)	2736		4104	
Medium-range (|i-j| < 4)	1816		2724	
Long-range (|i-j > 5)	1692		2538	
Ambiguous	1298		1947	
Hydrogen bonds	50		75	
Total dihedral angle restraints				
phi	214		321	
psi	188		282	
Radius of gyration (Å)	22.5		28.7	
Statistics	W_2	W_2 with Rg	W_3	W_3 with Rg
Energies (kcal/mol*rad^2)				
E_{total}	-922.1	-884.9	-1200	-948.7
E_{NOE}	212.7	200.2	351.9	434.8
$E_{improper}$	67.27	61.95	97.78	107.1
E_{bonds}	64.96	64.17	99.5	115.1
E_{CDIH}	0.97	0.81	1.08	2.03
RMSD aa 12-140, 212-340, 412-540				
backbone (Å)	0.83, 0.81	0.73, 0.86	0.76, 0.89, 0.88	0.91, 0.99, 0.97
All atoms (Å)	1.28, 1.27	1.22, 1.29	1.25, 1.39, 1.36	1.37, 1.44, 1.42
Ramachandran (%)				
favoured regions	65.1	64.7	64.6	63.9
generously and allowed regions	31.5	31.8	31.9	32.4
disallowed	3.4	3.5	3.5	3.8
# of NOE violations				
0.2-0.5	532	527	974	1172
>0.5	53	37	80	141

Since the diffusion coefficient (D_C) is inversely proportional to the R_g, the experimentally derived W_2 D_C was compared to the predicted D_C of each W_2 ensemble member using the program HYDROPRO (García de la Torre et al. 2000). On average, the D_C derived from structures calculated without the presence of an R_g was 6.48 ± 0.01 $x10^{-11}$ m^2/s, resulting in a 12% discrepancy relative to the experimental D_C ($7.37x10^{-11}$ m^2/s). When the R_g was included in the simulated annealing calculations, the average D_C of the structural ensemble increased to 6.97 ± 0.005 $x10^{-11}$ m^2/s, a 5.4% difference and demonstrating increased compaction through increased translational diffusion. Hence, the predicted D_C from the ensemble of structures calculated with an R_g match the experimental D_C more closely than that calculated without an R_g restraint. Since there are no major violations of other experimental restraints induced by incorporation of an R_g restraint, this implies that the structures in Figure 39b and d are closer to the representative solution structure of W_2 then panels a and c from the same figure. But, there is still a discrepancy between the experimentally derived D_C and the average predicted D_C for the structural ensemble. This could result from some preferred orientation of the domains relative to one another, which could have an impact on the diffusion of the protein, or it could result from oversimplifying assumptions made during calculations.

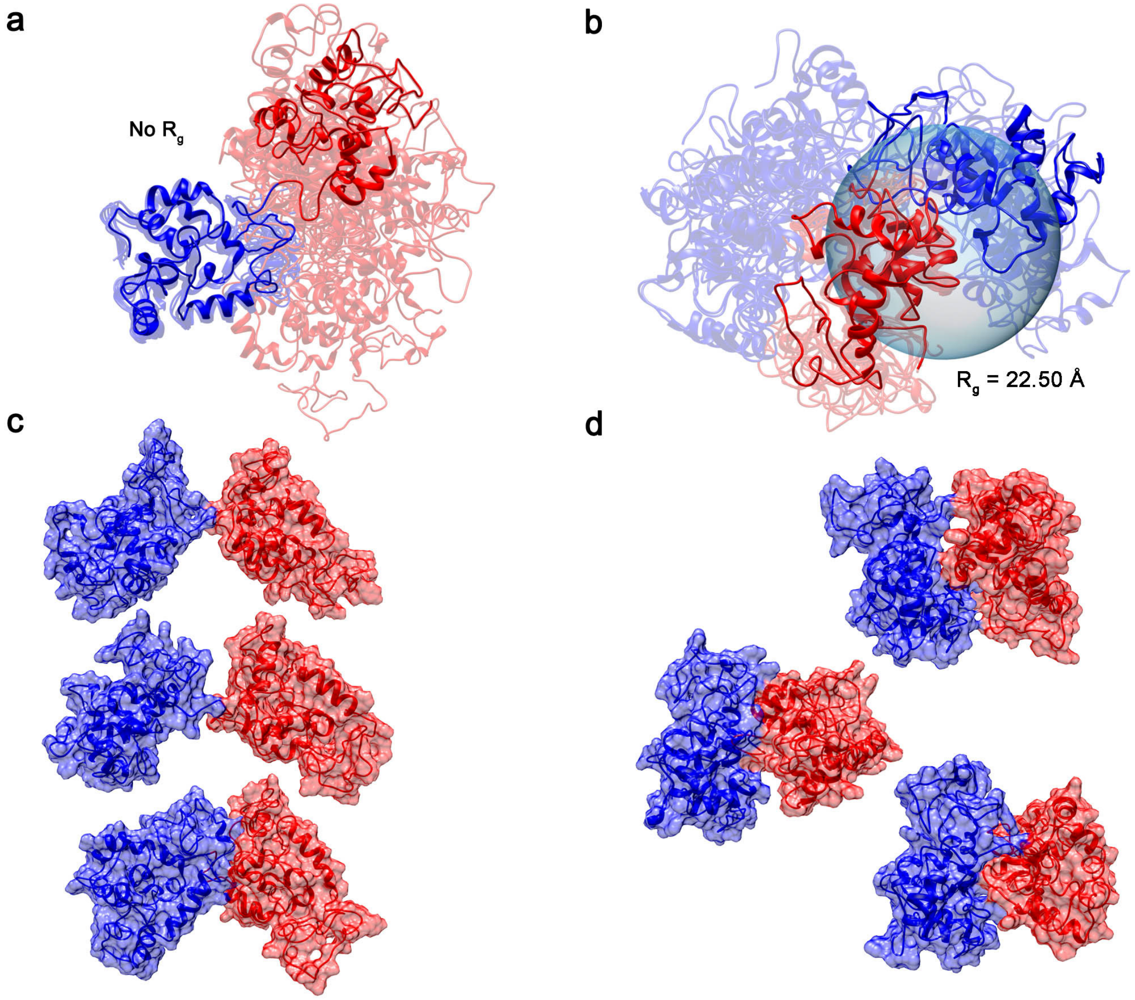


Figure 39 | The effects of adding a radius of gyration restraint during W_2 structure calculation. (a) 10 lowest energy structural ensemble (semi-transparent colouring) underlying the lowest energy structure (solid colouring). (b) 10 representative structures (semi-transparent) with a representative structure having the R_g closest to that determined by DOSY NMR for W_2 (solid colouring). The sphere represents the R_g. (c-d). Three representative structures with surface representation either without R_g (c) and with R_g (d). For a-d, blue represents W_{2-1} and red W_{2-2}.

4.5. EXTRAPOLATING W_2 TO A HIGHER NUMBER OF REPEATS

W_1, W_2, and W_3 were evaluated in the same experimental conditions and concentration. 1D ^{1}H-NMR experiments strongly indicate that W_3 has a similar atomic-level arrangement and, hence, “beads-on-a-string” architecture to W_2 (Figure 40a). Specifically, the spectra are indistinguishable with the exception of intensity, which increases linearly with increasing W repeat units. As demonstrated by the zoom of the amide region of W_{1-3} (Figure 40a), it is evident that the W_3 1D proton spectrum perfectly overlays those of W_2 and W_1. ^{1}H-^{15}N HSQCs were not collected for W_n greater than W_2, primarily due to the poor stability of W_3 and W_4 in solution; both aggregate within a few hours at protein concentrations amenable to multidimensional NMR. W_4 was not even sufficiently stable in soluble form for a high-quality 1D ^{1}H-NMR spectrum, with fibre formation evident in the NMR tube during the experiment (Figure 41b). Nevertheless, W_4 shows excellent recapitulation of the key spectral features through the amide proton region (Figure 41a) providing further evidence that the domains behave independently of each other even at higher W repeats.

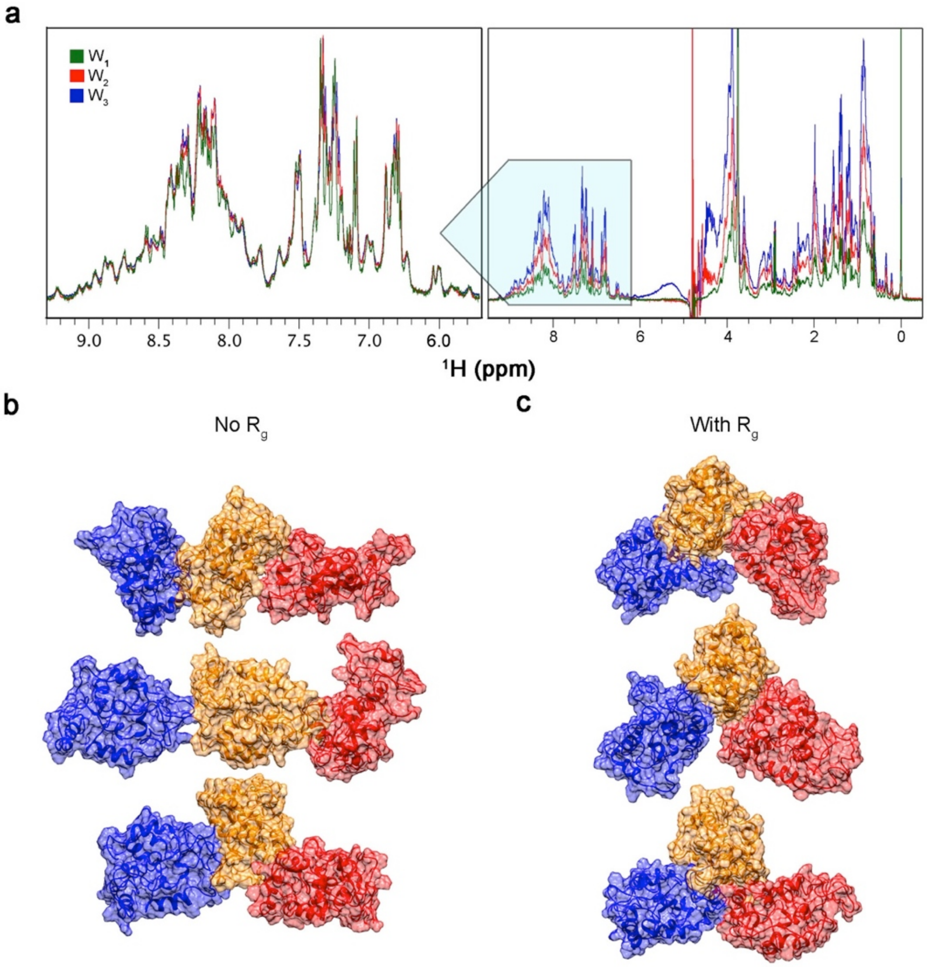


Figure 40 | W_3 follows the same trend as W_2. (a) Overlay of the 1D ^{1}H-NMR spectra for equimolar samples of indicated W_n sample. The intensity of the full spectral overlay on the right is not scaled, while the zoom of the amide region on the left is normalized. (b) Three randomly selected W_3 structures from the 20 lowest energy members without the addition of R_g. (c) W_3 structures with their R_g closest to the calculated value from experimental data.

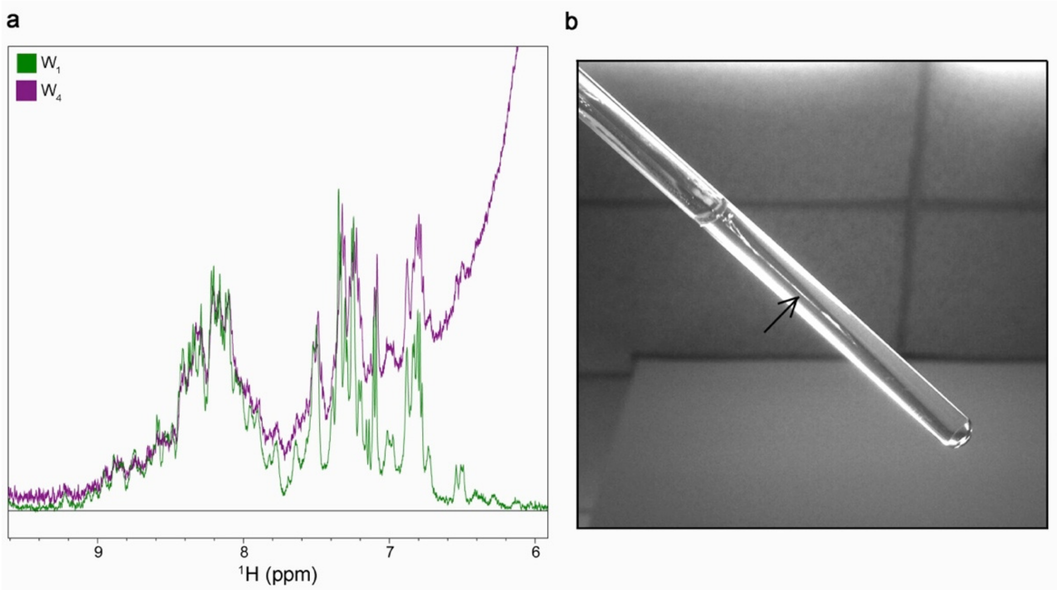


Figure 41 | (a) Overlay of W_1 and W_4 in the amide region. (b) Photograph of a W_4 fibre formed in solution (arrow) during an NMR experiment.

W_3 structure ensembles were calculated with and without R_g. The NMR statistics are outlined in Table 15. As with W_2, the structures calculated without the presence of an R_g restraint sampled a greater distribution of conformations due to the flexible linker and lack of restraints between domains. But, when the R_g was included as a restraint, the folded domains were compacted, creating a tighter structure. This fact is clearly visible from a side-by-side comparison of 3 representative ensemble members each from the two conditions (Figure 40b) although much less visible looking at the entire ensemble (Figure 42). Interestingly, the R_g in the case of W_3 had a greater impact on the agreement of the structural ensemble to the experimentally observed D_C than in W_1 and W_2 with an ensemble average D_C of 5.78 ± 0.004 $x10^{-11}$ m^2/s giving 0.3% divergence to the experimental D_C of 5.77 $x10^{-11}$ m^2/s. Without the use of an R_g restraint, the ensemble D_C was 5.46 ± 0.009 $x10^{-11}$ m^2/s, or a 5.3% deviation.

Unlike W_2 and W_3 structural ensembles, the incorporation of an R_g restraint for W_1 actually led to decreased agreement between the experimental D_C and the structural ensemble. The DOSY-derived D_C was $9.83x10^{-11}$ m^2/s and the predicted D_C for the structures calculated without an R_g was 9.26 ± 0.006 $x10^{-11}$ m^2/s (5.8% difference). Adding an R_g to the structure calculations resulted in an average of 9.15 ± 0.006 $x10^{-11}$ m^2/s for the ensemble, giving rise to a 7% difference. This is likely due to the fact that the flexible portions of W_1 (residues 1-11, 150-199) were already compactly tucked in around the globular domain rather then freely flowing as an extended "arms" in solution. The value of using an R_g in structure calculations would likely arise with linear proteins, like AcSp1, where flexible linkers give rise to many structural conformations with R_g serving to help restrict conformational variability.

As a whole, it is evident that W_2 and W_3 structures in solution are not linear; rather, these tend to coil in on themselves. It is difficult at this point to extrapolate to any construct larger than W_4 due to the lack of experimental data. Given that the D_C and R_g of W_{1-3} all imply that AcSp1 travels in solution as a more compact, globular-like structure rather than linear, this trend will likely extend to the larger W_n. The solution structures presented herein provide an indication to the pre-fibrillogenesis conformation and likely before protein-protein association according to the diffusion data all other supporting NMR data (particularly relaxation analysis). Since fibres can be produced from my NMR samples, the explanation for inter-protein interactions, which is not evident from the NMR experiments performed to date, is likely the result of a simple phenomenon shifting some physical equilibrium between monomers and oligomeric form, such as shear forces and protein dehydration (Hardy et al. 2008; Askarieh et al. 2010). Without this phenomenon, the population of monomeric AcSp1 available for protein-protein interaction and fibrillogenesis is too scarce.

The structures that are presented in this chapter serve as a starting point towards solving a more refined structure of W_2 and W_3 in solution. Further experiments would be required to refine the structure and confirm the observed compactness. It would be possible with FRET to experimentally determine the distance of the W domains by strategically attaching fluorophores on each domain. Since W lacks both lysines and cysteines, both amino acids could be included by strategic mutation of residues in the loops of W_2: a lysine on the first W and a cysteine on the 2nd W. In this manner, fluorophores can selectively be attached through derivatization.

Concomitantly, the same derivatization principle could be applied to paramagnetic resonance experiments (PRE). The advantage of PREs is that only one label is needed and this paramagnetic species, covalently attached to the protein, would attenuate the signals of nearby residues. By selectively labeling one domain with ^{15}N and the other with ^{13}C and ^{15}N labels, and using isotopically discriminated (IDIS) NMR experiments, the proximity of the domains could theoretically be determined. Obtaining side chain information on both domains in this case would be highly valuable.

4.6. SUMMARY

The use of inteins in NMR has not been widely applied thus far (Volkmann and Iwaï 2010, Skrisovska et al. 2010), but with the advent of studying larger complexes by NMR, the need to reduce spectral complexity will likely render segmental labeling methods increasingly routine. In this chapter, I demonstrated the power of split-intein *trans*-splicing technology (Wu et al. 1998) for NMR isotope labeling. With the two fibre forming W_2 constructs, $W_{2\text{-}1}$ and $W_{2\text{-}2}$, I was able to demonstrate that the repeated 200 amino acid sequence actually consist of a pair of individual, non-interacting domains separated by a flexible linker. This type of architecture best resembles a "beads-on-a-string" where the "beads" consist of the folded domain and the "string" is the linker. As mentioned earlier, W_2 readily forms fibres in solution, even from our NMR sample! The structures that are presented in this chapter are therefore representative, functionally relevant structures of recombinant AcSp1 (Figure 42). By observation, Dr. Lingling Xu and I determined that W_2 fibrillogenesis is controlled by a delicate equilibrium between the soluble form presented in this chapter, its aggregates, which originate over time in absence of shear forces, and fibres formed under shear forces. The cause of the domain-domain interactions responsible for driving fibre formation are still unknown at this time and can only be speculated upon.

In addition, I was able to probe the degree of W_n (n=2 and 3) protein linearity or curvature by observing the effects of the addition of a radius of gyration restraint during simulated annealing calculations. This resulted in a more overall compact structure, providing a better agreement to the experimental translational diffusion coefficients obtained from DOSY-NMR experiments. As a result, I was able to produce more reliable structural models of both W_2 and W_3 in a manner readily extendable to W_n. Future work could be aimed at refining those higher-order W models using additional inter-domain restraints, such as residual dipolar couplings to orient one domain relative to the other (See Appendix C). In addition, combining molecular dynamics simulations with the experimental NMR restraints presented herein could potentially improve refinement of these structures and lead to a more accurate model of a larger AcSp1 native structure.

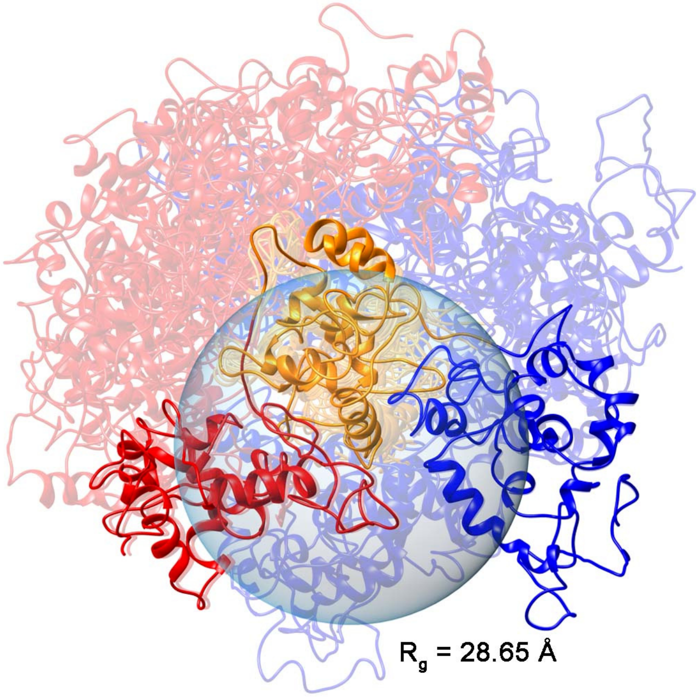


Figure 42 | 10 member structural ensembles (semi-transparent colouring) underlying representative structure (solid colouring) with R_g closest to that determined by DOSY NMR for W_3. Ensembles were calculated using concatenated sets of W_1 NMR restraints with an R_g restraint estimated from the observed DOSY-derived D_C.

CHAPTER 5. BACKBONE MOTIONS OF W_1 AND W_2

5.1. INTRODUCTION TO RELAXATION ANALYSIS

5.1.1. Proteins In Motion

It is now well established that proteins are not simply well-defined three-dimensional structures but are, instead, adaptive to their surroundings and can adopt many functionally relevant conformations (Fuxreiter and Tompa 2012). Although determination of protein structure is clearly valuable, the reality is that most proteins exist as a complex ensemble of conformers that continuously interconvert with only a subset of these conformations being functionally relevant (Jarymowycz and Stone 2006). The populations of states that exist over time are dependent on their relative thermodynamic stability and the rates at which they interconvert. With NMR, we can determine, at the atomic level, the motion within a protein at a wide variety of time scales (d'Auvergne and Gooley 2003) (Figure 43).

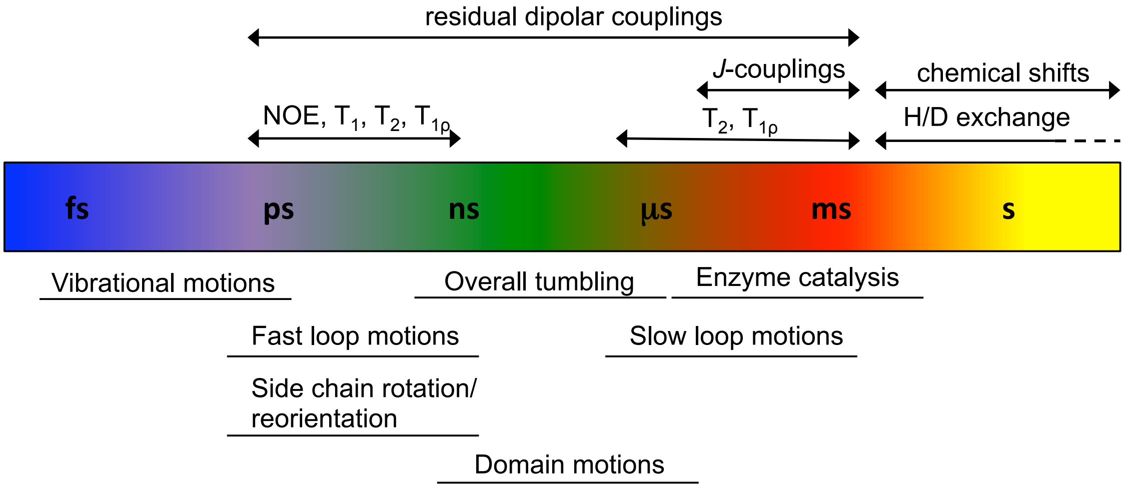


Figure 43 | Relative time scale of protein dynamics examinable by NMR. Above the time bar are the measurable NMR parameters and below are the types of protein motion. *This figure was reproduced from slides obtained during NMR Bootcamp 2010 in Birmingham, UK.*

5.1.2. NMR Relaxation

NMR relaxation is the process by which the nuclear spins in a system return to thermal equilibrium (Keeler 2002). The spin population at each energy level is determined by the Boltzmann distribution at equilibrium. At equilibrium, there is no transverse magnetization, or more generally, no coherence in the system. The application of a pulse in an NMR experiment breaks this equilibrium, producing a state, which leads to the spins undergoing a time-dependent re-establishment of equilibrium through relaxation processes.

Motion of and within a protein (or any NMR active molecule) affects experimental parameters in a manner that depends upon whether the process in question is faster (ps-ns) or slower (μs-ms) than the rotational correlation time (Keeler 2002). Relaxation measurements can therefore be used to characterize motion by determining the relaxation rate constants unique to each nuclear spin in the system. These rates are dependent on the architecture and physical properties of the protein.

Two types of relaxation processes act to restore Boltzmann equilibrium: 1) longitudinal relaxation is the rate that describes the restoration of the net longitudinal magnetization observed at the Boltzmann equilibrium and is characterized by a rate constant R_1 (1/T_1); 2) transverse relaxation refers to the system's loss of coherence over time due to local field fluctuations and is characterized by the rate constant R_2 (1/T_1) (Reddy and Rainey 2010). Molecular correlation time and magnetic fields both effect T_1 and T_2, with the magnetic field having a much greater effect on T_1 (Figure 44). Relaxation rates are measured by recording spectra with variable delay length for auto- or cross-relaxation of the selected coherences. The intensities of the peaks at all delays are then fitted by a mono-exponential curve:

$$I(t) = I_o e^{\left(-t/T_1\right)} \tag{5.1}$$

where t is the delay time, $I(t)$ is the intensity of the peak at time t, I_0 is the original intensity, and T_1 (or T_2) is the relaxation rate.

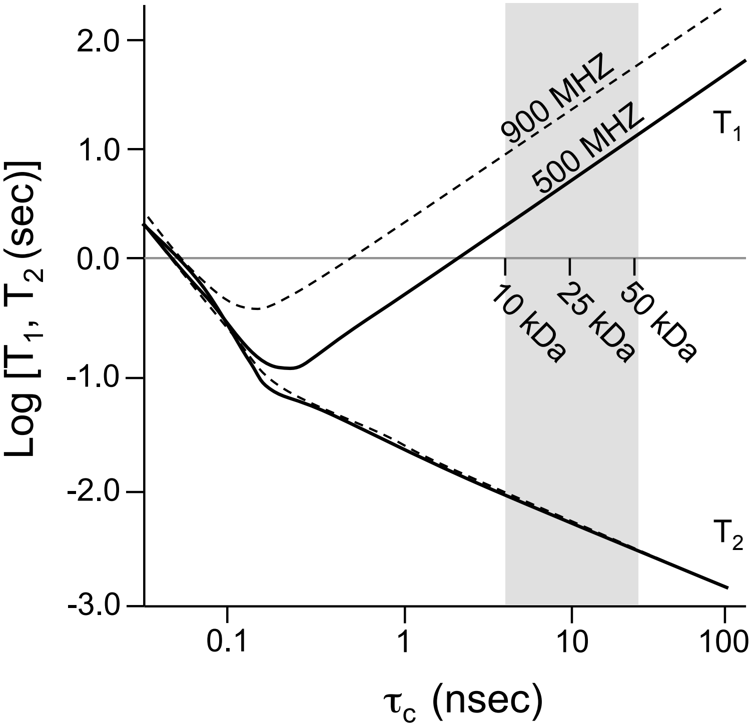


Figure 44 | Effects of molecular weight and field strength on proton relaxation. The graphical representation shows the dependence of T_1 and T_2 relaxation times on rotational correlation time τ_c, assuming a spherical protein. The dotted curves represent data computed at ~900 MHz and the solid line represent data at 500 MHz. The gray rectangle represents the molecular weights typically expected for routine analysis of proteins by NMR. *This figure was reproduced from Rule and Hitchens* (Rule and Hitchens 2006)*, pg 330.*

A third relaxation parameter, typically examined in biomolecular NMR relaxation experiments, is the [^{1}H]-^{15}N heteronuclear NOE (hetNOE) (Gust et al. 1975). The hetNOE equals the ratio of steady state ^{15}N signal intensities recorded with and without ^{1}H saturation (I_{sat}/I_{ref}) (Ishima 2012). The NOE enhancement factor (hetNOE) is defined as the irradiation of a spin population *I*, causing the nearby-irradiated spins *S* (*I* and *S* being two types of nuclei) to change *via* a dipole-dipole mechanism and is dependent on the relaxation rate constants R_1 and R_2 (Keeler 2002). The magnetization of spin *I* is equal to its equilibrium value during a zero mixing time, but at longer mixing time, spin *I* magnetization additionally, proportionally contributes to the mixing time and the cross-relaxation rate σ_{IS} of spins *I* and *S*. The change in intensity of spin *I* is the NOE enhancement. A positive enhancement of typically >0.65 signifies order while smaller or negative enhancements are indicative of backbone disorder. Together, the R_1, R_2, and hetNOE relaxation rates characterize motions in a protein or small molecule.

Although relaxation measurements can be made with any type of NMR active nucleus (Jarymowycz and Stone 2006), only ^{15}N relaxation in the context of *J*-coupled ^{1}H-^{15}N spin pairs will be presented herein. Measurement of R_1, R_2 and hetNOE as a function of amino acid residue allow describing the intramolecular motion at ps-ns timescales (Figure 43) at the backbone level for each non-proline peptide bond in the protein. The ^{1}H-^{15}N bond has been extensively characterized and therefore serves as a good model for the description of motions in proteins (Jarymowycz and Stone 2006; Kadeřávek et al. 2014).

Only two mechanisms significantly contribute to the ^{15}N relaxation, namely the chemical shift anisotropy of the ^{15}N nucleus and the dipole–dipole interaction between ^{15}N and ^{1}H. Molecular motions are characterized by a time dependent correlation function ($G(\tau)$), which measures the particular arrangement or rearrangement (longer time scales) of spins in a sample (Keeler 2002). The correlation function, just like the free induction decay (FID), can be Fourier transformed producing the frequency domain spectral density function. The spectral density function provides a measure of the relative amount of motion present as a function of frequency (Keeler 2002). The measured relaxation parameters are defined by the spectral density function $J(\omega)$ at five frequencies: ω_o, ω_N, ω_H, (ω_H - ω_N), and (ω_H + ω_N) (Viles et al. 2001; Wen et al. 2010). The relation between the relaxation rates of the ^{1}H-^{15}N bond and the spectral densities applied herein are as follows (Kay et al. 1989):

$$R_1 = \frac{1}{4}d^2\left[3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)\right] + c^2 J(\omega_N) \tag{5.2}$$

$$\begin{aligned} R_2 = {} & \frac{1}{8}d^2\left[4J(0) + 3J(\omega_N) + J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\right] \\ & + \frac{c^2}{6}[4J(0) + 3J(\omega_N)] + \mathrm{R_{ex}} \end{aligned} \tag{5.3}$$

$$NOE = 1 + \frac{d^2}{4R_1}\left(\frac{\gamma_H}{\gamma_N}\right)\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right] \tag{5.4}$$

where *d* and *c* are the dipolar and chemical shift anisotropy coefficients, where

$d = \left(\frac{\mu_0 h \gamma_H \gamma_N}{8\pi^2} \right) \left\langle \frac{1}{r_{HN}^3} \right\rangle$ and $c = \frac{\omega_X}{\sqrt{3}} (\sigma_{//} - \sigma_\perp)$, '$J$' is the spectral density at the indicated frequency, ω_N and ω_H are the Larmor frequencies of the ^{15}N and ^{1}H nuclei respectively, γ_H and γ_N are their corresponding gyromagnetic ratios, μ_0 is the permeability of free space, h is Plank's constant, r_{NH} is the length of the amide bond vector, R_{ex} is the conformational exchange contribution between states, and $\sigma_{//}$ and $\sigma_\perp$ are the parallel and perpendicular components of the (assumed to be axial) chemical shift tensor (Jarymowycz and Stone 2006).

5.1.3. Model-Free Analysis Of Relaxation Data

Once the relaxation parameters are determined for each backbone amide, these may be fit using a model describing the internal motions of a protein. This fitting is popularly performed through model-free analysis (Lipari and Szabo 1982) or reduced spectral density mapping (Farrow et al. 1995; Ropars et al. 2007).The Lipari-Szabo model-free analysis is the most popular formalism for analysis of localized dynamics of well-ordered proteins. In this treatment, the spectral density is written in terms of an order parameter (S^2) that is indicative of the degree of spatial restriction for a given inter-nuclear bond vector (^{1}H-^{15}N in this case). Each bond may be treated with either one or two correlation times: τ_e, the effective correlation time for fast internal motions, and τ_m, the slower, isotropic correlation time, assuming that there is a single, overall molecular rotation correlation time τ_c, dependent on solvent viscosity and protein shape and size (d'Auvergne and Gooley 2008). The overall rotation correlation time, τ_c, is defined as (Clore et al. 1990a; Jarymowycz and Stone 2006):

$$\tau_c^{-1} = \tau_m^{-1} + \tau_e^{-1} \qquad (5.5)$$

The model-free formalism was extended by Clore and coworkers (Clore et al. 1990a; Clore et al. 1990b), to allow for internal motions occurring at fast and slow time scales, each with a corresponding order parameter combined as $S^2 = S_s^2 S_f^2$, where the s and f subscripts designate slow and fast motions, respectively. An additional term used in model-free analysis is R_{ex}, which is the term used to describe relaxation due to chemical exchange and is included to account for motions on micro to millisecond timescale (d'Auvergne and Gooley 2003). Several versions of the model-free formalism have been adopted and differentiated by their assumptions and parameters:

Model 1 is referred to as the simplified model free formalism, with S^2 being the only parameter that needs to be optimized using equation 5.6 (Lipari and Szabo 1982):

$$J(\omega)=\frac{2}{5}\left(\frac{S^2\tau_m}{1+(\omega\tau_m)^2}\right) \tag{5.6}$$

Model 2 is the original model free formalism, with S^2 and τ_e optimized using equation 5.7 under the assumptions that $\tau_e < 500$ ps and $R_{ex} \approx 0$:

$$J(\omega)=\frac{2}{5}\left(\frac{S^2\tau_m}{1+(\omega\tau_m)^2}+\frac{(1-S^2)\tau}{1+(\omega\tau)^2}\right) \tag{5.7}$$

Model 3 and 4 use equations 5.6 and 5.7 of models 1 and 2 respectively but with the addition of a non-zero R_{ex} term (Lipari and Szabo 1982; Clore et al. 1990b).

Model 5 is the extended model-free formalism, with optimization of S^2 , S_f^2 , and $\tau_s^{'}$ under the assumptions that $\tau_f \ll \tau_m$, $\tau_s \geq 500$ ps, and $R_{ex} \approx 0$ using equation 5.8 (Jarymowycz and Stone 2006).

$$J(\omega)=\frac{2}{5}\left(\frac{S^2\tau_m}{1+(\omega\tau_m)^2}+\frac{(S_f^2)\tau_s^{'}}{1+(\omega\tau_s^{'})^2}\right) \tag{5.8}$$

Model-free analysis consists of fitting the relaxation data by minimizing the χ^2 function (5.9) for models 1-5 (d'Auvergne and Gooley 2003):

$$\chi^2=\sum_{i-1}^{N}\sum_{j-1}^{M}\left[\left(R_{ij}-\hat{R}_{ij}\right)^2 \Big/ \sigma_{ij}^2\right] \tag{5.9}$$

where N is the total number of spins, M is the number of experimental parameters, R_{ij} is the j^{th} experimental parameter, $\hat{R}_{ij}$ is the j^{th} theoretical relaxation parameter calculated from the putative values of the dynamic parameters, and σ_{ij} is the experimental uncertainty in the j^{th} relaxation parameter (Jarymowycz and Stone 2006).

The overall goal of the model-free approach is to describe the motion of each amide bond under some optimal overall rotational correlation time ($\tau_c \approx \tau_m$) and a generalized order parameter S^2. The amplitude of motion reflected in S^2 can take a value between 0 (completely disordered bond vector) and 1 (static bond vector). The approach of this analysis is to model the data with Model 1, adding more parameters and moving progressively toward Model 5. The model is deemed appropriate when most of the residues in the protein satisfy the statistical approach of Akaike's information criteria (AIC), where AIC = χ^2 +2*k*, where *k* is the number of parameters in the model (d'Auvergne and Gooley 2003; Spyracopoulos 2006).

5.1.4. Reduced Spectral Density Mapping

Reduced spectral density mapping allows characterization of the degree of motion of each ^{1}H-^{15}N bond at the amplitude at the zero-frequency motion reflective of the total internal motions, ω_N, the amplitude of motion at the ^{15}N frequency, and (0.87)ω_H, amplitudes of motion at 87% of the ^{1}H frequency (Viles et al. 2001; Reddy and Rainey 2010). Due to the magnitude of difference in gyromagnetic ratio between ^{1}H (γ_H= 267 rad/sT) and ^{15}N (γ_H=-27.2 rad/sT) (Lide 2000), the high frequency values of $J(\omega_H - \omega_N)$ and $J(\omega_H + \omega_N)$ are treated as being approximately equivalent to $J(0.87\omega_H)$ (Farrow et al. 1997; Wen et al. 2010). The relaxation rates R$_1$, R$_2$, and the hetNOE are parameterized as:

$$R_1 = \frac{1}{4}d^2\left[3J(\omega_N) + 7J(0.87\omega_H)\right] + c^2 J(\omega_N) \tag{5.10}$$

$$R_2 = \frac{1}{8}d^2\left[4J(0) + 3J(\omega_N) + 13J(0.87\omega_H)\right] + \frac{1}{6}c^2\left[4J(0) + 3J(\omega_N)\right] \tag{5.11}$$

$$NOE = 1 + \frac{1}{4}T_1 d^2\left(\frac{\gamma_H}{\gamma_N R_1}\right)\left[5J(0.87\omega_H)\right] \tag{5.12}$$

and incorporated into the two field-dependent reduced spectral densities $J(0.87\ \omega_H)$ and $J(\omega_N)$ and one field-independent reduced spectral density $J(0)$:

$$J(0) = \frac{1}{3d^2 + 4c^2}\left(6R_2 - R_1\left(3 + \frac{18}{5}(\gamma_N/\gamma_H)(NOE - 1)\right)\right) \tag{5.13}$$

$$J(\omega_N) = \frac{4}{3d^2 + 4c^2}\left(R_1\left(1 - \frac{7}{5}(\gamma_N/\gamma_H)(NOE - 1)\right)\right) \tag{5.14}$$

$$J(0.87\omega_H) = \frac{4}{5d^2}\left(R_1\left((\gamma_N/\gamma_H)(NOE - 1)\right)\right) \tag{5.15}$$

The motions of the ^{1}H-^{15}N bond are reflected in $J(0)$; the smaller the value of $J(0)$, the greater the sub-nanosecond flexibility of the ^{1}H-^{15}N bond (Viles et al. 2001) (Figure 45). Slow micro- to millisecond motions reflected in R_{ex}, obtained in model-free analysis (Lipari and Szabo 1982) may also be apparent in $J(0)$. R_{ex} doesn't affect $J(0.87\omega_H)$ or $J(\omega_N)$ (Twomey et al. 2012). The spectral density at $J(\omega_N)$ and $J(0.87\omega_H)$ reflect fast picosecond motions at high frequency; $J(\omega_N)$ and $J(0.87\omega_H)$ will decrease with increased dynamics of the ^{1}H-^{15}N bond. $J(0.87\omega_H)$ is directly proportional to the cross-relaxation rate constant (hetNOE), whereas $J(\omega_N)$ is dominant for the longitudinal relaxation rate R_1 and $J(0)$ is dominant for the transverse relaxation rate R_2 (Keeler 2002; Twomey et al. 2012).

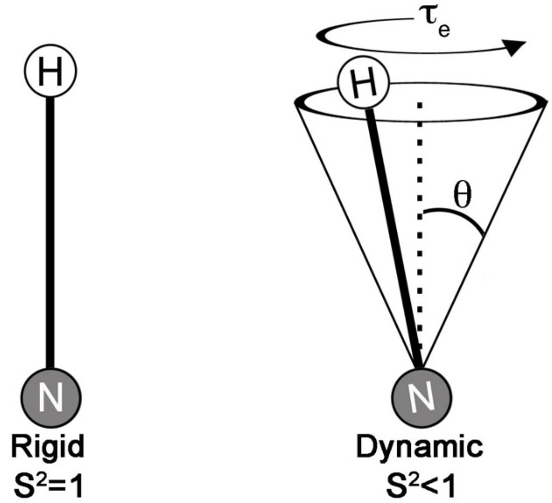


Figure 45 | Illustration of the ^{1}H-^{15}N bond motions, its relation to $\tau_{e,}$, and its relation to the order parameter.

Reduced spectral density mapping is often applied when model-free analysis fails to converge with a global motional model with coherent parameters. This is frequently the case with proteins containing intrinsically disordered regions (IDR) (Kadeřávek et al. 2014). It should also be noted, though, that caution must be used when applying the reduced spectral density mapping approach with proteins containing IDRs because the high-frequency values of the spectral density have a tendency to have increased contribution to ^{15}N relaxation in IDR in comparison to well-ordered segments. Also, the peak widths in IDR are narrower and more intense than ordered regions, leading to more precise data in IDR. Finally, anisotropic molecular motions are more difficult to assess using this model of motion.

5.1.5. Project Goal

Molecular motion were characterized for the ^{15}N-^{1}H bonds within W_1, W_{2-1}, and W_{2-2} at 16.4 T and W_1 at 11.7 T using relaxation analysis. R_1 and R_2 relaxation rates and the hetNOE were determined as a function of residue and analyzed using both the model-free formalism and reduced spectral density mapping. The model-free approach failed to converge to a global motional model for W_1 and W_2. Reduced spectral density mapping revealed that the repetitive W units displayed identical trends in motions. The linker regions show a higher degree of sampling of faster motions than the folded domains, strongly supporting the "beads-on-a-string" AcSp1 architecture discussed in chapter 4.

5.2. METHODS

5.2.1. NMR Spectroscopy

Uniformly ^{15}N-enriched W_1 (~0.2 mM), and segmentally ^{15}N-enriched W_{2-1} and W_{2-2} (~0.2 mM) NMR samples were prepared in sodium acetate buffer (20 mM d_3-acetate in H_2O:D_2O at 90%:10% (v/v), 1 mM NaN_3, 1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid; pH 5). Experiments were carried out at 30.0°C on either a Bruker Avance III operating at 16.4 T (Bruker Canada, Milton, ON) equipped with a 5 mm indirect detection TCI cryoprobe or at 11.7 T Avance II spectrometer (Bruker Canada, Milton, ON) equipped with a 5 mm broadband observed (BBO) Smartprobe. The experimental details and delays used for relaxation are summarized in Table 16. Data sets were processed using NMRpipe (Delaglio et al. 1995) and imported into CcpNMR Analysis 2.3 (Vranken et al. 2005) for peak analysis.

Table 16 | Summary of NMR relaxation experiments for W_1, $W_{2\text{-}1}$, and $W_{2\text{-}2}$

Experiment	Bruker pulse sequence	Scans	Recycle Delay (s)	Increments	Spectral width (ppm)	Center (ppm)	Delay (ms)	^{1}H Frequency (MHz)
W1				1H | 15N	1H | 15N	1H | 15N		
T1	hsqct1etf3gpsi	24	1.5	4096|96	16|23	4.70|115.5	50, 100, 250, 500, 750, 1000, 1300, 1700	500
T2	hsqct2etf3gpsi	20	1.5	4096|96	16|23	4.70|115.5	17, 34, 51, 85, 119, 152, 187, 238	500
NOE	hsqcnoef3gpsi	58	5	4096|192	16|23	4.70|115.5	-	500
T1	hsqct1etf3gpsi	16	1.5	2048|192	16|23	4.70|115.5	50, 100, 250, 500, 750, 1000, 1300, 1700	700
T2	hsqct2etf3gpsi	16	1.5	2048|192	16|23	4.70|115.5	17, 34, 51, 85, 119, 152, 187, 238	700
NOE	hsqcnoef3gpsi	32	5	2048|256	16|23	4.70|115.5	-	700
W21								
T1	hsqct1etf3gpsi	16	1.75	2048|128	16|23	4.70|115.5	50, 100, 250, 500, 750, 1000, 1300, 1700	700
T2	hsqct2etf3gpsi	16	1.75	2048|128	10|23	4.70|115.5	17, 34, 51, 85, 119, 152, 187, 238	700
NOE	hsqcnoef3gpsi	32	5	2048|256	16|23	4.70|115.5		700
W22								
T1	hsqct1etf3gpsi	16	1.75	2048|128	16|23	4.70|115.5	50, 100, 250, 500, 750, 1000, 1300, 1700	700
T2	hsqct2etf3gpsi	16	1.75	2048|128	10|23	4.70|115.5	17, 34, 51, 85, 119, 152, 187, 238	700
NOE	hsqcnoef3gpsi	32	5	2048|128	16|23	4.70|115.5	-	700

5.2.2. Determination Of ^{15}N Relaxation Parameters

The R_1 and R_2 values with their errors were extracted using the *Mathematica* (Wolfram, version 8.0.4, Champaign, IL, USA) notebook '*Relaxation Decay*' developed by Dr. Leo Spyrcopoulos (Spyracopoulos 2006). Errors were estimated based on the average spectral noise. The [^{1}H]-^{15}N heteronuclear NOE enhancement factor was measured as the ratio of the intensities observed in a saturated spectrum to the reference spectrum for each residue, $I_{\text{sat}}/I_{\text{ref}}$, where I_{sat} and I_{ref} are the intensities of the peaks in the ^{1}H-^{15}N HSQC spectra, with and without proton saturation, respectively. Non-linear fits were used to minimize the statistical value of χ^2. The χ^2 goodness-of-fit test per residue was used and compared to the critical χ^2 (9.146) calculated over 100 Monte Carlo simulations for a single residue at a 95% confidence interval.

5.2.3. ^{15}N Relaxation Analysis

The ^{15}N R_1, R_2, and [^{1}H]-^{15}N heteronuclear NOE data were further analyzed using Dr. Spyracopoulos' suite of *Mathematica* notebooks for protein main chain relaxation analysis (Spyracopoulos 2006). Main chain amide T_1 (R_1 (s^{-1}) = $1/T_1$) and T_2 (R_2 (s^{-1}) = $1/T_2$) relaxation rates are determined from nonlinear least-square fits to a two-parameter mono-exponential decay (eq. 5.1). The axially symmetric diffusion tensor and the anisotropic overall rotational correlation time were determined using the *Diffusion* notebook from residues with a hetNOE > 0.65. A per-residue estimate of correlation time was calculated from T_1/T_2 ratios obtained from R_1 and R_2 measurements. Model-free analysis was completed using the *ModelFree* notebook (Spyracopoulos 2006) and the analysis based on the anisotropic diffusion tensor D_z and D_x obtained from the *Diffusion* notebook. The order parameter, S^2, was determined as a function of residue using the *ModelFree* notebook. Per-residue plots of $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ were generated using the *SpectralDensity* notebook.

5.3. DYNAMICS OF W_1 AND W_2

5.3.1. *The {^{1}H}-^{15}N NOE*

[^{1}H]-^{15}N heteronuclear NOE (hetNOE) measurements (Gust et al. 1975) show positive enhancement throughout the folded protein core and negative or minimal positive enhancement in the linker (Figure 25,Figure 39,and Figure 46), supporting the localization of folded vs. disordered domains observed in chapter 3 and 4 (Barbato et al. 1992; Eliezer et al. 1998). The similarity in hetNOE enhancement factors between both W domains in W_2 ($W_{2\text{-}1}$ and $W_{2\text{-}2}$) and in W_1 supports my combined chemical shift analysis results, presented in chapter 4 (Figure 38), demonstrating the independence between W domains in the ps-ns timeframe. The similarity between domains extends into the disordered C-terminal tail (W_1, $W_{2\text{-}2}$) or linker ($W_{2\text{-}1}$) where an increase in order is observed spanning residues 175-185 in both W_1 and W_2 due to the many prolines (6 of the 8 found in the W sequence) (Sequence: …^{171}GPSGPSTGPS ^{181}GYPGPLGGGA ^{191}PFGQSGFGG) (Figure 46). Also notably, increasingly positive hetNOE enhancement factors for residues 190-199 in $W_{2\text{-}1}$ and 1-10 in $W_{2\text{-}2}$ are indicative of attenuated ps-ns dynamics upon covalent linkage of the domains in W_2.

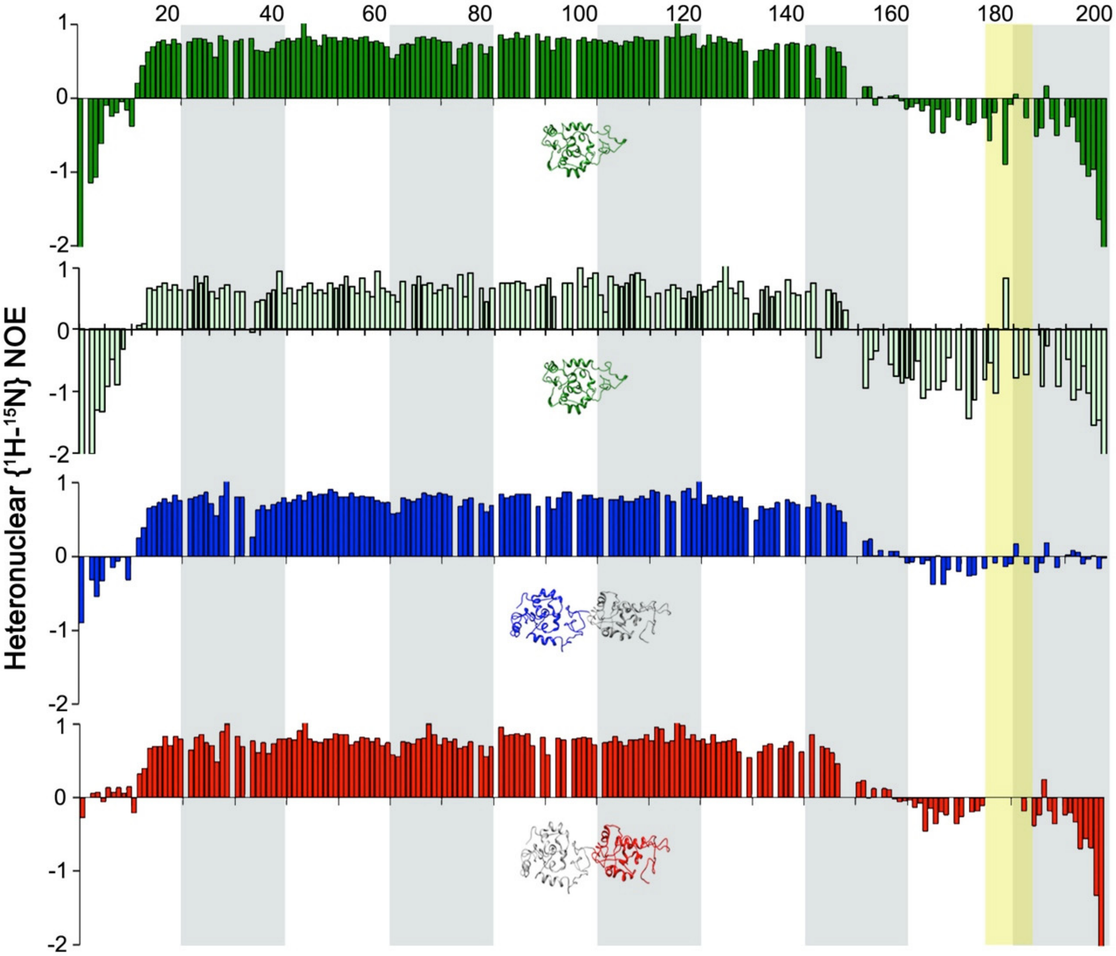


Figure 46 | ^{1}H-^{15}N hetNOE enhancement factors for W_1 (dark green), $W_{2\text{-}1}$ (blue), and $W_{2\text{-}2}$ (red) at 16.4 T and W_1 (light green) at 11.7 T. The coloured domain of the lowest energy structure represents the domain that is ^{15}N labeled. Residue numbers are located at the top, with $W_{2\text{-}2}$ residue numbering from 1-200 instead of 201-400. The yellow highlight represents the region of the C-terminal tail with slightly dampened in comparison to the remaining other ~40 residues.

5.3.2. Peak Assignment In W_1 For T_1/T_2 Relaxation Analysis

The NMR relaxation data were assigned with Analysis, exported, and properly formatted for use in the *Mathematica* notebooks. In all cases, peaks with closely overlapped intensities were omitted in the analysis: 12, 22, 31, 34, 74, 82, 89, 91, 94, 97, 103, 131, 137, 141, 145, 150, 157, 170, 177, 189 (+200 for $W_{2\text{-}2}$ residue numbering). The residues 10T, 48V, 49Q, 117L, 121V, 126G, 148S in W_1, $W_{2\text{-}1}$, and $W_{2\text{-}2}$ showed strong chemical exchange as the delay in T_2 increased but not in T_1 (Figure 48). The exchanging residues were omitted from model-free and reduced spectral density mapping analysis. It is quite possible that the samples are heterogeneous, where one peak is initially weaker but since it relaxes more slowly it becomes more evident at the longer T_2 delays. But, if the phenomenon corresponds to chemical exchange, than one could "simply" perform CPMG relaxation dispersion experiments (Carr and Purcell 1954; Meiboom and Gill 1958) to determine the exchange rate of this residue.

In addition, it should be noted that an issue was encountered with the W_1 data acquisition at 11.7 T for T_2 data, but oddly not the T_1 or hetNOE data. Specifically, all peaks were coupled in the ^{1}H direct dimension with a value of about ~10 Hz, which was possibly a result of slightly less-than-ideal tuning of the ^{15}N frequency (Figure 47). Although this situation is not ideal, the data were assignable, choosing only the left peak of the coupled T_2 peaks for assignment. The T_1 and T_2 data fit the exponential decay quite well and were used for subsequent analysis.

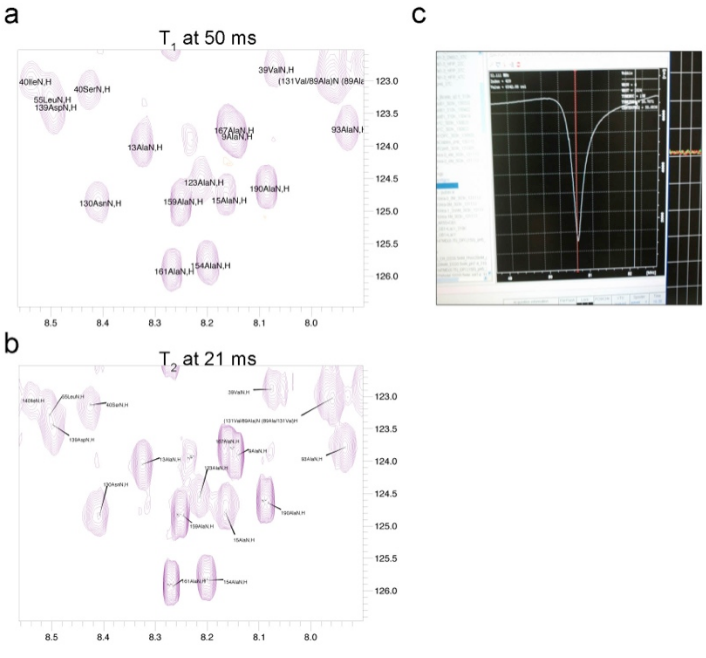


Figure 47 | ^{15}N tuning mismatch during the acquisition of the W_1 11.7 T relaxation data. (a) Normal T_1 data from the 50 ms delay spectrum. (b) The coupled T_2 data with a ~10Hz splitting, and (c) The mismatched tuning.

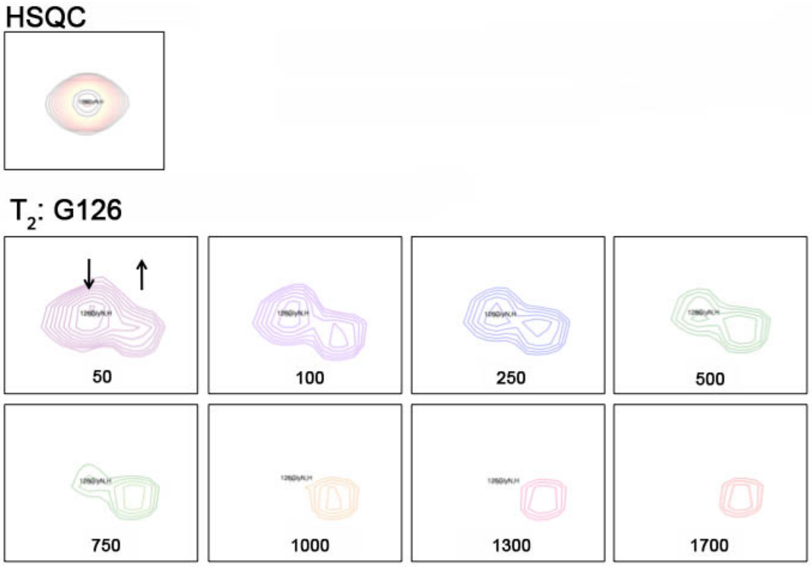


Figure 48 | Slow chemical exchange apparent in the T_2 data set for W_1 and W_2. Presented is Gly126, a nice example of the chemical exchange that is happening with certain residues in the domain of wrapping silk.

5.3.3. Determination Of W_1 And W_2 Relaxation Parameters

The folded domain (12-149) and the 'linker' region (residues 1-11 and 150-199) (Table 17 and Figure 50) of the W unit display very obvious segregation with regards to dynamics. The linker region of W_1 displays much slower transverse relaxation than the folded domain, which is more evident in R_2 ($1/T_2$) and T_1/T_2 at 11.7 and 16.4 T than in the R_1 relaxation (Figure 50). There is a difference between W_1 R_1 data gathered at 11.7 and 16.4 T: the rates are much slower at 11.7 T than at 16.4 T, which is expected given that R_1 ($1/T_1$) depend directly on field strength (Keeler 2002). Regions around residues ~10 and ~150 have the lowest T_1 values at 16.4 T in all W units (Figure 46). These regions were correspondingly observed to rapidly exchange with D_2O (Figure 23). T_1 is decreased slightly compared to all other residues in the protein, which is a feature that does not appear at 11.7 T for W_1. Instead, W_1 at 11.7 T demonstrates the difference in dynamics between the linker and the folded domain that is not portrayed at 16.4 T (Table 17).

Since T_2 is practically independent of field strength (Keeler 2002), the observation that T_2 at 11.7 T and 16.4 T show the same trend is expected (Figure 50). The different dynamics between the linker and the folded domain are clearly represented by T_2. S144 in W_1 at both field strengths, but not in W_2, has a longer T_2 compared to the other residues in the folded domain. Oddly, the S144 side chain is pointed towards the interior of W_1 and has NOE contacts to the side chain of T28 (Figure 49). Residue S180 also displays odd dynamics in both W_1 datasets compared to the adjacent amino acids. It is situated after P179 and three amino acids away from P183. Again, this phenomenon is not observed in W_2.

Both domains in W_2 exhibit practically identical flexibilities and structure; the similarities between the domains are quite remarkable. The folded domains in W_2 have slightly higher T_1 values compared to the linker and tails (Table 17 and Figure 50). Upon inspection of Figure 50, the decrease in dynamics of the proline-rich C-terminal region from 173-185 observed in Figure 46 is not reflected in T_2 but rather in T_1. But, unlike in W_1, residues G188 and G388 in W_2 exhibit shorter T_2 values than the remainder of the linker. The reason for this is not clear at the present time.

T_1/T_2 is proportional to τ_c (Lee et al. 2006). Overlaying T_1 and T_2 demonstrates the same trend for W_1, $W_{2\text{-}1}$, and $W_{2\text{-}2}$ without any striking difference between these labeled W units. An overly of the T_1/T_2 ratio at 16.4 T demonstrates that the W_2 repetitive units demonstrate very similar differences in dynamics between the linker and the folded domain, with a greater variation than that observed for W_1. This is reasonable because the smaller W_1 protein tumbles faster in solution in comparison to W_2. Also of note is the difference between W_1 T_1/T_2 at 11.7 T and 16.4 T. Although the N- and C-terminal regions are identical in intensity, the folded domain is not.

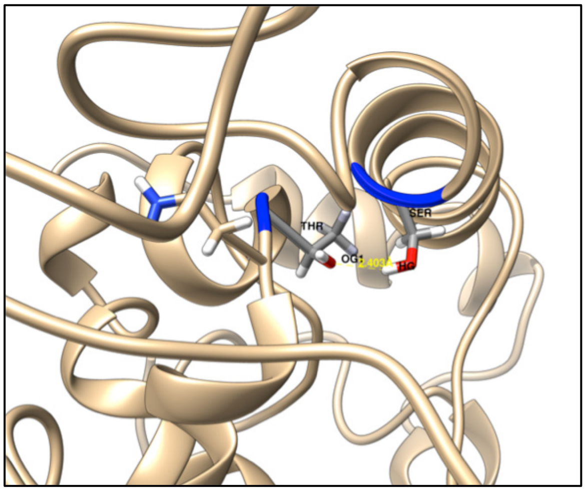


Figure 49 | Residue S144 in contact with T28. The top right helix is H5.

Table 17 | Average T_1, T_2, and T_1/T_2 relaxation of the overall W unit, for the linker, and the folded domain.

Frequency (MHz)		T_1			T_2		
		Total	Linker	Domain	Total	Linker	Domain
500	W_1	635.7 ± 157	848.0 ± 246	575.6 ± 51.7	210.8 ± 190	485.1 ± 188	118.8 ± 54.6
700	W_1	805.0 ± 117	832.9 ± 216	795.2 ± 44.9	156.2 ± 139	341.2 ± 162	90.7 ± 28.6
	W_{21}	813.8 ± 77.4	725.5 ± 77.4	837.6 ± 53.4	140.4 ± 116.6	319.2 ± 73.3	79.3 ± 41.4
	W_{22}	809.2 ± 82.1	716.2 ± 77.6	841.8 ± 78.0	138.3 ± 134	339.8 ± 109	71.09 ± 12.8

Frequency (MHz)		$T_1/T_2 \propto \tau_c$		
		Total	Linker	Domain
500	W_1	4.57 ± 1.9	1.99 ± 0.9	5.34 ± 1.4
700	W_1	7.75 ± 2.9	3.08 ± 2.1	8.86 ±1.8
	W_{21}	9.24 ± 4.3	2.47 ± 0.8	11.66 ± 1.6
	W_{22}	10.31 ± 4.9	2.31 ± 0.8	12.90 ± 1.8

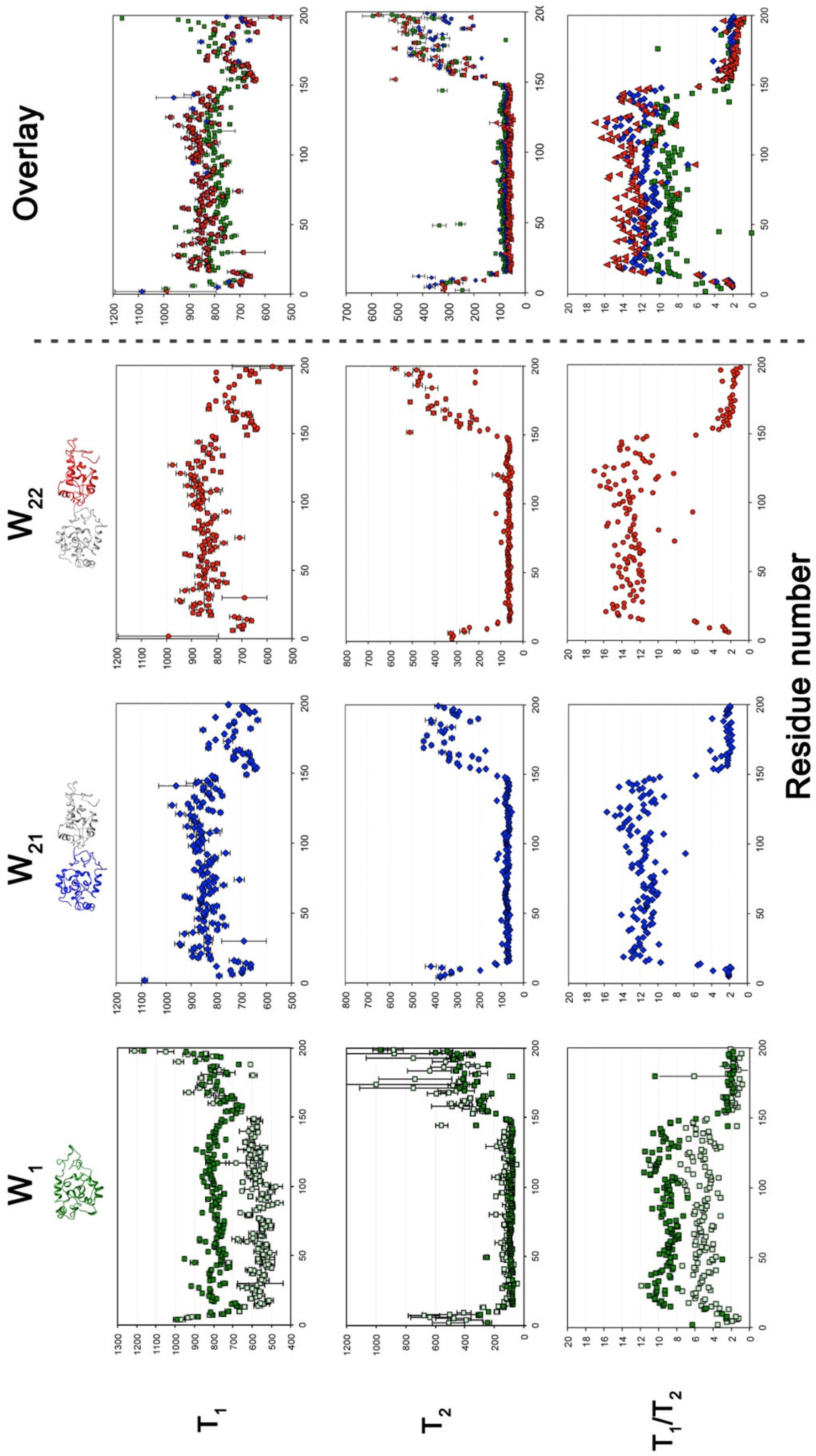


Figure 50 | Relaxation data summary from W_1 (dark green), $W_{2\text{-}1}$ (blue), and $W_{2\text{-}2}$ (red) and their overlay at 16.4 T. W_1 relaxation data at 11.7 T is in light green overlaid on top of the W_1 16.4 T data. The y-axis units are in sec.

5.3.4. Rotational Correlation Times For W_1, $W_{2\text{-}1}$, And $W_{2\text{-}2}$

The selected T_1, T_2, and the hetNOEs from residues fitting the chosen χ^2 statistical cut-off value in the relaxation analysis were used for the determination of the correlation times of W_1 and W_2. About 90 residues were employed for the W_1 diffusion calculation, with good T_1 and T_2 exponential fittings, and a hetNOE value of >0.65 (Yao et al. 1998). The diffusion model for W_1 was determined through Spyracopoulos' Mathematica *Diffusion* notebook (Spyracopoulos 2006). The rotational correlation time τ_c of W_1 was determined to be 7.9 ns with a $\chi^2 = 556$. The diffusion model was best described as axially symmetric, and the diffusion tensor parameters were estimated as $D_x = D_y = 2.03\times10^7$ m^2/s and $D_z = 2.3\times10^7$ m^2/s, with angles of $\theta = 43.3°$ Φ=4.41° at a $\chi^2 = 506.9$.

At 11.7 T, only 45/155 W_1 residues were employed for the determination of the anisotropic diffusion parameters with the *Diffusion* notebook (Spyracopoulos 2006). Again, only residues with a hetNOE > 0.65 for W_1 at 11.7 T were used for the calculation. The rotational correlation time τ_c of W_1 was determined to be 8.5 ns at a $\chi^2 = 203$. The diffusion model is best described as axially symmetric due to the lower χ^2, and its diffusion tensor parameters are $D_x = 1.8\times10^7$ m^2/s and $D_z = 2.3\times10^7$ m^2/s, with an angle of $\theta = 48°$ $\Phi = 30°$, with $\chi^2 = 166°$. The rotational correlation time of W_1 at 11.7 T differs from the correlation time at 16.4 T (7.9 ns) by 0.6 ns. W_1 can be described as oblate.

For both $W_{2\text{-}1}$ and $W_{2\text{-}2}$, ~90 residues with a hetNOE value > 0.65 (Yao et al. 1998) were employed for the calculation. The rotational correlation time τ_c of $W_{2\text{-}1}$ was determined to be 9.05 ns with a $\chi^2 = 1173$. The diffusion model, based on hydrodynamics and the overall shape of the protein (chapter 4), was best described as axially symmetric with tensor parameters $D_x = 1.9\times10^7$ m^2/s and $D_z = 1.7\times10^7$ m^2/s, and angles of $\theta = 63°$ $\Phi = 102°$, fitted at $\chi^2 = 1065$. The rotational correlation time τ_c of $W_{2\text{-}2}$ was determined to be 9.8 ns with a $\chi^2 = 1687$. The $W_{2\text{-}2}$ diffusion tensor parameters are $D_x = 1.8\times10^7$ m^2/s and $D_z = 1.6\times10^7$ m^2/s, with angles of $\theta = 76°$ $\Phi = 97°$, with $\chi^2 = 1565$. $W_{2\text{-}1}$ and $W_{2\text{-}2}$ are prolate.

The W_1 and W_2 rotational correlation times both imply that the W unit is non-spherical and moves much faster in solution than would be expected for a globular protein of the comparable molecular mass. Using Stoke's law along with the viscosity, the hydration radius, and the diffusion coefficient obtained from NMR diffusion experiments, one would expect τ_c to be ~ 24.5 ns, not 9 ns. A spherical W_1 was predicted to have a τ_c of 9.4 ns rather than ~8 ns.

5.3.5. Model-Free Analysis

The model-free formalism is one approach to infer internal motions of a protein from nuclear spin relaxation data. The most appropriate way to model the spectral density is highly dependent on whether the rotational diffusion tensor is isotropic or anisotropic. As detailed above, an anisotropic tensor was most appropriate for both W_2 and W_1. The order parameter S^2 (Figure 51) was obtained from the *Diffusion Mathematica* notebook (Spyracopoulos 2006). As detailed in section 5.1.3 and 5.1.4, S^2 quantitatively describes the amplitude of the internal motion on a ps-ns time scale. Residues with $S^2 > 0.65$ are mostly observed in the folded domains of W_2 and W_1. As expected, the linker and tail regions display larger amounts of disorder, correlated to an average S^2 value < 0.3.

Of all models, the extended model-free formalism, or model 5 (equation 5.7) (Clore et al. 1990a; Clore et al. 1990b), was the best fit, likely because it contains fast and slow motions, as reflected in the hetNOE, T_1, and T_2 data. But since only about half of the residues in either W_1 or W_2 fit model 5, it therefore cannot fully describe the overall motions of these proteins. Essentially, all three W_1, $W_{2\text{-}1}$, and $W_{2\text{-}2}$ data sets could not be fit in model-free analysis according to AIC value over all residues (Results presented in Appendix C). The assumption of statistical independence of global and local motions, which is the basis of model-free analysis of well-ordered proteins, is not necessarily valid for proteins containing long disordered segments. So, as a remedy, reduced spectral density mapping was explored next. This treatment is usually much more appropriate for proteins containing intrinsically disordered regions (Kadeřávek et al. 2014).

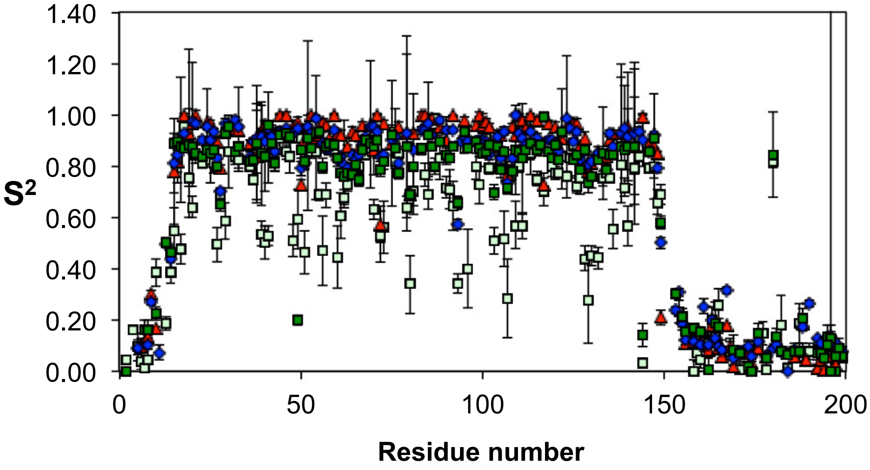


Figure 51 | The per-residue calculated S^2 values of W_1 at 11.7 T (light green) and W_1 (dark green), $W_{2\text{-}1}$ (blue), and $W_{2\text{-}2}$ (red) at 16.4 T obtained from the *Diffusion* notebook (Spyracopoulos 2006).

5.3.6. Reduced Spectral Density Mapping

For a protein containing intrinsically disordered regions, the spectral density is generally the better-represented interpretation for motion for each ^{15}N-^{1}H bond at 0, ^{15}N, and ^{1}H frequencies without knowing the nature of the global rotational diffusion (Bernadó et al. 2002; Twomey et al. 2012). As can be seen in Figure 52, the linker region samples a different frequency of motion than the folded core of the W unit, just as described in section 5.3.3.

The most interesting feature of the spectral density mapping analysis for all W units, more pronounced in W_2 than W_1, is uniformly elevated high frequency motion observed from residues 11-17 and 146-160. In the structures presented in chapters 3 and 4 (Figure 25, Figure 39, Figure 42), these two regions correspond to a portion of the linker and a portion of the first helical segment H1 (Figure 25). Residues 146-160 constitute part of H5 and the disordered tail. Hence, based solely on dynamics, in my opinion this region would be the most willing to adopt a different conformation and perhaps greatly helps in seeding protein-protein interaction for fibrillogenesis. Most intriguingly, this phenomenon is completely independent of protein concatenation.

Residues ~180-199 also seem to display greater differences in dynamics, but greatly increases or decreases from one residue to the next. Comparatively, W_1 $J(\omega_N)$ at 11.7 T differs from $J(\omega_N)$ at 16.4 T: the data at 11.7 T demonstrate higher values of spectral density throughout the folded core and lower values in the tails, in contrast to at 16.4 T which is fairly uniform throughout the protein.

$J(0.87\omega_H)$ and $J(0)$ both strongly illustrate the differential dynamics between the folded core and the linker, mirroring the behaviour of the individual T_1, T_2 and hetNOE parameters detailed in section 5.3.3. $J(0.87\omega_H)$ is very sensitive to differences in motion at the high frequency scale (Kadeřávek et al. 2014). Notably, in the $J(0.87\omega_H)$ function at 16.4 T, W_1, W_{2-1}, and W_{2-2} (Figure 52) show small localized increases of the spectral density values observed in the folded core that correlate very well with the location of loops within the W unit (Figure 53) presented herein in chapters 3 and 4. But, the loops or unstructured regions that are embedded within the core and not solvent exposed, such as the stretch between residues 40-60, do not display different dynamics compared to the helical segments. The loops also show different behaviour in the 11.7 T W_1 $J(\omega_N)$ data, with dips in the value of spectral density not visible at 16.4 T.

In the case of $J(0)$, the behaviour of W_1 and W_2 differ; W_1 has the lowest average frequency over the folded core, followed by W_{2-1}, than W_{2-2}. Oddly, W_{2-2} has a greater average than W_{2-1} over the folded core but is actually not statistically different (Figure 54). The only statistical difference in frequency is between W_1 at 11.7 T and both domains in W_2. The standard deviation of the W_1 11.7 T spectral density mapping is quite large compared to the spectral density mapping at 16.4 T. Perhaps the variance leads to the conclusion that the mismatched tuning at 11.7 T induced some error in peak intensity during data collection.

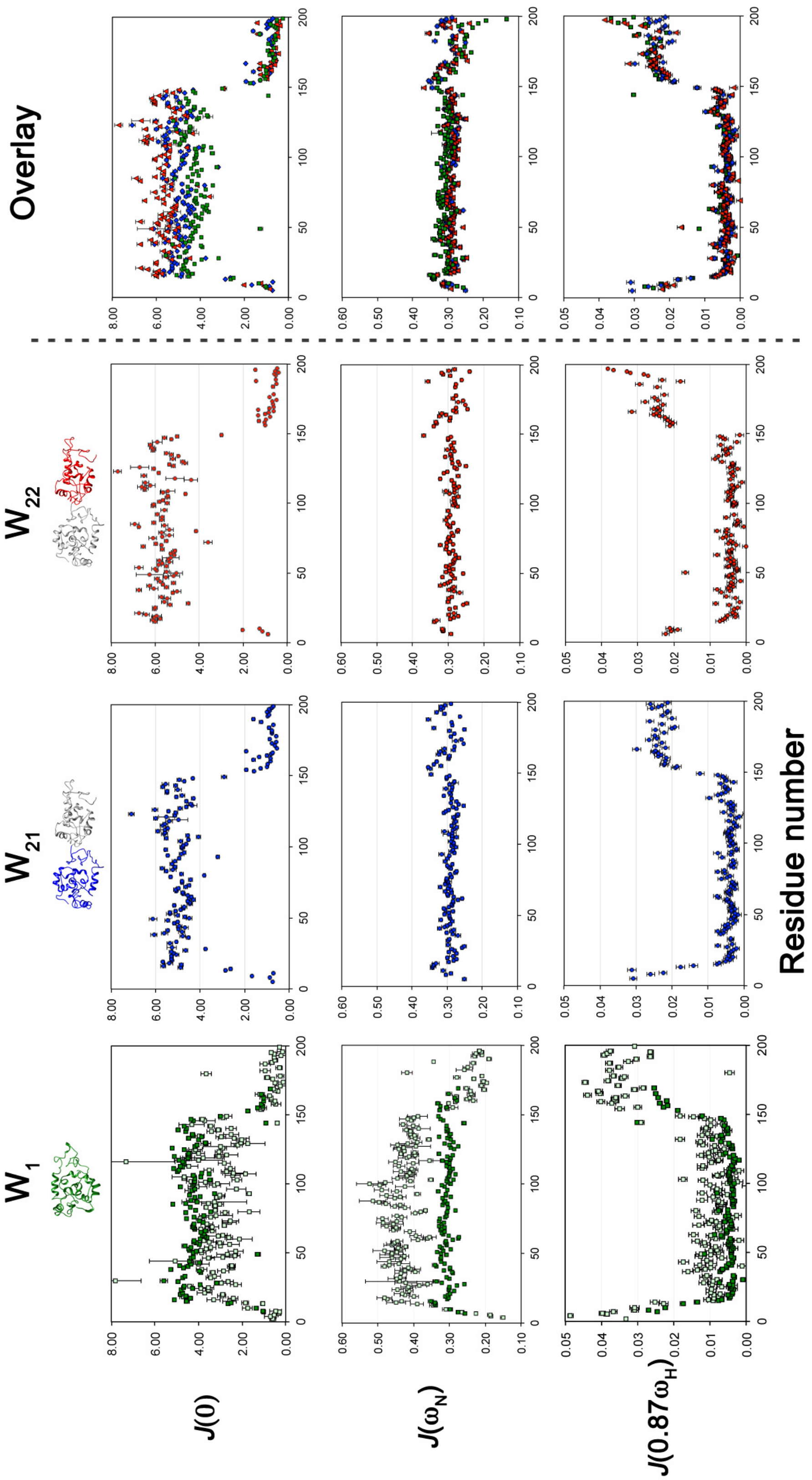


Figure 52 | Reduced spectral density mapping summary of W_1 (dark green), $W_{2\text{-}1}$ (blue), and $W_{2\text{-}2}$ (red) at 16.4 T, their overlay, and W_1 spectral density mapping at 11.7 T (light green)

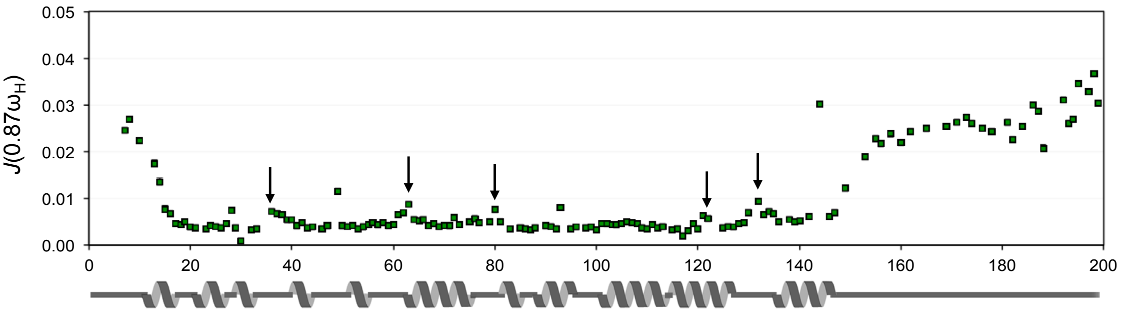


Figure 53 | W$_1$ $J(0.87\omega_H)$ reduced spectral density at 16.4 T. The arrows denote the slight increase in the high frequency values (marked by an arrow) that correlate with loops within W$_1$. W$_{2\text{-}1}$ and W$_{2\text{-}2}$ also demonstrate the same trend.

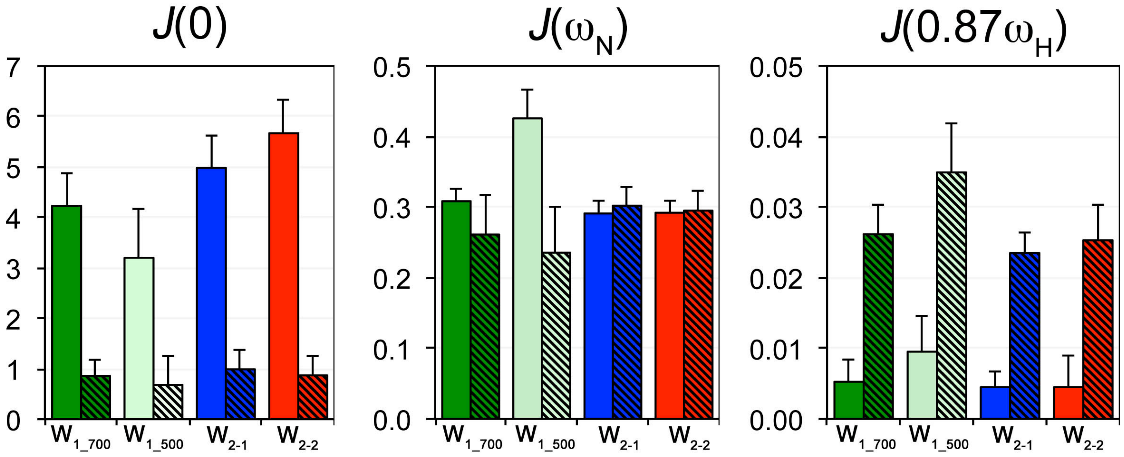


Figure 54 | Average frequency difference between the folded domain (aa12-150) (full colour) and the linker (aa1-11, 150-200) (diagonal hash) of the spectral density mapping. W$_{2\text{-}1}$ and W$_{2\text{-}2}$ data are at 16.4 T.

5.4. SUMMARY

The molecular motions calculated using reduced spectral density mapping do not limit the number of molecular modes a protein can adopt, which is highly appropriate for describing motions in flexible or disordered proteins. This is particularly true when the model-free approach fails to provide a global model for motion, as was the case for the W repetitive unit. Reduced spectral density mapping provided a better model for describing the amplitudes of motion observed in the T$_1$, T$_2$ and hetNOE relaxation data of W$_1$ (at 11.7 T and 16.4 T) and of W$_{2\text{-}1}$ and W$_{2\text{-}2}$ (at 16.4 T). I demonstrated that the folded domain (residues 12-149) was more rigid and much less dynamic than the linker or tails (residues 1-11, 150-199). The linker exhibited properties expected of an intrinsically disordered domain. In the linker region, both W$_1$ and W$_2$ have similar behaviour, with some damping of the motional amplitude at the junction between W domains in comparison to the free termini. The rotational correlation time (τ_c) of both W_1 and W_2 support non-spherical structuring, as presented in the hydrodynamics analysis in chapter 4. This implies that both W_1 and W_2 don't behave like a globular protein in solution, with faster tumbling being apparent.

CHAPTER 6. THE INTERMEDIATE STATE OF W PROVIDES INSIGHT INTO FIBRILLOGENESIS

6.1. PROTEIN STABILIZATION

6.1.1. Protein Stabilization Forces

Denaturation of proteins involves the disruption and unfolding of secondary, tertiary, and quaternary structures present in the native state by application of an external force (Dill 1990; Szilagyi et al. 2007; Baldwin 2007). From a thermodynamic point of view, protein stabilization comes from four major contributions: 1) the hydrophobic effect, 2) the energy of hydrogen bonds, 3) the energy of electrostatic interactions, and 4) van der Waals interactions (Chen et al. 2000).

Hydrophobic forces, which guide the rapid collapse of an unfolded polypeptide chain, are thought to be the main driving force in protein folding (Kauzmann 1959; Dill and Shortle 1991). The hydrophobic effect, conferring stability to a water-soluble protein, can be explained by the rearrangement of the hydrogen bonds between the solvent molecules around the apolar groups of the protein (Baldwin 2007). It is energetically unfavourable to have apolar groups interacting with polar solvent; therefore, by decreasing the apolar solvent-exposed area, the free energy of the system decreases. The stability provided from the hydrophobic effect depends highly on the temperature of the system and the hydrocarbon number and shape within the protein and ranges in energies of <10 kcal/mol (Prevost et al. 1991).

van der Waals (or London dispersion) forces arise from fixed or induced dipoles and contribute to the packing of the protein in a solvent (Baldwin 2005) with energy of ~0.1 kcal/mol per interaction (Roth et al. 1996). Following a Coulombic-type interaction, the energy of van der Waal forces generally decays with the inverse sixth power of separation distance. Together, the hydrophobic effect and van der Waals forces constitute weaker but very abundant forces that stabilize folded protein structure.

Hydrogen bonds are present in fewer number, but the energy contribution per bond is high leading to individually significantly contributions to protein stability of 1.5 ± 1.0 kcal/mol, depending on the electronegativity of the contributing partners (Vinogradov and Linnell 1971; Myers and Pace 1996). The dominant component of the H-bond is electrostatic, but it also comprises of dispersion, charge-transfer, and steric repulsion interactions (Dill 1990). H-bonds are found mainly in secondary structure elements like α-helices and β-sheets.

Electrostatic interactions provide significant energy (~10 kcal/mol) to protein stability. These mainly consist of salt bridges and ion pairing. Electrostatic interactions may be broken or decreased by changes in solvent pH and/or ionic strength (Dill 1990). The pH determines the charge of the moieties involved in the interaction while the ionic strength determines the extent of interaction between these charges (Friend and Gurd 1979; Dill 1990). Individually, the energy of H-bonds and electrostatic interactions contribute heavily to the stability of a protein.

All four factors of protein stabilization combine together to give rise to a single stabilization energy. The relation of Gibbs free energy to enthalpy and entropy may in turn be used to define protein stability,

$$\Delta G = \Delta H - T\Delta S \tag{6.1}$$

where, ΔG represents Gibbs free energy, ΔH represents enthalpy, ΔS represents entropy, and T is the absolute temperature of the system (in Kelvin). Proteins typically maintain native structuring because a favourable enthalpic term overcomes the entropic term. The stability is dependent on protein-solvent, solvent-solvent, and protein-protein interactions. Upon denaturation, the energy of the native state increases (Szilagyi et al. 2007).

In the context of AcSp1, conversion of the soluble protein into a fibre involves the conversion of the native soluble state into a solid-state fibrous conformation concomitant with a structural transition from a 2:3 ratio of α-helix:random coil to ~1:1:1 mixture of α-helix:β-sheet:random coil. The structural transition must therefore involve partial or complete denaturation of the W unit, with subsequent refolding into a different structure with decreased α-helical content and introduction of β-sheet character. In dragline silk, this transition may be initiated by pH change (Hagn et al. 2010b; Jaudzems et al. 2012), salting out by kosmotropic ions (Hardy et al. 2008), shear force (Heim et al. 2009), and/or dehydration (Silvers et al. 2010).

6.1.2. The Molten Globule

It is difficult to discuss protein intermediate states without mentioning the molten globule. The molten globule is a state where the polypeptide chain is compactly packed, just as it is in the native state, but the intramolecular motions of atoms and bond torsion angle rotation, especially in side chains, are extensively relaxed (Ohgushi and Wada 1983). This state still contains a prominent amount of secondary structuring, but a decrease of specific tertiary structure due to the release of side chain packing, an increase in the radius of gyration by 10%–30% compared to the native state, and accessibility of a protein's hydrophobic core to solvent molecules (Arai and Kuwajima 2000).

6.1.3. Protein Denaturation Factors

Proteins can be denatured by heat (Tsai et al. 2002; Prabhu and Sharp 2005), cold (Privalov 1990; Kunugi and Tanaka 2002), high pressure (Heremans and Smeller 1998; Kamatari et al. 2004; Meersman et al. 2006), extreme pH (Anderson et al. 1990; Fitch et al. 2006), or the addition of salts and other chemicals or solvents (O'Brien et al. 2007; Stumpe and Grubmüller 2007; Mohan and Hosur 2008; Nick Pace et al. 2010). With relation to the experiments described in this chapter, only chemically- and pressure-based denaturation processes will be addressed in detail.

In chaotropic denaturation, the addition of chemical agents, like urea or guanidinium chloride (GdmCl), disrupts surface hydrogen bonds and electrostatic charges (Makhatadze 1999; O'Brien et al. 2007; Stumpe and Grubmüller 2007). Urea molecules, for example, denature proteins by attacking the first solvation shell and disrupting surface electrostatic interactions, intramolecular hydrogen-bonding, and significantly disturbing dispersion forces (Das and Mukhopadhyay 2009). The intramolecular contacts become less stable, leading to a shift of the equilibrium away from the folded state toward the denatured state. As the first solvation shell is removed, the water network and urea molecules penetrate the interior of the protein, disrupting intramolecular hydrophobic forces, solvating apolar groups, increasing the protein-solvent interaction, and diminishing overall intramolecular interactions. The greater surface accessibility of the unfolded protein to solvent is ultimately the factor that causes unfolded protein structures to be favoured in concentrated urea solutions by increasing protein-solvent interactions and decreasing intramolecular interactions (Makhatadze 1999; O'Brien et al. 2007).

GdmCl's denaturation principle is similar to urea's with the exception of GdmCl lacking the ability to effectively hydrogen bond with peptide backbone like urea (Lim et al. 2009). The efficiency of GdmCl is 4 fold higher than urea in unfolding helices relying on planar amino acid side chain for stability (Dempsey et al. 2005). In contrast, urea is much more efficient than GdmCl in destabilizing salt bridges. Dempsey *et al.* (Dempsey et al. 2005) discovered that GdmCl is considerable more efficient at denaturing indole-indole interactions in protein than hydrogen-bonding.

Detergent micelles can also act as protein denaturants (Otzen 2002). In the case of sodium dodecylsulfate (SDS), the monomeric detergent typically binds to the native state of proteins (Reynolds and Tanford 1970; Bordbar et al. 1997). The amphiphilic properties that are shared by both the protein and detergent are the driving force for denaturation. Proteins denatured in SDS retain some folded characteristics, albeit non-native (Otzen 2002). Above detergent critical micelle concentration (CMC) (Jones et al. 1975; Turro et al. 1995; Lu et al. 2005), water competes with micelles for access to hydrophobic regions in proteins, increasing the entropy of the system.

Applying pressure to a system requires a slightly different conceptual treatment, since physical forces are at play instead of chemical forces. Increasing the pressure of a system leads to an increase in density and rearrangement of molecules, leading to a decrease of "empty space" between atoms. Increased pressure also shifts the conformational equilibrium of protein molecules toward lower volume conformers, therefore decreasing the partial molar volume of a protein (Kamatari et al. 2004). Pressure strongly modifies the hydrogen bond network of water, which adopts a hexagonal arrangement rather than the more commonly observed tetrahedral arrangement (Starr et al. 1999; Cai et al. 2005) around polar and nonpolar residues. In other words, the water lattice contains about 4.3 molecules at standard atmospheric pressure, while at pressures of 1000 MPa the number of water molecules increases to 10 per lattice unit (Hayakawa et al. 1996). This hexagonal arrangement ultimately results in the amplification of the hydrophobic effect by a large decrease in H-bonding energy between water and the protein (Starr et al. 1999; Cai et al. 2005). At high water density, the boundary between the organized water in the neighbourhood of a nonpolar group and the bulk water becomes less distinct, merging into a single arrangement that minimizes the overall entropy of the system (Grigera and McCarthy 2010). The high-density water in its new arrangement decreases the thermodynamic advantage of the system, destabilizes the nonpolar group, and exposes these nonpolar groups to the bulk water, leading to protein unfolding.

Intermediate states of proteins can be monitored using various biophysical techniques, such as fluorescence spectroscopy and CD spectropolarimetry. Another very effective way is through heteronuclear 2D NMR spectroscopy, such as the trusted protein backbone-fingerprinting tool: the ^{1}H-^{15}N HSQC. NMR provides atomic resolution information on the structure of the protein at different stages of the folding process and chemical shifts can be followed as a function of position in the protein through the unfolding process.

6.1.4. Project Goals

The aim of this project was to monitor the changes that occur in W_1 and W_2 when destabilizing agents are added in solution: the chaotropic agents urea and guanidinium chloride (GdmCl), the detergent dodecylphosphocholine (DPC), and pressure. These were performed with the goal of gaining insight into the local structural changes that occur upon initialization of fibrillogenesis. The changes were monitored residue per residue in uniformly labeled ^{15}N W_1 via ^{1}H-^{15}N HSQCs and identically reflected in W_2 with the addition of DPC. Amid all the denaturing factors employed, one local structural change is common: H5 is the first helix to denature. The functional relevance of unfolding of H5 is supported by a stronger propensity for β-sheet character in this region, as determined through secondary structure propensity (SSP) analysis (Marsh et al. 2006).

6.2. MATERIALS AND METHODS

6.2.1. Samples (Supplied By Lingling Xu)

Uniformly ^{15}N- or ^{13}C/^{15}N-enriched W_1 was produced as described in chapter 3. Three samples with ^{15}N-W_1 (600 μL, 0.34 mM) in NMR buffer (20 mM AcOH, 1mM DSS, 1 mM NaN_3, pH 5) in 5 mm NMR tubes were made for in titrations. ^{13}C/^{15}N-W_1 (0.3 mM in NMR buffer) with 20 mM dodecylphosphocholine (DPC-d_{38}) was used for the acquisition of 3D backbone NMR data. Segmentally labeled W_2 ({^{15}N}$W_{2\text{-}1}$ and {^{15}N/^{13}C}$W_{2\text{-}2}$) was expressed, purified, and intein-spliced as described in chapter 4. The W_2 NMR sample concentration was 0.17 mM in our standard NMR buffer. Stocks of 1, 10, and 100 mM DPC, 10 M urea, and 8 M guanidinium chloride (GdmCl) in NMR buffer at pH 5 were made for the titration experiments

6.2.2. NMR Spectroscopy

DPC was added to ^{15}N-enriched W_1 (0.34 mM in NMR buffer) at an increasing molar ratio (W_1:DPC of 10:1, 5:1, 1:1, 1:5, 1:10, 1:50), with ^{1}H-^{15}N HSQC experiments (16 scans, 2048x128 points, recovery delay of 1.5 s), acquired at each titration point using a 16.4 T Bruker Avance III spectrometer equipped with a 5 mm indirect detection TCI cryoprobe at 303.15 K. The change in W_1 concentration due to DPC volume increase was negligible. 3D backbone NMR experiments (Table 18) were acquired at the DPC endpoint (0.34mM W_1, 20 mM DPC; DPC CMC: 1.1 mM (Palladino et al. 2010)) to facilitate assignment. Perturbation of W_2 (intein-spliced variant the first W-unit uniformly ^{15}N-enriched and the second W-unit ^{13}C- and ^{15}N-enriched; 0.17 mM in NMR buffer) was monitored using the isotopically discriminated (IDIS) ^{1}H-^{15}N HSQC experiment (Golovanov et al. 2007) at 0 mM and 20 mM DPC to simultaneously acquire data on both W domains in a single sample. It should be noted, for reproducibility, that o2p was set at the centre of the CO region (175 ppm) as the IDIS experiment will otherwise fail.

W_1 (0.34mM in NMR buffer) was titrated with GdmCl (0, 0.4, 0.6, 0.8, 1, 1.2, and 2 M) and urea (0, 0.5, 1, 1.2, 1.4, 2.0, 2.3, 2.5 M) (0.34mM W_1 in NMR buffer), and monitored by HSQC spectra acquired at 11.7 T on a Bruker Avance spectrometer equipped with a 5 mm BBFO SmartProbe (details of HSQC?). Concentration variation due to volume change was accounted for the denaturants at each titration point but not for W_1, for which a 25% decrease in concentration (0.25 mM) was observed at both GmdCl and urea endpoints.

The pressure data were acquired by Drs. Scott Prosser and Jan Rainey on a 600 MHz Varian Inova spectrometer (Agilent Technologies) equipped with a 5 mm triple-resonance cryoprobe in a 3.4 mm i.d. sapphire NMR tube (Daedalus Innovations, Aston, PA, USA). High pressures were achieved using the Xtreme 60 high-pressure NMR cell (Daedalus Innovations, Aston, PA, USA). ^{1}H-^{15}N HSQC experiments were acquired with a relaxation delay of 1.5 s and 1024 x 256 transients.

All acquired spectra were processed with NMRPipe (Delaglio et al. 1995), adding linear prediction in the ^{15}N dimension, and assigned with CcpNMR Analysis (Vranken et al. 2005). Combined chemical shift differences, weighed by gyromagnetic ratio for ^{1}H and ^{15}N nuclei (Schumann et al. 2007), were calculated between the samples containing no titrant and those at 20 mM DPC, 0.8M GdmCl (0.30 mM W_1), and 2.5M urea. The spectra chosen for analysis of GdmCl and urea perturbation were not based upon titration endpoints; rather, these were chosen as the most assignable spectra before denaturation made backbone amide chemical shifts indiscriminable. All urea titration spectra were re-referenced to the urea NH_2 peak at the lowest concentration (0.5 M) (^{1}H: 5.765 ppm and ^{15}N: 122.87 ppm) to account for the small change in chemical shift in the presence of urea. Similarly, the GdmCl titration spectra were referenced to the GdmCl NH_2 peak (^{1}H: 6.721 ppm and ^{15}N: 119.76 ppm). All shifts were placed into an NMR series in Analysis and were followed using the tool in Data Analysis > Follow intensity changes. The shifts were exported to Microsoft Excel for chemical shift comparison (Schumann et al. 2007) using equation 4.1 (as reiterated here):

$$\Delta\delta_{comb} = \sqrt{\frac{1}{N_a}\sum_{i=1}^{N_a}(w_i\Delta\delta_{ij})^2} \qquad (4.1)$$

All structures were visualized and figures produced with Chimera (Pettersen et al. 2004).

Table 18 | List of 3D NMR experiments for W_1 in 20 mM DPC

<table>
<tr><th>Experiment</th><th>Bruker pulse sequence</th><th>Relaxation delay (s)</th><th>Number of scans</th><th>Number of points
F1|F2|F3</th><th>Sweep width (ppm)
F1|F2|F3</th><th>Center position (ppm)
F1|F2|F3</th><th>^{1}H frequency</th><th>Facility</th></tr>
<tr><td>HSQC</td><td>hsqcetf3gpsi</td><td>1.75</td><td>8</td><td>2048|128</td><td>14|25</td><td>4.7|116.5</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNCA</td><td>hncagp3d</td><td>1.75</td><td>8</td><td>2048|48|80</td><td>16|22.5|24</td><td>4.7|115.25|53.0</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNcoCA</td><td>hncocagp3d</td><td>1.75</td><td>8</td><td>2048|48|80</td><td>16|22.5|24</td><td>4.7|115.25|53.0</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNCO</td><td>hncogp3d</td><td>1.75</td><td>8</td><td>2048|48|64</td><td>14|22.5|10</td><td>4.7|115.25|176.0</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>HNcaCO</td><td>hncacogp3d</td><td>1.75</td><td>8</td><td>2048|48|64</td><td>14|22.5|10</td><td>4.7|115.25|176.0</td><td>700</td><td>NRC-IMB</td></tr>
<tr><td>CBCANH</td><td>cbcanhgp3d</td><td>1.75</td><td>8</td><td>2048|48|128</td><td>14|22.5|58</td><td>4.7|115.25|41.0</td><td>700</td><td>NRC-IMB</td></tr>
</table>

6.2.3. Simulations Of The Putative W_1 Intermediate State

The distance restraint and dihedral angle files generated at the last iteration of my W_1 structure calculations (Table 15) were modified by commenting out all restraints associated with residues 135-155. Xplor-NIH (Schwieters et al. 2006) was employed for one round of calculations using these restraint and the 20 lowest energy structures out of 100 were used for the evaluation of an average surface hydrophobicity. Accessible surface areas for both W_1 and the W_1 intermediate state (W_1IS) were obtained via the DSSP web server (http://www.cmbi.ru.nl/xssp/) (Kabsch and Sander 1983). The average area of the most hydrophobic amino acids (I, V, L, A and F) was totalled and compared between W_1 and W_1IS. The Kyte-Doolittle hydrophobicity scale (Kyte and Doolittle 1982) (embedded within Chimera (Pettersen et al. 2004)) was used to assess W_1 hydrophobicity using surface visualization.

6.3. RESULTS

6.3.1. Project Initiation

The idea for this chapter originated from the production of a new protein construct where a W_1 was be attached to the AcSp1 non-repetitive C-terminal domain in an attempt to solve the structure of a C-terminal domain (C_{ac}) when attached to the repetitive domain. This protein wasn't very soluble and was prone to rapid aggregation. A previous NMR structural study showed that solubilization of a C-terminal domain was achieved through the addition of the detergent dodecylphosphocholine (DPC) (Wang et al. 2014). The AcSp1 W_1-C_{ac} protein dissolved and stayed in solution for days, allowing 3D NMR data collection. CD spectropolarimetry demonstrated little perturbation of the overall structure of W_1-C_{ac} in the presence vs. absence of DPC, with demonstration of a slight decrease in negative ellipticity for the helical band at 222 nm in the absence of

DPC but unchanged ellipticity at 208 nm (Figure 55). In comparison, W_1 in NMR buffer shows a similar overall CD spectrum at a lower intensity. A significant perturbation in this case would be expected to involve denaturation and a shift towards random coil structuring, which is definitely not apparent with retention of the minima characteristic of an α-helix (Greenfield and Fasman 1969) (Figure 55).

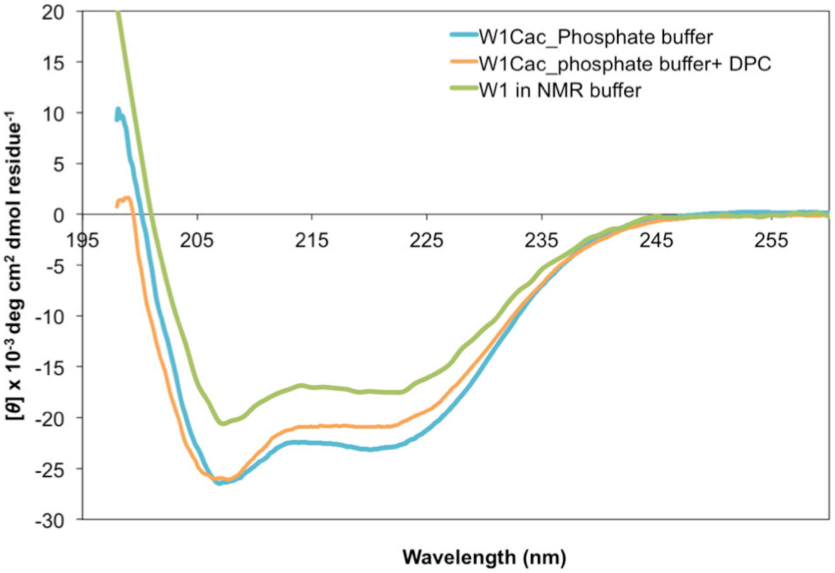


Figure 55 | CD of W_1+C_{ac} with and without 50 mM DPC. W_1 in NMR buffer was added for comparison. Data provided by Dr. Lingling Xu.

To test for the possibility that DPC interacts with W_1 and to see if the C_{ac} domain affected the W_1 structure, a ^{1}H-^{15}N HSQC experiment was acquired on a segmentally labeled W_1C_{ac} construct where the W_1 unit was uniformly ^{15}N-enriched and the C_{ac} was at natural abundance (Figure 56). As can be seen in Figure 56, W_1 differs between buffer and DPC; therefore, DPC is interacting with W_1. But, curiously, DPC doesn't completely unfold the structure since there is still peak dispersion and some of the peaks don't shift. Therefore, we decided to look at W_1-DPC interactions as a potential means to gain insight into the initial steps of fibre formation by determining which portions of W_1 were most susceptible to denaturation. Beyond these results, W_1C_{ac} is not discussed further in this thesis as these initial experiments were followed up by an undergraduate student (Kathleen Orrell) rather than me.

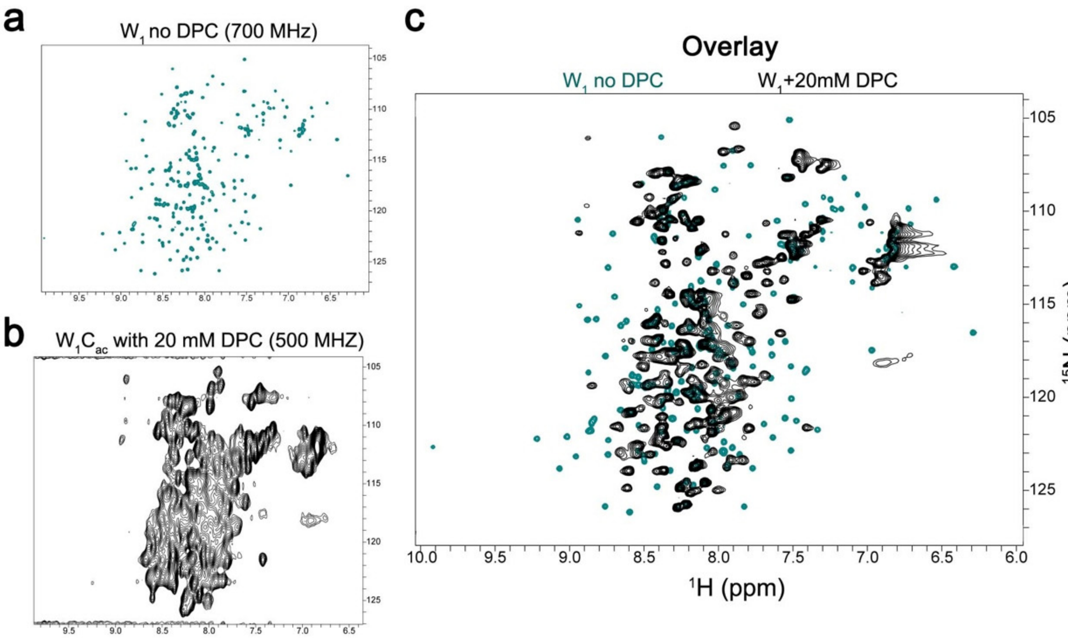


Figure 56 | ^{1}H-^{15}N HSQCs acquired in DPC to screen for the effect of detergent addition. (a) W_1 in NMR buffer (0.3 mM) containing no DPC acquired at 700 MHz. (b) Segmentally labeled W_1C_{ac}, where W_1 is uniformly ^{15}N-labeled and C_{ac} is at natural abundance, in 20 mM DPC, collected at 500 MHz. (c) Overlay of the spectrum in (a) with ^{15}N-W_1 in 20 mM DPC collected at 500 MHz demonstrating the decreased spread of chemical shifts in W_1 in the presence of DPC.

6.3.2. DPC Titration With W_1

Perturbations to ^{1}H-^{15}N HSQC peak positions upon DPC addition were quite easily followed over the first three titration points (0.034, 0.068, and 0.34 mM DPC) since most peaks moved only slightly. At a 1:1 ratio of W_1:DPC (0.34 mM DPC), peaks for some residues disappeared and could not be followed, such as those for L25, S29, T30, and L100. These are residues with large secondary chemical shifts and are located either far upfield or downfield on the HSQC spectrum. Interestingly, they are also part of structural elements of W_1 (Figure 57b). About 60 peaks could be followed all the way to the 20 mM DPC endpoint without requiring sequential assignment, with only mild confidence for some of the assignments. The peaks that could be followed corresponded to the N- and C-terminal regions of W_1, which are the intrinsically disordered regions. With the aid of 3D backbone NMR data acquired at the 20 mM endpoint, 153/199 residues were assigned with high confidence, despite the typical overlap in ^{1}H, ^{15}N, and ^{13}C chemical shifts pertaining to glycine and serine residues. Unfortunately, assignment of the serines in the polyserine region from residues 141-153 of H5 (helix nomenclature in Figure 25) was not possible.

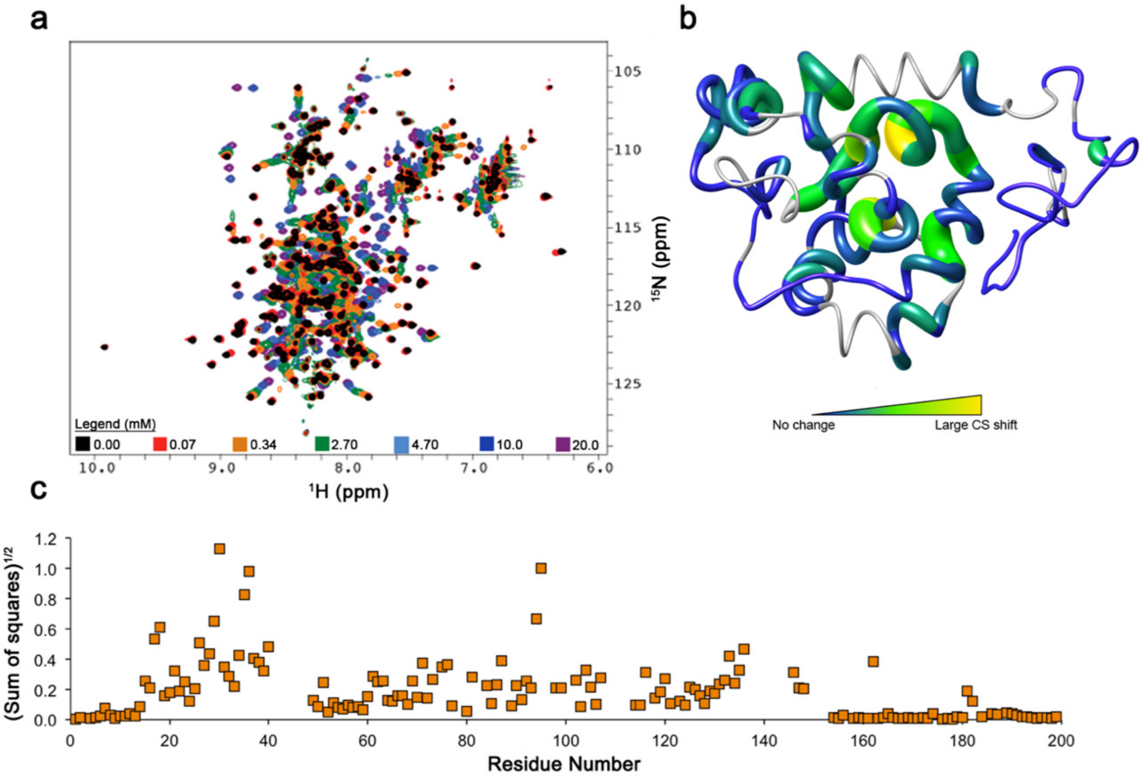


Figure 57 | Summary of the DPC titration with W_1. (a) ^{1}H-^{15}N HSQCs collected at each DPC addition. The W_1 concentration was 0.34 mM and the DPC concentrations are located at the bottom of the HSQC. (b) Per residue chemical shift changes represented on the lowest energy structure of W_1. (c) Graphical summary of the DPC titration (endpoint 20 mM) represented as a square root of the sum of squares, weighted by gyromagnetic ratio (Schumann et al. 2007).

Naïvely, one might expect the disordered regions to be most prone to interaction with DPC molecules or micelles. Somewhat counter-intuitively, therefore, the disordered "linker" regions of W_1 were not perturbed by DPC and, in fact, overlap with the same intensity throughout the titration. The folded domain, conversely, undergoes a transition to a different structural state, with peaks broadening and losing intensity and subsequently regaining intensity at some later point in the titration (Figure 57b-c). The residues that underwent the largest perturbations were 1) H5, from residues ~130-150; 2) those in the region located proximal in space to H5, from residues ~25-38; and, 3) the central helix from 90-95. Coincidentally, region (2) is also the region of W_1 that was the most difficult to assign under normal NMR conditions due to low intensity and with chemical shifts that deviated most significantly from the reported BMRB statistical ranges (see section 3.3.2).

The implication that H5 is structurally weaker than the rest of the folded structure of W_1 also brings to mind the H/D exchange experiments presented in chapter 3 (Figure Y). There too, H5 was the most readily exchanged portion of the structured domain with D_2O, confirming the local vulnerability of this segment to unfolding.

6.3.3. Guanidinium Chloride And Urea Titration With W_1

Similarly to the DPC titration just detailed, the goal of the chaotropic denaturation experiments for W_1 were to assess whether chaotropes induced unfolding of regions of W_1 in a specific order, allowing mapping of the structurally weaker regions of W_1. Unlike the DPC endpoint, where a structural transition was apparent, GdmCl and urea titrations would be expected to denature W_1. In both of these titrations, the individual peaks in the HSQC were extremely difficult to follow as a function of titrant due to peak broadening. This effect was exacerbated with GdmCl relative to urea. The point in each titration where peak positions became difficult to follow was coincident with the initial loss of α-helical character evidenced by a decrease in intensity of the characteristic negative band at 222 nm by CD (Figure 58). About 130 peaks were assignable in the GdmCl condition at 0.8 M (Figure 59). Unfortunately, during the titration, some peaks became overlapped, and discriminating between them became impossible. The same trend was observed for the urea titration, but this time the transition to the denatured state was smoother and slower, allowing for 172/191 peaks to be followed (Figure 60).

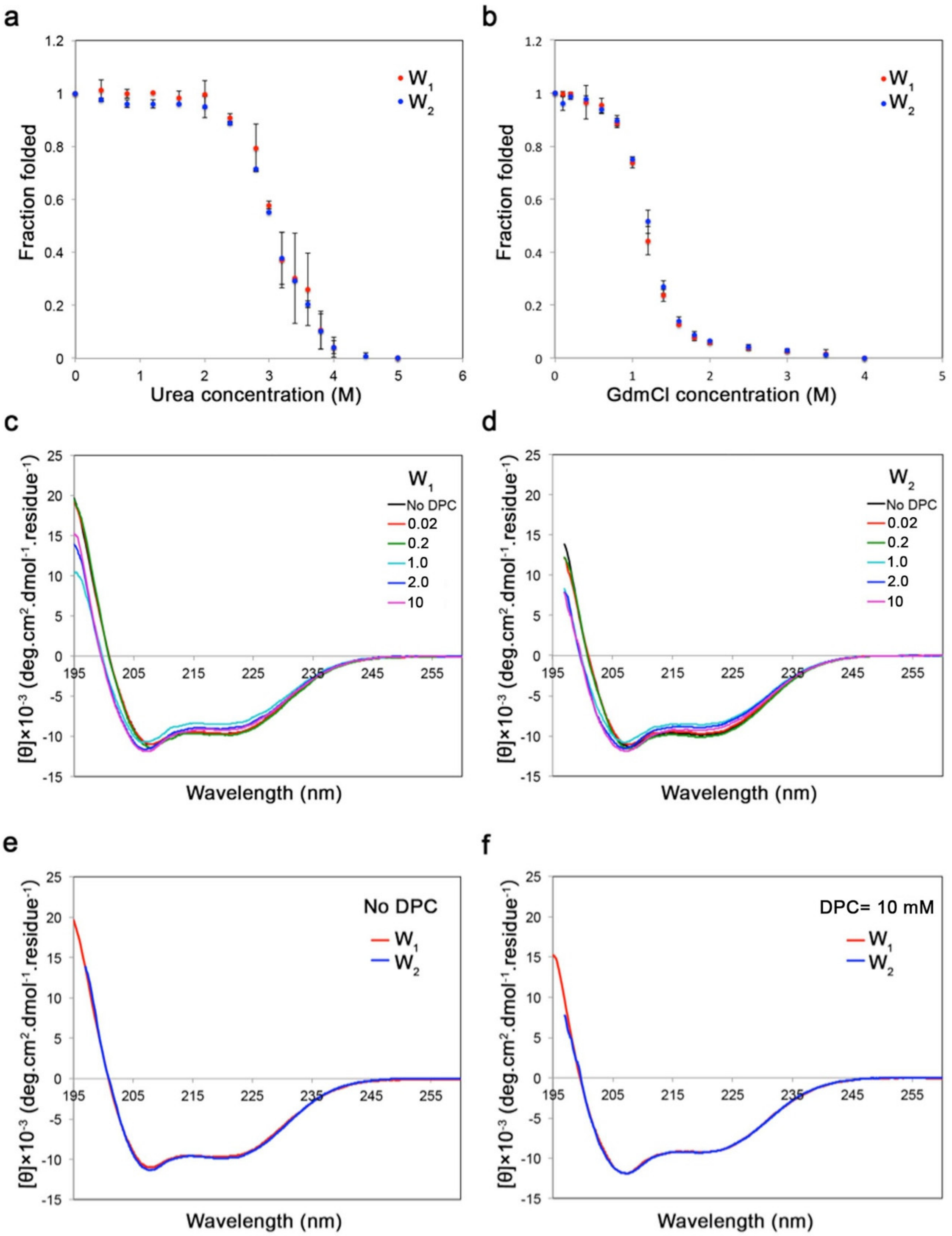


Figure 58 | Far-UV CD spectroscopy comparing effects of addition of the chaotropes urea and guanidium chloride (GdmCl) or the detergent dodecylphosphocholine (DPC) upon W$_1$ and W$_2$. a. Urea and b. GdmCl titration-based denaturation curves, with fraction folded by ellipticity at the characteristic α-helical band at 222 nm at 22±2˚C. c-f. Titration of W$_1$ and W$_2$ with DPC at indicated molar ratio. [*Data collected and figure made by Lingling Xu that helped in determining concentration for the NMR experiments*]

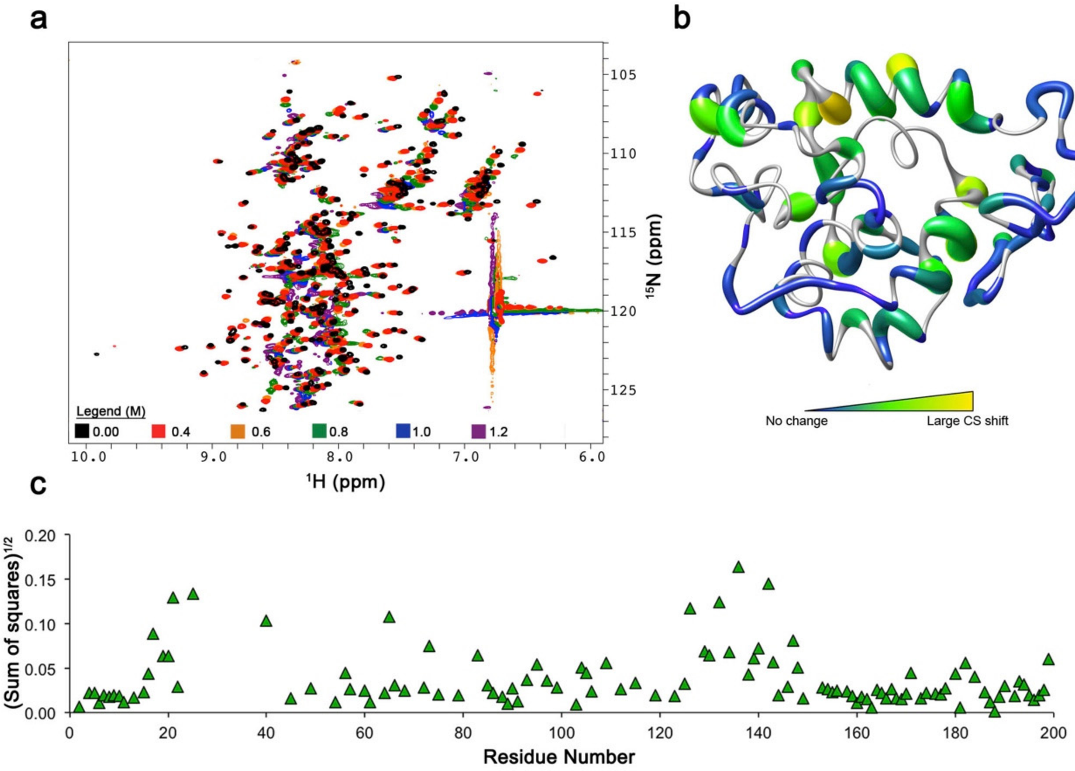


Figure 59 | Summary of the GdmCl titration with W_1. (a) ^{1}H-^{15}N HSQCs collected at each GdmCl addition. The W_1 concentration was 0.34 mM and the GdmCl concentrations are located at the bottom of the HSQC. (b) Per residue chemical shift changes (0 to 0.8 M) represented on the lowest energy structure of W_1. (c) Graphical summary of the GdmCl titration (0 to 0.8 M) represented as the square root of the sum of squares, weighted by gyromagnetic ratio (Schumann et al. 2007).

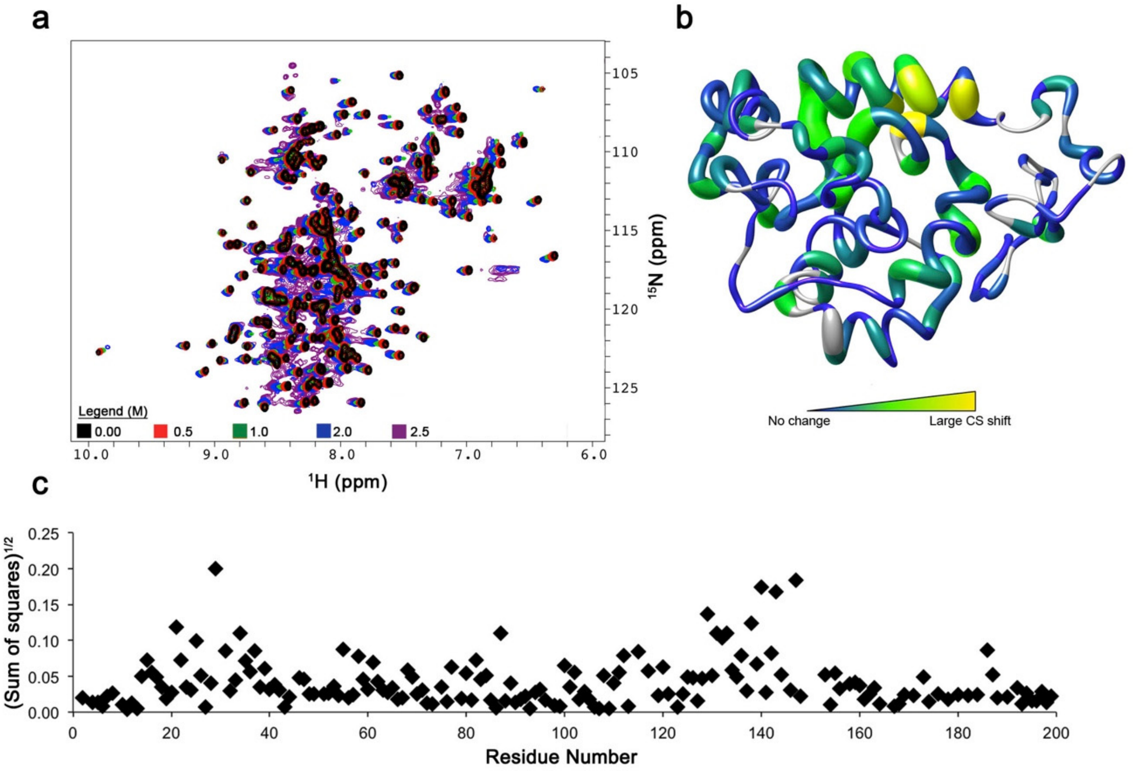


Figure 60 | Summary of the urea titration with W_1. (a) ^{1}H-^{15}N HSQCs collected at each urea addition. The W_1 concentration was 0.34 mM and the urea concentrations are located at the bottom of the HSQC. (b) Per residue chemical shift changes between 0 - 2.5 M are represented on the lowest energy structure of W_1. (c) Graphical summary of the urea titration (0 to 2.5 M) represented as the square root of the sum of squares, weighted by gyromagnetic ratio (Schumann et al. 2007).

6.3.4. W_1 Under Pressure

The effect of pressure on W_1 was used to simulate the forces involved in fibrillogenesis. The series of ^{1}H-^{15}N HSQC experiments performed under increasing pressure mimicked the effect observed with chemical denaturation, but to a lesser extent (Figure 61). The chemical shifts for H5 and the region proximal to it around S29 and T30 changed the most between 1-1200 bars. Past this point, all peaks start to change position, obscuring the perturbations to the peaks observed during the initial movement.

Four distinct residues undergo exchange, evidence by two sets of peaks instead of one, as soon as pressure is applied, with slowly exchange evident as the pressure increases: 10T, 48V, 93A, 103S (in magenta in Figure 61b and Figure 62). Interestingly, referring back to section 5.3, two of these residues don't show slow exchange at long R_2 delays (93A and 103S) while two do (10T and 48V). These slowly exchanging residues are also located on the face of the protein where residues don't show a great chemical shift disturbance upon denaturation (Figure 57, Figure 59, Figure 60, Figure 62). But, quite remarkably, pressure induces minimal overall chemical shift perturbations throughout W_1, similar to urea denaturation. Notably, W_1 regains its original conformation when the pressure is returned to 1 bar, meaning that inter-conversion between the two states is reversible.

Rather than unfolding the structure, like the effect generated by the chemical denaturants, pressure would be predicted to compress the protein, forcing a transition to a more compact conformation. As discussed in chapter 3, W_1 would have to assume a different conformation upon compression to decrease the volume that it occupies.

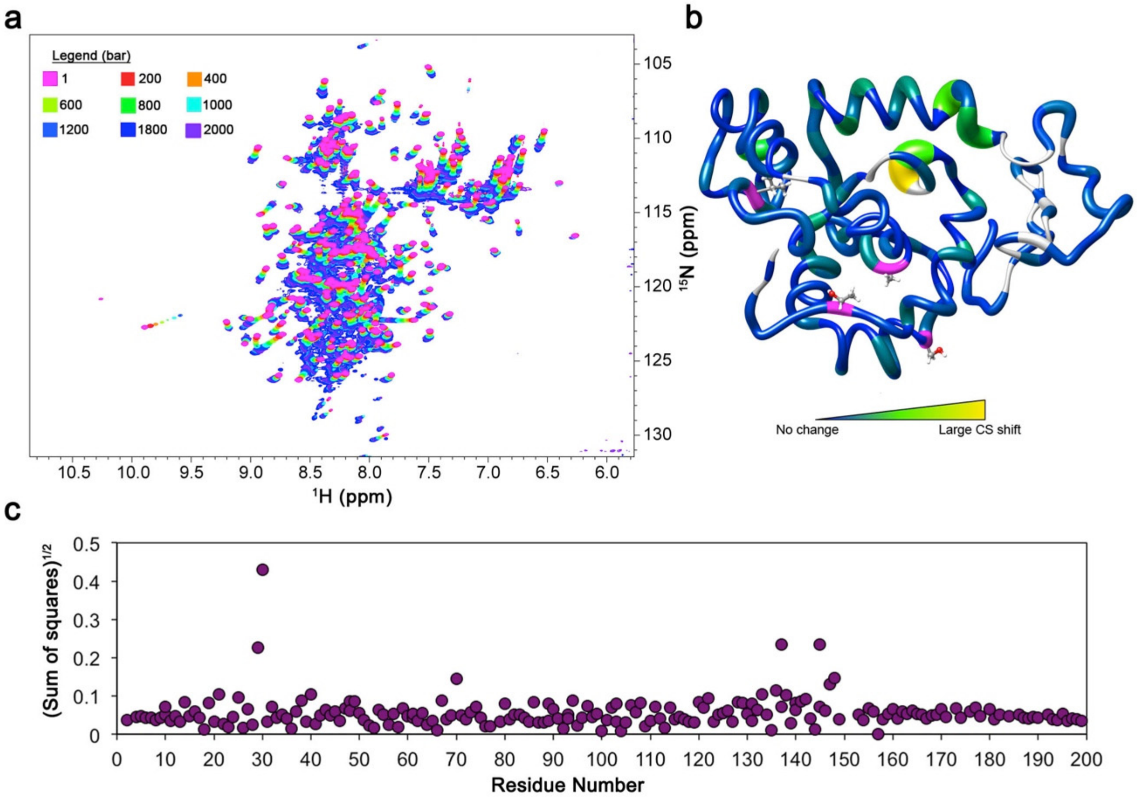


Figure 61 | Summary of the pressure titration with W_1 to 1200 bar. (a) Overlaid ^{1}H-^{15}N HSQCs collected at each pressure point (coloured as in the legend on the top). (b) Per residue chemical shift changes at 1200 bars represented on the lowest energy structure of W_1. In magenta are the residues that show slow exchange upon application of pressure. (c) Graphical summary of the response of W_1 to pressure represented as a square root sum of squares between at 1200 bars relative to 1 bar, weighted by gyromagnetic ratio (Schumann et al. 2007).

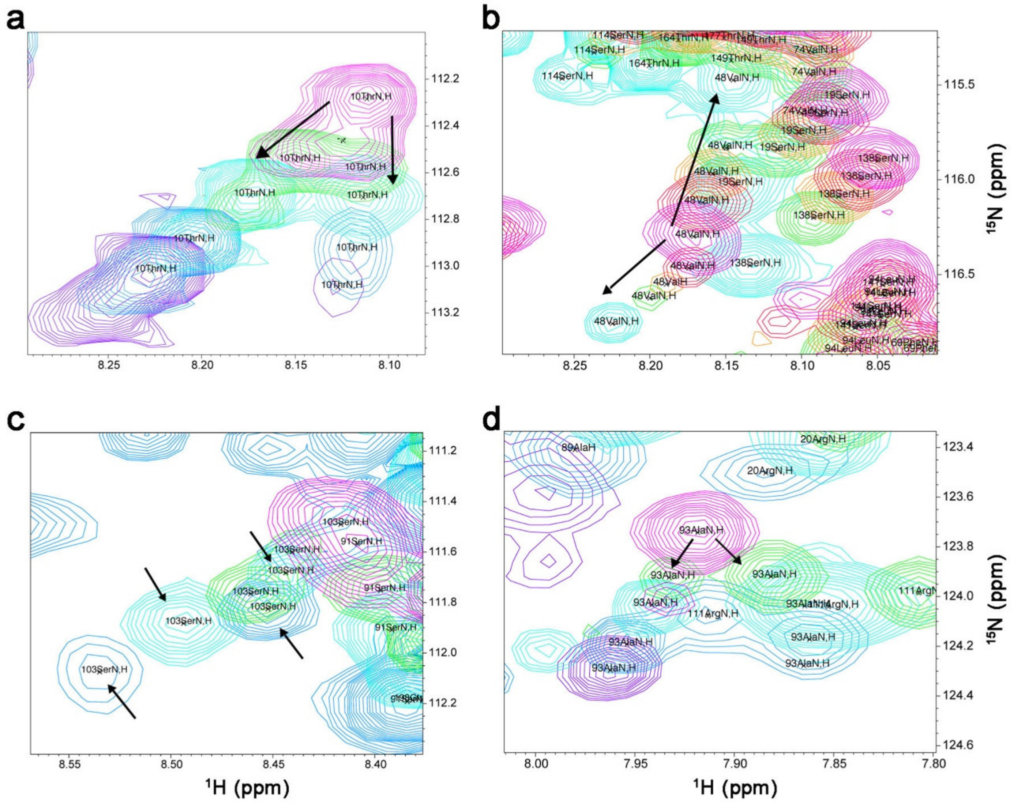


Figure 62 | Apparent slow exchange in W_1 at elevated pressures for residues 10T (a), 48V (b), 103S (c), 93A (d). The arrows indicate the direction of peak movement from the spectrum at 1 bar (pink peaks) between 600 bars (green), 1000 bars (cyan), 1200 bars (light blue), and 2000 bars (purple). In (c), both inter-converting peaks move in the same direction, but one peak displays a greater chemical shift change than the other. The down and up arrows indicate pairs of peaks at 1000 bar and 1200 bar respectively. In all of these cases, non-degenerate peaks start appearing at 200 bars.

6.3.5. W_2 DPC Titration

Although the results for W_1 are very exciting in terms of implying a potential route for the initial fibrillogenesis step, W_1 cannot form fibres and these results may therefore not be directly relevant to the fibre formation process when the full linker is intact. Therefore, as a remedy, we returned to our intein toolbox to create a construct where one W unit was uniformly ^{15}N-labeled and the other was uniformly ^{13}C/^{15}N-labeled. Using isotopically discriminated (IDIS) NMR experiments (Golovanov et al. 2007), it was possible to observe and distinguish both W domains in W_2 simultaneously in solution. In IDIS ^{1}H-^{15}N HSQC experiments, data are acquired in an interleaved manner for ^{1}H-^{15}N-^{12}C' vs. ^{1}H-^{15}N-^{13}C' correlation spectra. Once the interleaved data are split and processed, two HSQC spectra are produced, one specific for the first labeling scheme and a second for the other.

Figure 63 demonstrates that $W_{2\text{-}1}$ overlays perfectly with $W_{2\text{-}2}$; both units of W_2, in turn, overlay beautifully with W_1 (with exception of the terminal residues, of course) at the 20mM DPC endpoint. This finding reinforces my confidence that the W domains behave completely independently from each other. In this case, assignments were not made for W_2 in DPC since the HSQC fingerprint matched that already observed and assigned for W_1.

As in both the initial W_1C_{ac} experiments (section 6.3.1) and with W_1, W_2 stayed soluble in solution in DPC as long as the DPC concentration was above its CMC. W_2 typically had difficulty staying in solution when the DPC concentration was below CMC, and the sample kept aggregating during acquisition.

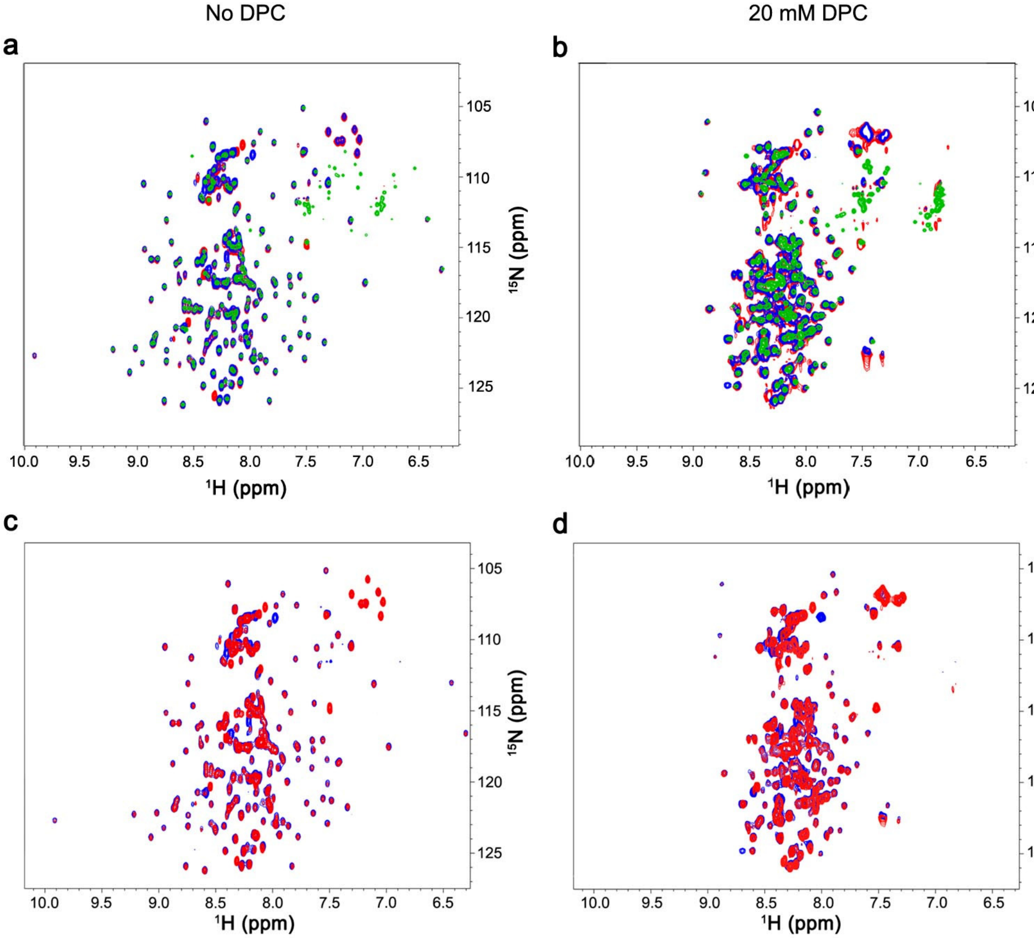


Figure 63 | Isotopically discriminated (IDIS) ^{1}H-^{15}N HSQCs of W_2. These were acquired with an intein spliced W_2 protein in such a way that the first W in W_2 (blue: $W_{2\text{-}1}$) is ^{13}C-^{15}N labeled and the second W is only ^{15}N labeled (red:$W_{2\text{-}2}$). In panels a and b, I show the HSQCs for the control sample and for the end point at 20 mM DPC with W_1 (green) spectra overlaid. In panels c and d, only W_2 is shown.

6.4. IS INTERCONVERSION BETWEEN SOLUBLE PROTEIN AND FIBRE SEEDED AT H5?

The purpose of these studies was to look for regions of W_1 more readily perturbed in response to a variety of perturbants, allowing me to propose an initial conformational change step for the fibrillogenesis pathway. All titrations presented in this chapter lead to the same conclusions: H5 and the area proximal to it between residues ~25-35 are the most readily perturbed in the face of destabilizing factors. These data are combined in Figure 65, which clearly demonstrates that one face of W_1 is more extensively perturbed than the other. In other types of spider silks (Hijirida et al. 1996; Lewis 2006; Lefèvre et al. 2011), A_n regions are known to form β-sheets in fibres (Lewis 2006; Lefèvre et al. 2011). AcSp1 does not contain such a sequence, even though it has ~27% β-sheet content in its fibrillar form (Figure 67). SSP analysis reveals that the region in W_1 and W_2 that shows the greatest propensity for β-sheet structure is H5 and the region C-terminal to this (Figure 66). Interestingly, H5 also contains the only serine-rich portion in W (in bold, Figure 64).

131 141 151 161

VDSGSVQSDI **SSSSS**FL**STS** **SSSASYSQAS** **ASSTS**GAGYT

Figure 64 | The serine rich (in bold) amino acid region of W_1.

Polyserine regions (S_n) are not very often ascribed to a particular structural feature in protein structure. Most of the literature discussion of S_n motifs relates to Huntington's disease, and, even then, the roles of A_n and Q_n are more heavily discussed than that of S_n in the disease. An egg-case spider silk paper from Zhao *et al* (Zhao et al. 2005) mentioned in the abstract that "The presence of Ser-rich and GVGAGASA motifs suggests the formation of a β-sheet", but no reference to or evidence to back up this statement was given. Although this is motivating that I am not the only one suggesting that Ser-rich regions have β-sheet propensity, the lack of literature on the subject renders definite conclusions difficult. Only structural studies, possibly via ssNMR, can resolve my hypothesis unambiguously.

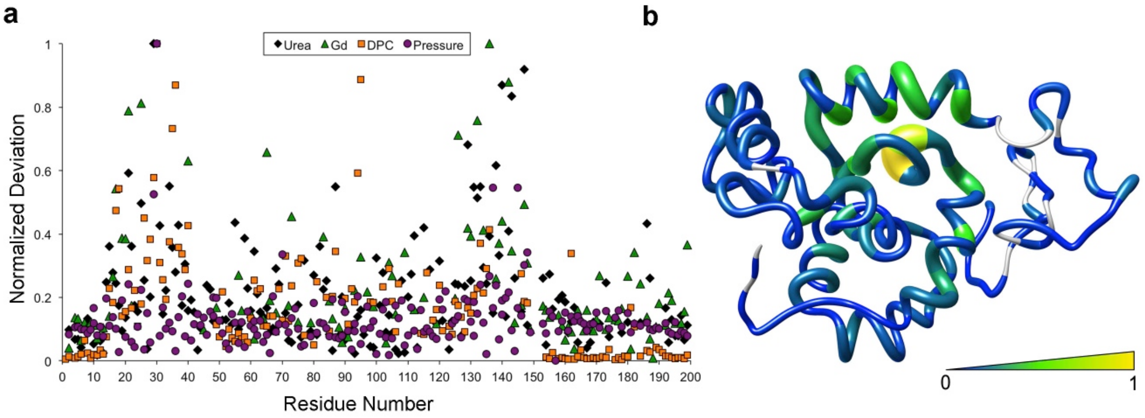


Figure 65 | Site-specific perturbation of W_1 upon denaturation, detergent treatment, and pressure. (a) Normalized combined chemical shift displacement as a function of amino acid position in W_1 for titrations with the chaotropic denaturants urea (black diamond; 2.5 M point) and GdmCl (green triangle; 0.8 M point), the detergent dodecylphosphocholine (DPC; orange squares; 20 mM endpoint), and pressure (purple circles, 1200 bar). (b) Representation of the average displacement for urea, GdmCl, DPC, and pressure by thickness and colour on a cartoon representation of the lowest-energy member of the W_1 structural ensemble.

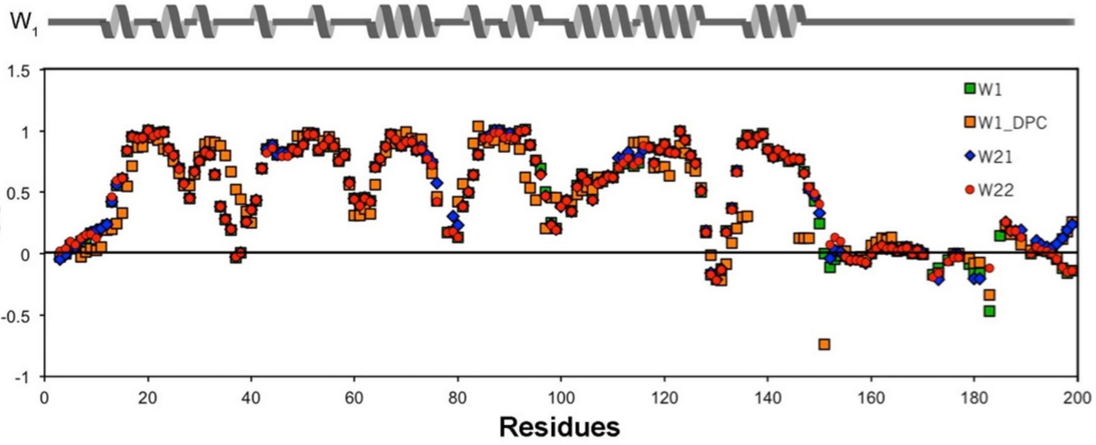
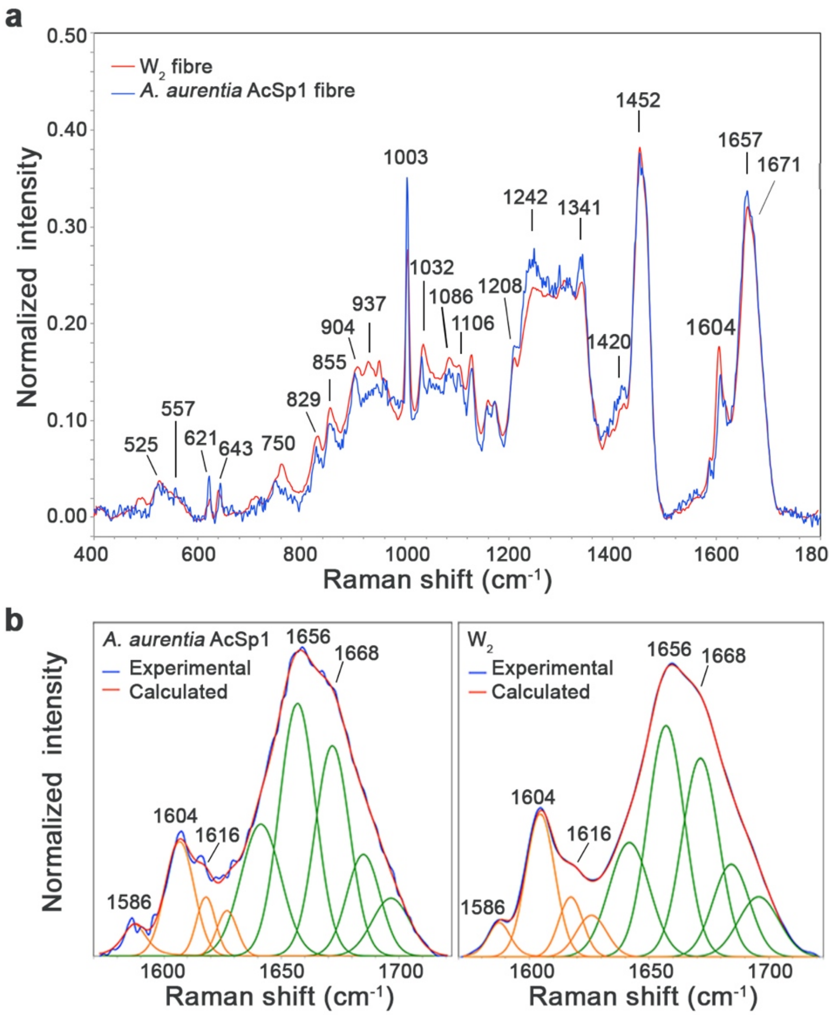


Figure 66 | SSP analysis of W_1, $W_{2\text{-}1}$, $W_{2\text{-}2}$ and W_1 in DPC with CA, CO, N, and H nuclei. Values above 0 have a propensity for forming helices and below 0 have a propensity for forming β-sheets.



Protein type	Dope		Fibre		
	W_2	*N. clavipes*	W_2	*A. aurentia*	*N. clavipes*
α-helices (%)	39	49	32	33	24
β-sheets (%)	0	0	28	27	30

Figure 67 | Protein secondary structure comparison of native *Argiope aurantia* aciniform silk fibre to fibres drawn from W_2 solution. (a) Overlaid orientation-insensitive Raman spectra (Lefèvre et al. 2006) (averaged over multiple fibres/positions of the indicated type of aciniform silk fibre. (b) Spectral decomposition (Lefèvre et al. 2007) of the indicated amide I band. (c) Comparison of secondary structure: W_2 in dope is based on the W_1 NMR ensemble in NMR buffer (recall that W_2 in NMR buffer is fibre-forming competent); all other proportions are based upon amide I band Raman spectral decomposition (with ±3% accuracy (Lefèvre et al. 2007)), with *N. clavipes* data previously published (Lefèvre et al. 2011). The amide I decomposition component at 1656 cm^{-1} was assigned to α-helices; that at 1668 cm^{-1} to β-sheets. The α-helical and β-sheet content were calculated from the ratio of the area of the given band to the total amide I band area. *Data acquired and figure made by Theirry Lefèvre.*

I simulated the putative partially denatured W_1 intermediate with an unfolded H5 by calculating structures with Xplor-NIH (Schwieters et al. 2006) where I removed all NOE and dihedral angle restraints from 135-155 to remove the driving force for H5 folding. The hope was to mimic the unfolding events mentioned in this chapter. Briefly, my attempt is by no means the *real* structure of the intermediate state of W_1, but serves as a representation of the possible events that might subsequently happen upon denaturation (Figure 68). First and foremost, this strategy was successful, with H5 being completely unfolded in all 20 lowest energy structures. The orientation and conformation of that formerly helical segment varies from conformer to conformer. Some restraints, around residue ~180, have NOE contacts with residues ~17-21, so these keep the C-terminal trail from flailing away form the domain. The rest of W_1 retains its folded state.

With the absence of H5, W_1 became a very loose structure with "holes" spanning the protein. The surface hydrophobicity is also changed (Figure 69). W_1 in the native state is fairly evenly distributed with equal amounts of hydrophilic and hydrophobic surface residues as examined according to the Kyte-Doolittle hydrophobicity scale (Kyte and Doolittle 1982). The intermediate state of W_1 fits the description of a molten globule. Removal of H5 exposes the hydrophobic interior of the protein and increases surface accessibly to hydrophobic moieties, while retaining original secondary structuring. The accessible hydrophobic surface area increases from 1453 $Å^2$ in the native state to 1710 $Å^2$ in the intermediate state by summing the most hydrophobic area the amino acids L, V, I, A, and F (more hydrophobic than glycine on the Kyte-Doolittle scale (Kyte and Doolittle 1982)). This is most obvious in the bottom left structure in the intermediate state (Figure 69 arrow), compared to the native structure, where H5 blocks that hydrophobic patch. Given that the hydrophobic effect is a major driving force for protein folding, exposing the hydrophobic core in the intermediate state of W_1 would be in accordance to general principles of protein destabilization (Dill and Shortle 1991; Tsai et al. 2002; Kamatari et al. 2004).

What has been presented thus far provides insight into an unfolding pathway that may or may not be linked to the pathway that AcSp1 goes through during fibrillogenesis. However, this route clearly represents one highly plausible mechanism for fibre initiation for structural transition in fibre formation. It is impossible to hypothesize a correct pathway for fibre formation until we obtain a glimpse of the final atomic resolution structure in the fibre. For now, the pathway I propose is highly promising given that it promotes solvent exposure of the hydrophobic core of W, facilitating protein-protein association in fibrillogenesis. With this knowledge, I believe that future work must aim to elucidate the function of H5 in fibre formation.

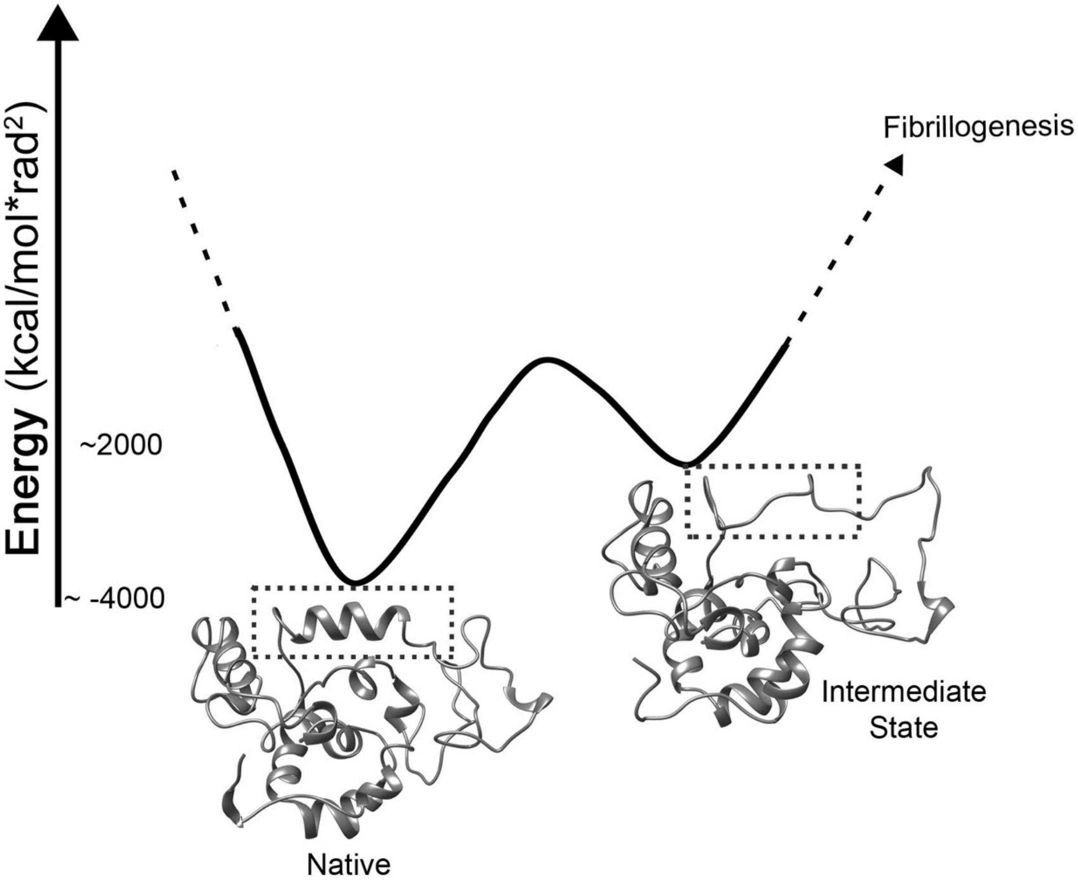


Figure 68 | Hypothetical energy landscape of the conversion of the native state of W_1 to fibre formation if more W domains were present. The energies are the Xplor-NIH (Schwieters et al. 2006) output for the respective structure calculation for the 20 lowest energy members. The intermediate state was calculated with Xplor-NIH (Schwieters et al. 2006) by removing restraints associated to residues 135-158 from the distance restraint file of W_1. The dotted boxes demonstrate the disappearance of the H5. Both structures are in the same orientation.

6.5. SUMMARY

Titrations with DPC, GdmCl, urea, and pressure presented in this chapter all point toward a clear conclusion: the W_1 denaturation pathway includes an intermediate where the less structurally stable H5 unfolds and exposes the region underneath it from residues 25-35, which in turns perturbs the central helix from residues 90-95 (Figure 65). In this study, I used a variety of perturbants to probe the unfolding pathway of W_1, with the goal of gaining insight into a possible route for fibrillogenesis. Structure calculations were then performed to simulate an intermediate state upon the basis of the disappearance of H5. When H5 unfolds, it exposes normally buried hydrophobic regions of W_1 (Figure 69). In an aqueous environment, this would likely enhance fibrillogenesis by inducing protein-protein association. Given that W_2 behaves like W_1 in every single aspect of this thesis, it can only be assumed that events similar to W_1 unfolding (beyond DPC interactions, which I proved) would occur, even in larger fibre-forming W domains concatenation. In summary, in this chapter, I have demonstrated a highly plausible pathway for W protein structural transition that exposes a hydrophobic patch and would promote protein-protein association in fibre formation.

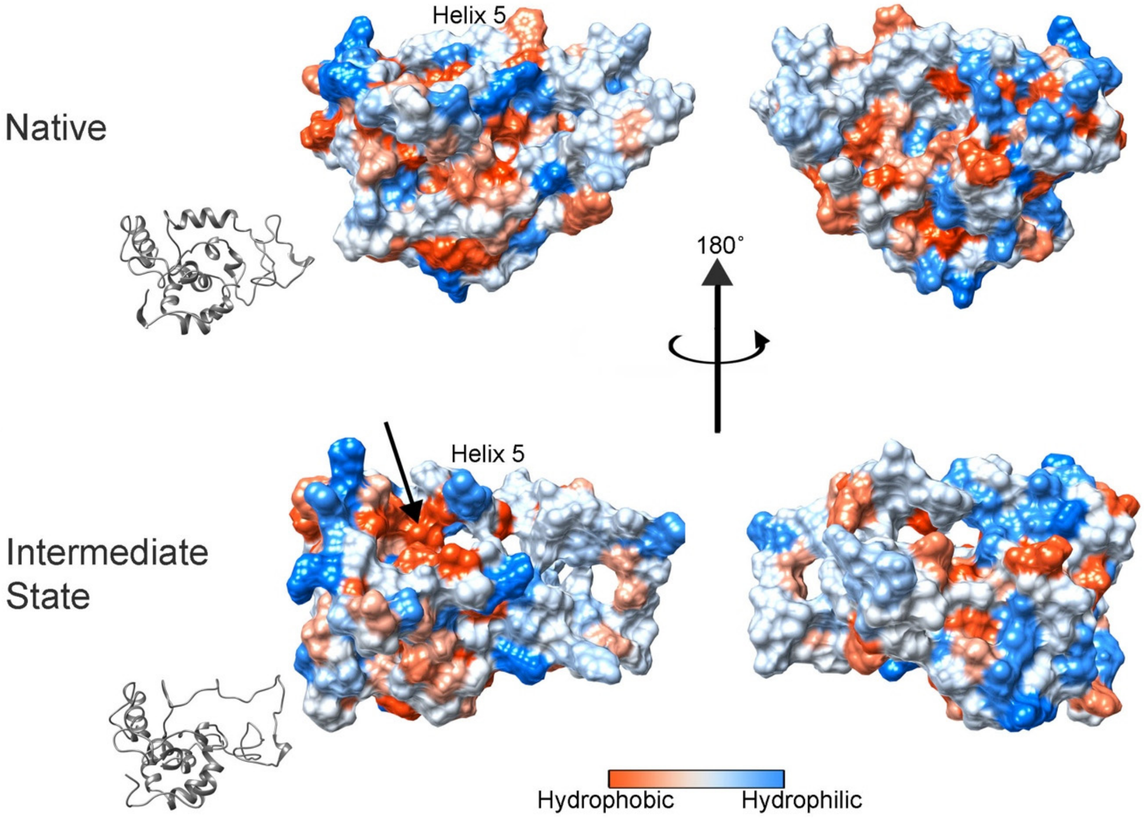


Figure 69 | Comparing the Kyte-Doolittle hydrophathy scale between the native and intermediate state of W_1. In grey are the ribbon structures W_1 in the native and intermediate conformation displayed as the surface on the left. Unfolding of the 5th helix reveals the hydrophobic interior of W_1.

CHAPTER 7. CONCLUSIONS

Silks are strong, tough, and biodegradable material with applications in electronics (Steven et al. 2011); in medicine as carrier particles for drug delivery (Lammel et al. 2011; Hofer et al. 2012, Florczak et al. 2014, Wendt et al. 2011); and for cell scaffolding and tissue engineering (MacIntosh et al. 2008; Zhang et al. 2008; Lu et al. 2011). Despite the remarkable mechanical properties of silk, little is known about its atomic-level structure. My PhD thesis served to shed light on the structure of the toughest spider silk, wrapping silk from the species *Argiope trifasciata*. Thus far, biomolecular NMR has played the most important role in the structural characterization of spider silks at the molecular and atomic level (Zhao and Asakura 2001; Asakura et al. 2013a; Asakura et al. 2013b).

Biomolecular nuclear magnetic resonance (NMR) spectroscopy frequently employs secondary chemical shifts to estimate local secondary structuring in advance of obtaining a full 3D structure. I first set out to obtain two new sets of random coil chemical shifts (RCCS), one derived in DMSO and the other in the popular membrane mimetic solvent chloroform:methanol:water (1:4:4). Extensive analysis was performed to determined the effect of solvent dielectric on protein NMR chemical shifts by assessing the accuracy of the Chemical Shift Index (CSI) (Wishart et al. 1992) with various sets of RCCS. Most of the previously published RCCS data sets were obtained in aqueous or polar solvents, rather than lower dielectric solvents potentially more relevant to membrane proteins or to the interior of a protein fibre. Through design and development of a Python program, CS-CHEMeleon, I evaluated the effect of changing solvent environment in which RCCS are deterimined upon the ability to correctly assess protein secondary structure for a large, non-redundant set of non-membrane proteins and for all membrane proteins available in 2009 in the PDBTM.

Major differences were observed between RCCS in different solvent environments that surpassed the CSI structural thresholds of 0.1 ppm for ^{1}H and 0.7 ppm for ^{13}C, but the ability to assess secondary structure with different data sets was generally and surprisingly similar. Noticeably, $\Delta\delta$ predictive accuracy seems much more dependent on the type of secondary structure (with helices being the most accurately predicted structure) than solvent environment. I hence concluded that great care must be taken when using this class of experimental restraint.

I then set out to solve the first solution-state structure of the full 200 amino acid repetitive unit of AcSp1 of *Argiope trifasciata* in solution. This is key to understanding the starting-point for aciniform silk fibrillogenesis. W_1, the 200 amino acid repetitive unit of AcSp1 was enriched with ^{13}C and/or ^{15}N isotopes through recombinant over-production in *E.coli*. Chemical shifts were sequentially assigned, followed by the assignment of NOE restraints. Due to the highly ambiguous nature of the NOE data set, structure calculations were initially performed with ARIA2 (Rieping et al. 2007) for NOE restraint filtering. The final structural ensemble was calculated with Xplor-NIH (Schwieters et al. 2006) using NOE distance restraints filtered from ARIA2, TALOS+ (Shen et al. 2009a) dihedral angles, and H/D exchange derived H-bonds. The 20-member structural ensemble shows that W_1 is composed of a 5-helix globular core with intrinsically disordered N- and C-terminal tails.

To test if the disorder of N- and C-terminal tails was an artefact of protein truncation, NMR experiments were performed on the selectively-labeled fibre-forming W_2. This was achieved by my colleague Dr. Lingling Xu using split-intein technology. I was therefore able to unambiguously study the effect of tandemerization on the repetitive domain in W_2 where only one of the two W repetitive units is ^{13}C and/or ^{15}N labeled. I demonstrated that the chemical shifts were identical for both domains in W_2, with exception of the residues in the termini of W_1 vs. at the connection point between repeat units in W_2. This proved that the flexibility observed in the termini of W_1 was not a result of protein truncation but, rather, due to intrinsic disorder, regardless of repeat unit concatenation. Reduced spectral density mapping analysis further supported W unit independence in W_2. The motional sampling in the folded core of W_1 was very similar to that observed in both folded domains in W_2, supporting a model of structural architecture with a pair of individual, non-interacting domains separated by a flexible linker. On the basis of these data, I proposed a model for the soluble structure of native AcSp1 composed of a “beads-on-a-string” architecture, where the “beads” are the folded domains and the “strings” are the linkers.

The shape and degree of compactness of W_2 and W_3 was probed by incorporating hydrodynamics data into structure calculations. Structural models of W_2 and W_3 were initially produced using Xplor-NIH through concatenation of W_1 NOE, dihedral, and H-bond restraints. In a direct comparison, structural ensembles were calculated with an additional radius of gyration restraint, determined from hydrodynamics data. These latter ensembles showed better agreement in hydrodynamics behaviour without violation of the other experimental restraints, allowing me to show that W_2 and W_3 prefer to adopt a compact structure, where both domains are close in space, rather than far apart. This approach will be readily extendable to a higher domain concatenation.

Investigating the fibrillogenesis mechanism and the ultimate structure of AcSp1 within a fibre were also goals of this project. The first step in the conversion of the soluble protein to the fibrous state seems likely to be seeded at H5, given that this helix is the most readily perturbed portion of the protein in the presence of denaturants, detergent and under high pressure. In support of this model, I also found that the H5 backbone amide N-H groups were the most readily exchanged with deuterium and that SSP analysis of chemical shift demonstrates that the DPC-denatured H5 region shows β-sheet propensity.

Since residues ~135-160 appear to be the most structurally plastic, I suggest that future work should focus on the investigation of the aciniform silk fibrous state by ssNMR spectroscopy as a method to elucidate β-sheet location within the repeat unit sequence. One of the limitations in ssNMR during my thesis work was our inability to efficiently produce enough fibres for a sample. Developing the capability to produce large amounts of fibres with NMR-active isotope labels is therefore a critical precursor to these suggested studies.

In conclusion, I have presented the most detailed characterization of aciniform silk to date, a material with strong potential value for a wide variety of biomedical and biomaterial applications. A detailed structural and biophysical understanding of this class of silk significantly enhances our knowledge of and ability to engineer spider silks, and provides the groundwork for future understanding of the aciniform silk structure within the fibre.

APPENDIX A. LIST OF PDBS AND BMRB FILES USED FOR THE ANALYSIS OF SECONDARY CHEMICAL SHIFTS FROM CHAPTER 2

Table 1. Matched PDB_TM and BMRB entries containing $^{1}H^{\alpha}$ in the TM dataset.

PDB_TM	BMRB
1AFO	BMR7208
1BCT	BMR1162
1JO5	BMR4303
1L6T	BMR5326
1MOT	BMR5607
1RKL	BMR6056
1WU0	BMR6489
1XRD	BMR6349
1Z65	BMR6598
1ZZA	BMR6715
2CPB	BMR4197
2CPB	BMR4209
2HTG	BMR7245
2JO1	BMR16168
2JTW	BMR11056
2K1K	BMR15728
2K3C	BMR15747
2K9P	BMR15995
2KA1	BMR16012
2KBV	BMR16056

Table 2. Matched PDB_TM and BMRB entries containing $^{13}C^{\alpha}$ in the TM dataset.

PDB_TM	BMRB
1AFO	BMR7208
1L6T	BMR5326
1MM4	BMR5557
1MM4	BMR6234

1WU0	BMR6489
1XRD	BMR6349
1ZZA	BMR6715
2CPB	BMR4197
2CPB	BMR4209
2HTG	BMR7245
2JO1	BMR16168
2K0L	BMR15651
2K1K	BMR15728
2K3C	BMR15747
2K73	BMR15966
2K9P	BMR15995
2KA1	BMR16012
2KBV	BMR16056

Table 3. Matched PDB_TM and BMRB entries containing $^{13}C^{\beta}$ in the TM dataset.

PDB_TM	BMRB
1AFO	BMR7208
1L6T	BMR5326
1MM4	BMR5557
1MM4	BMR6234
1WU0	BMR6489
1XRD	BMR6349
2CPB	BMR4197
2CPB	BMR4209
2HTG	BMR7245
2JO1	BMR16168
2K0L	BMR15651
2K1K	BMR15728
2K3C	BMR15747

2K73	BMR15966
2K9P	BMR15995
2KA1	BMR16012
2KBV	BMR16056

Table 4. Matched PDB_TM and BMRB entries containing 2 or more of H^{α}, C^{α} and C^{β} in the TM data set.

PDB_TM	BMRB
1AFO	BMR7208
1L6T	BMR5326
1MM4	BMR5557
1MM4	BMR6234
1WU0	BMR6489
1XRD	BMR6349
1ZZA	BMR6715
2CPB	BMR4197
2CPB	BMR4209
2HTG	BMR7245
2JO1	BMR16168
2K0L	BMR15651
2K1K	BMR15728
2K3C	BMR15747
2K73	BMR15966
2K9P	BMR15995
2KA1	BMR16012
2KBV	BMR16056

Table 5. Matched PDB and BMRB entries containing $^{1}H^{\alpha}$ in the AQ dataset.

PDB	BMRB	PDB	BMRB	PDB	BMRB
1BV2	BMR4917	1TDP	BMR6211	2IZ3	BMR15037
1BW5	BMR4121	1UEP	BMR10100	2JMB	BMR15012
1C06	BMR4577	1UJV	BMR10124	2JMK	BMR15039
1C89	BMR4449	1VEX	BMR10002	2JN4	BMR15085
1CB3	BMR4483	1WFG	BMR10083	2JNP	BMR15120
1CB9	BMR4419	1WFT	BMR10008	2JOR	BMR15192
1CKV	BMR4431	1WFU	BMR10006	2JPU	BMR15265
1CKW	BMR4595	1WFW	BMR10009	2JQM	BMR15034
1CKX	BMR4597	1YCM	BMR6444	2JRM	BMR15339
1D0R	BMR4741	1YV8	BMR6455	2JS1	BMR15350
1E8L	BMR1093	1YWU	BMR6514	2JUO	BMR15451
1EZO	BMR4987	1YZA	BMR6553	2JVL	BMR15795
1FA4	BMR5858	1YZC	BMR6552	2JWK	BMR15459
1G9L	BMR4915	1Z0Q	BMR6554	2JY8	BMR15592
1GH9	BMR4740	1Z9B	BMR6577	2K1S	BMR15683
1H4B	BMR5707	2AAS	BMR443	2K5K	BMR15840
1H95	BMR5076	2B3W	BMR6782	2K5Q	BMR15846
1HOM	BMR1037	2COM	BMR10011	2K7H	BMR5605
1IE5	BMR5044	2CRD	BMR114	2KAX	BMR16033
1ILO	BMR4991	2D1U	BMR6803	2KAY	BMR16034
1ITY	BMR5361	2E0G	BMR10027	2KB1	BMR15677
1JEI	BMR5074	2E29	BMR10213	2KC7	BMR16064
1JQ4	BMR5148	2E5T	BMR11000	2KD2	BMR16103
1KZS	BMR5283	2E8D	BMR10022	2KDC	BMR15019
1LD6	BMR5375	2E9G	BMR10141	2KE5	BMR15230
1LKN	BMR5357	2E9I	BMR10142	2KHF	BMR16239
1MPE	BMR5654	2EC7	BMR15364	2KJ5	BMR16312
1N91	BMR5596	2EE4	BMR10144	2KL1	BMR16378
1NG7	BMR5753	2EE7	BMR10147	2KLC	BMR16390

1NMR	BMR5698	2EE9	BMR10159	2OMJ	BMR15168
1O7B	BMR6393	2EEA	BMR10160	2PXG	BMR6797
1OMT	BMR5518	2EN6	BMR10151	2RML	BMR5819
1ON4	BMR5742	2EOF	BMR10216	2RMO	BMR11026
1OQ6	BMR5768	2EOS	BMR10157	2RN7	BMR11017
1P68	BMR5687	2EOY	BMR10154	2RN9	BMR11019
1Q8K	BMR10023	2F40	BMR7073	2RNJ	BMR11024
1QXF	BMR5682	2FFT	BMR6926	2RO1	BMR11036
1R36	BMR5991	2FRW	BMR7035	2RO5	BMR11034
1R57	BMR5845	2FYH	BMR10004	2RQ1	BMR11065
1R8P	BMR5952	2G31	BMR7067	2YSE	BMR10214
1RFH	BMR6059	2G9L	BMR7065	2YTS	BMR10156
1RL5	BMR5989	2GBS	BMR7004	2ZAJ	BMR10215
1SRB	BMR1811	2HKY	BMR7206		
1SRZ	BMR6171	2HWT	BMR7112		

Table 6. Matched PDB and BMRB entries containing $^{13}C^{\alpha}$ in the AQ dataset.

PDB	BMRB	PDB	BMRB	PDB	BMRB
1BV2	BMR4917	1TDP	BMR6211	2IZ3	BMR15037
1BW5	BMR4121	1UEP	BMR10100	2JMB	BMR15012
1C06	BMR4577	1UJV	BMR10124	2JMK	BMR15039
1C89	BMR4449	1VEX	BMR10002	2JN4	BMR15085
1CB3	BMR4483	1WFG	BMR10083	2JNP	BMR15120
1CB9	BMR4419	1WFT	BMR10008	2JOR	BMR15192
1CKV	BMR4431	1WFU	BMR10006	2JPU	BMR15265
1CKW	BMR4595	1WFW	BMR10009	2JQM	BMR15034
1CKX	BMR4597	1YCM	BMR6444	2JRM	BMR15339
1D0R	BMR4741	1YV8	BMR6455	2JS1	BMR15350
1E8L	BMR1093	1YWU	BMR6514	2JUO	BMR15451
1EZO	BMR4987	1YZA	BMR6553	2JVL	BMR15795
1FA4	BMR5858	1YZC	BMR6552	2JWK	BMR15459

1G9L	BMR4915	1Z0Q	BMR6554	2JY8	BMR15592
1GH9	BMR4740	1Z9B	BMR6577	2K1S	BMR15683
1H4B	BMR5707	2AAS	BMR443	2K5K	BMR15840
1H95	BMR5076	2B3W	BMR6782	2K5Q	BMR15846
1HOM	BMR1037	2COM	BMR10011	2K7H	BMR5605
1IE5	BMR5044	2CRD	BMR114	2KAX	BMR16033
1ILO	BMR4991	2D1U	BMR6803	2KAY	BMR16034
1ITY	BMR5361	2E0G	BMR10027	2KB1	BMR15677
1JEI	BMR5074	2E29	BMR10213	2KC7	BMR16064
1JQ4	BMR5148	2E5T	BMR11000	2KD2	BMR16103
1KZS	BMR5283	2E8D	BMR10022	2KDC	BMR15019
1LD6	BMR5375	2E9G	BMR10141	2KE5	BMR15230
1LKN	BMR5357	2E9I	BMR10142	2KHF	BMR16239
1MPE	BMR5654	2EC7	BMR15364	2KJ5	BMR16312
1N91	BMR5596	2EE4	BMR10144	2KL1	BMR16378
1NG7	BMR5753	2EE7	BMR10147	2KLC	BMR16390
1NMR	BMR5698	2EE9	BMR10159	2OMJ	BMR15168
1O7B	BMR6393	2EEA	BMR10160	2PXG	BMR6797
1OMT	BMR5518	2EN6	BMR10151	2RML	BMR5819
1ON4	BMR5742	2EOF	BMR10216	2RMO	BMR11026
1OQ6	BMR5768	2EOS	BMR10157	2RN7	BMR11017
1P68	BMR5687	2EOY	BMR10154	2RN9	BMR11019
1Q8K	BMR10023	2F40	BMR7073	2RNJ	BMR11024
1QXF	BMR5682	2FFT	BMR6926	2RO1	BMR11036
1R36	BMR5991	2FRW	BMR7035	2RO5	BMR11034
1R57	BMR5845	2FYH	BMR10004	2RQ1	BMR11065
1R8P	BMR5952	2G31	BMR7067	2YSE	BMR10214
1RFH	BMR6059	2G9L	BMR7065	2YTS	BMR10156
1RL5	BMR5989	2GBS	BMR7004	2ZAJ	BMR10215
1SRB	BMR1811	2HKY	BMR7206		
1SRZ	BMR6171	2HWT	BMR7112		

Table 7. Matched PDB and BMRB entries containing $^{13}C^{\beta}$ in the AQ dataset.

PDB	BMRB	PDB	BMRB	PDB	BMRB
1BV2	BMR4917	1TDP	BMR6211	2IZ3	BMR15037
1BW5	BMR4121	1UEP	BMR10100	2JMB	BMR15012
1C06	BMR4577	1UJV	BMR10124	2JMK	BMR15039
1C89	BMR4449	1VEX	BMR10002	2JN4	BMR15085
1CB3	BMR4483	1WFG	BMR10083	2JNP	BMR15120
1CB9	BMR4419	1WFT	BMR10008	2JOR	BMR15192
1CKV	BMR4431	1WFU	BMR10006	2JPU	BMR15265
1CKW	BMR4595	1WFW	BMR10009	2JQM	BMR15034
1CKX	BMR4597	1YCM	BMR6444	2JRM	BMR15339
1D0R	BMR4741	1YV8	BMR6455	2JS1	BMR15350
1E8L	BMR1093	1YWU	BMR6514	2JUO	BMR15451
1EZO	BMR4987	1YZA	BMR6553	2JVL	BMR15795
1FA4	BMR5858	1YZC	BMR6552	2JWK	BMR15459
1G9L	BMR4915	1Z0Q	BMR6554	2JY8	BMR15592
1GH9	BMR4740	1Z9B	BMR6577	2K1S	BMR15683
1H4B	BMR5707	2AAS	BMR443	2K5K	BMR15840
1H95	BMR5076	2B3W	BMR6782	2K5Q	BMR15846
1HOM	BMR1037	2COM	BMR10011	2K7H	BMR5605
1IE5	BMR5044	2CRD	BMR114	2KAX	BMR16033
1ILO	BMR4991	2D1U	BMR6803	2KAY	BMR16034
1ITY	BMR5361	2E0G	BMR10027	2KB1	BMR15677
1JEI	BMR5074	2E29	BMR10213	2KC7	BMR16064
1JQ4	BMR5148	2E5T	BMR11000	2KD2	BMR16103
1KZS	BMR5283	2E8D	BMR10022	2KDC	BMR15019
1LD6	BMR5375	2E9G	BMR10141	2KE5	BMR15230
1LKN	BMR5357	2E9I	BMR10142	2KHF	BMR16239
1MPE	BMR5654	2EC7	BMR15364	2KJ5	BMR16312
1N91	BMR5596	2EE4	BMR10144	2KL1	BMR16378
1NG7	BMR5753	2EE7	BMR10147	2KLC	BMR16390

1NMR	BMR5698	2EE9	BMR10159	2OMJ	BMR15168
1O7B	BMR6393	2EEA	BMR10160	2PXG	BMR6797
1OMT	BMR5518	2EN6	BMR10151	2RML	BMR5819
1ON4	BMR5742	2EOF	BMR10216	2RMO	BMR11026
1OQ6	BMR5768	2EOS	BMR10157	2RN7	BMR11017
1P68	BMR5687	2EOY	BMR10154	2RN9	BMR11019
1Q8K	BMR10023	2F40	BMR7073	2RNJ	BMR11024
1QXF	BMR5682	2FFT	BMR6926	2RO1	BMR11036
1R36	BMR5991	2FRW	BMR7035	2RO5	BMR11034
1R57	BMR5845	2FYH	BMR10004	2RQ1	BMR11065
1R8P	BMR5952	2G31	BMR7067	2YSE	BMR10214
1RFH	BMR6059	2G9L	BMR7065	2YTS	BMR10156
1RL5	BMR5989	2GBS	BMR7004	2ZAJ	BMR10215
1SRB	BMR1811	2HKY	BMR7206		
1SRZ	BMR6171	2HWT	BMR7112		

Table 8. Matched PDB and BMRB entries containing at least 2 of H^{α}, C^{α} and C^{β} in the AQ dataset.

PDB	BMRB	PDB	BMRB	PDB	BMRB
1BW5	BMR4121	2COM	BMR10011	2KB1	BMR15677
1C06	BMR4577	2D1U	BMR6803	2KC7	BMR16064
1C89	BMR4449	2E0G	BMR10027	2KD2	BMR16103
1CKV	BMR4431	2E29	BMR10213	2KDC	BMR15019
1EZO	BMR4987	2E5T	BMR11000	2KE5	BMR15230
1G9L	BMR4915	2E8D	BMR10022	2KHF	BMR16239
1GH9	BMR4740	2E9G	BMR10141	2KJ5	BMR16312
1H4B	BMR5707	2E9I	BMR10142	2KL1	BMR16378
1H95	BMR5076	2EC7	BMR15364	2KLC	BMR16390
1IE5	BMR5044	2EE4	BMR10144	2OMJ	BMR15168
1ILO	BMR4991	2EE7	BMR10147	2PXG	BMR6797
1ITY	BMR5361	2EE9	BMR10159	2RML	BMR5819
1JQ4	BMR5148	2EEA	BMR10160	2RMO	BMR11026
1LKN	BMR5357	2EN6	BMR10151	2RN7	BMR11017
1MPE	BMR5654	2EOF	BMR10216	2RN9	BMR11019
1N91	BMR5596	2EOS	BMR10157	2RNJ	BMR11024
1NG7	BMR5753	2EOY	BMR10154	2RO1	BMR11036
1O7B	BMR6393	2F40	BMR7073	2RO5	BMR11034
1ON4	BMR5742	2FFT	BMR6926	2RQ1	BMR11065
1P68	BMR5687	2FRW	BMR7035	2YSE	BMR10214
1Q8K	BMR10023	2FYH	BMR10004	2YTS	BMR10156
1QXF	BMR5682	2GBS	BMR7004	2ZAJ	BMR10215
1R36	BMR5991	2HKY	BMR7206		
1R57	BMR5845	2HWT	BMR7112		
1R8P	BMR5952	2JMB	BMR15012		
1RFH	BMR6059	2JMK	BMR15039		
1SRZ	BMR6171	2JN4	BMR15085		
1TDP	BMR6211	2JNP	BMR15120		

1UEP	BMR10100	2JOR	BMR15192
1UJV	BMR10124	2JPU	BMR15265
1VEX	BMR10002	2JQM	BMR15034
1WFG	BMR10083	2JRM	BMR15339
1WFT	BMR10008	2JS1	BMR15350
1WFU	BMR10006	2JUO	BMR15451
1WFW	BMR10009	2JVL	BMR15795
1YCM	BMR6444	2JWK	BMR15459
1YV8	BMR6455	2JY8	BMR15592
1YWU	BMR6514	2K1S	BMR15683
1YZA	BMR6553	2K5K	BMR15840
1YZC	BMR6552	2K5Q	BMR15846
1Z0Q	BMR6554	2K7H	BMR5605
1Z9B	BMR6577	2KAX	BMR16033
2B3W	BMR6782	2KAY	BMR16034

APPENDIX B. W$_1$ *J*-COUPLING VALUES

B.1. J_{HH_α} HNHA MENTIONED IN CHAPTER 3

```
assign  ( resid  4    and name HN  ) ( resid  4    and name N   )
           ( resid  4    and name CA ) ( resid  4    and name HA   )  6.48282   0.09226
assign  ( resid  10    and name HN  ) ( resid  10    and name N   )
         ( resid  10    and name CA ) ( resid  10    and name HA   )  7.03116   0.08613
assign  ( resid  17    and name HN  ) ( resid  17    and name N   )
         ( resid  17    and name CA ) ( resid  17    and name HA   )  4.21612   0.66364
assign  ( resid  18    and name HN  ) ( resid  18    and name N   )
         ( resid  18    and name CA ) ( resid  18    and name HA   )  3.20451   1.05107
assign  ( resid  20    and name HN  ) ( resid  20    and name N   )
         ( resid  20    and name CA ) ( resid  20    and name HA   )  4.33965   0.65881
assign  ( resid  21    and name HN  ) ( resid  21    and name N   )
         ( resid  21    and name CA ) ( resid  21    and name HA   )  3.73983   0.78001
assign  ( resid  23    and name HN  ) ( resid  23    and name N   )
         ( resid  23    and name CA ) ( resid  23    and name HA   )  3.21533   0.77439
assign  ( resid  24    and name HN  ) ( resid  24    and name N   )
           ( resid  24    and name CA ) ( resid  24    and name HA   )  3.44627   0.72319
assign  ( resid  25    and name HN  ) ( resid  25    and name N   )
           ( resid  25    and name CA ) ( resid  25    and name HA   )  3.97430   1.05731
assign  ( resid  27    and name HN  ) ( resid  27    and name N   )
           ( resid  27    and name CA ) ( resid  27    and name HA   )  6.84334   0.39894
assign  ( resid  28    and name HN  ) ( resid  28    and name N   )
           ( resid  28    and name CA ) ( resid  28    and name HA   )  6.80058   0.36326
assign  ( resid  29    and name HN  ) ( resid  29    and name N   )
           ( resid  29    and name CA ) ( resid  29    and name HA   )  6.04207   1.54615
assign  ( resid  32    and name HN  ) ( resid  32    and name N   )
           ( resid  32    and name CA ) ( resid  32    and name HA   )  3.97311   0.98597
assign  ( resid  33    and name HN  ) ( resid  33    and name N   )
           ( resid  33    and name CA ) ( resid  33    and name HA   )  5.80582   0.67527
assign  ( resid  36    and name HN  ) ( resid  36    and name N   )
           ( resid  36    and name CA ) ( resid  36    and name HA   )  6.54174   0.81623
assign  ( resid  41    and name HN  ) ( resid  41    and name N   )
           ( resid  41    and name CA ) ( resid  41    and name HA   )  3.13129   1.30750
assign  ( resid  43    and name HN  ) ( resid  43    and name N   )
           ( resid  43    and name CA ) ( resid  43    and name HA   )  5.09535   0.60488
assign  ( resid  46    and name HN  ) ( resid  46    and name N   )
           ( resid  46    and name CA ) ( resid  46    and name HA   )  3.85073   0.55650
assign  ( resid  47    and name HN  ) ( resid  47    and name N   )
           ( resid  47    and name CA ) ( resid  47    and name HA   )  2.75461   1.41568
assign  ( resid  48    and name HN  ) ( resid  48    and name N   )
           ( resid  48    and name CA ) ( resid  48    and name HA   )  2.25736   0.69421
assign  ( resid  51    and name HN  ) ( resid  51    and name N   )
           ( resid  51    and name CA ) ( resid  51    and name HA   )  5.35736   1.05453
```

```
assign  ( resid  53    and name HN  ) ( resid  53    and name N   )
        ( resid  53    and name CA ) ( resid  53    and name HA   )  3.64922   0.64633
assign  ( resid  54    and name HN  ) ( resid  54    and name N   )
        ( resid  54    and name CA ) ( resid  54    and name HA   )  3.40819   0.40769
assign  ( resid  57    and name HN  ) ( resid  57    and name N   )
        ( resid  57    and name CA ) ( resid  57    and name HA   )  3.71266   0.42767
assign  ( resid  61    and name HN  ) ( resid  61    and name N   )
        ( resid  61    and name CA ) ( resid  61    and name HA   )  7.84452   0.28124
assign  ( resid  64    and name HN  ) ( resid  64    and name N   )
        ( resid  64    and name CA ) ( resid  64    and name HA   )  3.68104   0.25400
assign  ( resid  65    and name HN  ) ( resid  65    and name N   )
        ( resid  65    and name CA ) ( resid  65    and name HA   )  4.25700   0.40973
assign  ( resid  70    and name HN  ) ( resid  70    and name N   )
        ( resid  70    and name CA ) ( resid  70    and name HA   )  3.84738   0.63012
assign  ( resid  71    and name HN  ) ( resid  71    and name N   )
        ( resid  71    and name CA ) ( resid  71    and name HA   )  3.39504   0.86971
assign  ( resid  73    and name HN  ) ( resid  73    and name N   )
        ( resid  73    and name CA ) ( resid  73    and name HA   )  3.91748   0.54349
assign  ( resid  76    and name HN  ) ( resid  76    and name N   )
        ( resid  76    and name CA ) ( resid  76    and name HA   )  7.50020   0.39152
assign  ( resid  77    and name HN  ) ( resid  77    and name N   )
        ( resid  77    and name CA ) ( resid  77    and name HA   )  4.95769   0.69334
assign  ( resid  79    and name HN  ) ( resid  79    and name N   )
        ( resid  79    and name CA ) ( resid  79    and name HA   )  2.65075   1.09584
assign  ( resid  84    and name HN  ) ( resid  84    and name N   )
        ( resid  84    and name CA ) ( resid  84    and name HA   )  2.92009   1.06694
assign  ( resid  85    and name HN  ) ( resid  85    and name N   )
        ( resid  85    and name CA ) ( resid  85    and name HA   )  4.55807   0.59097
assign  ( resid  92    and name HN  ) ( resid  92    and name N   )
        ( resid  92    and name CA ) ( resid  92    and name HA   )  2.63545   1.28104
assign  ( resid  93    and name HN  ) ( resid  93    and name N   )
        ( resid  93    and name CA ) ( resid  93    and name HA   )  2.47773   0.57905
assign  ( resid  96    and name HN  ) ( resid  96    and name N   )
        ( resid  96    and name CA ) ( resid  96    and name HA   )  3.57821   1.09353
assign  ( resid  99    and name HN  ) ( resid  99    and name N   )
        ( resid  99    and name CA ) ( resid  99    and name HA   )  6.20246   0.57214
assign  ( resid  100    and name HN  ) ( resid  100    and name N   )
        ( resid  100    and name CA ) ( resid  100    and name HA   )  9.01609   0.73741
assign  ( resid  101    and name HN  ) ( resid  101    and name N   )
        ( resid  101    and name CA ) ( resid  101    and name HA   )  7.05308   0.49921
assign  ( resid  102    and name HN  ) ( resid  102    and name N   )
        ( resid  102    and name CA ) ( resid  102    and name HA   )  2.48927   1.68690
assign  ( resid  103    and name HN  ) ( resid  103    and name N   )
        ( resid  103    and name CA ) ( resid  103    and name HA   )  4.66742   0.21084
assign  ( resid  107    and name HN  ) ( resid  107    and name N   )
        ( resid  107    and name CA ) ( resid  107    and name HA   )  9.39838   0.30618
assign  ( resid  110    and name HN  ) ( resid  110    and name N   )
```

```
          ( resid  110    and name CA ) ( resid  110    and name HA   )  2.99525   0.85918
assign  ( resid  111    and name HN  ) ( resid  111    and name N   )
          ( resid  111    and name CA ) ( resid  111    and name HA   )  5.19829   0.60705
assign  ( resid  115    and name HN  ) ( resid  115    and name N   )
          ( resid  115    and name CA ) ( resid  115    and name HA   )  4.08630   0.52927
assign  ( resid  120    and name HN  ) ( resid  120    and name N   )
          ( resid  120    and name CA ) ( resid  120    and name HA   )  4.65351   1.20280
assign  ( resid  127    and name HN  ) ( resid  127    and name N   )
          ( resid  127    and name CA ) ( resid  127    and name HA   )  7.25615   0.43701
assign  ( resid  130    and name HN  ) ( resid  130    and name N   )
          ( resid  130    and name CA ) ( resid  130    and name HA   )  4.50742   0.23111
assign  ( resid  132    and name HN  ) ( resid  132    and name N   )
          ( resid  132    and name CA ) ( resid  132    and name HA   )  5.62588   0.35731
assign  ( resid  135    and name HN  ) ( resid  135    and name N   )
          ( resid  135    and name CA ) ( resid  135    and name HA   )  5.21272   0.28512
assign  ( resid  136    and name HN  ) ( resid  136    and name N   )
          ( resid  136    and name CA ) ( resid  136    and name HA   )  3.74825   0.82913
assign  ( resid  142    and name HN  ) ( resid  142    and name N   )
          ( resid  142    and name CA ) ( resid  142    and name HA   )  4.98627   0.64822
assign  ( resid  148    and name HN  ) ( resid  148    and name N   )
          ( resid  148    and name CA ) ( resid  148    and name HA   )  5.56705   0.36994
assign  ( resid  158    and name HN  ) ( resid  158    and name N   )
          ( resid  158    and name CA ) ( resid  158    and name HA   )  5.76670   0.12144
assign  ( resid  169    and name HN  ) ( resid  169    and name N   )
          ( resid  169    and name CA ) ( resid  169    and name HA   )  6.47293   0.07053
assign  ( resid  186    and name HN  ) ( resid  186    and name N   )
          ( resid  186    and name CA ) ( resid  186    and name HA   )  6.02891   0.10824
assign  ( resid  190    and name HN  ) ( resid  190    and name N   )
      ( resid  190    and name CA ) ( resid  190    and name HA   )  4.98763   0.11787
```

B.2. J_{NH_β} HNHB MENTIONED IN CHAPTER 3

```
assign  ( resid  17    and name N  ) ( resid  17    and name CA  )
          ( resid  17    and name CB ) ( resid  17    and name HB*  )  2.724  1.00
assign  ( resid  17    and name N  ) ( resid  17    and name CA  )
          ( resid  17    and name CB ) ( resid  17    and name HB*  )  5.773  1.00
assign  ( resid  18    and name N  ) ( resid  18    and name CA  )
          ( resid  18    and name CB ) ( resid  18    and name HB*  )  2.508  1.00
assign  ( resid  20    and name N  ) ( resid  20    and name CA  )
          ( resid  20    and name CB ) ( resid  20    and name HB*  )  3.977  1.00
assign  ( resid  23    and name N  ) ( resid  23    and name CA  )
          ( resid  23    and name CB ) ( resid  23    and name HB*  )  5.437  1.00
assign  ( resid  23    and name N  ) ( resid  23    and name CA  )
          ( resid  23    and name CB ) ( resid  23    and name HB**  )  3.127  1.00
assign  ( resid  24    and name N  ) ( resid  24    and name CA  )
          ( resid  24    and name CB ) ( resid  24    and name HB*  )  6.273  1.00
```

```
assign  ( resid  25    and name N  ) ( resid  25    and name CA  )
           ( resid  25    and name CB ) ( resid  25    and name HB*  )  5.836  1.00
assign  ( resid  25    and name N  ) ( resid  25    and name CA  )
           ( resid  25    and name CB ) ( resid  25    and name HB*  )  3.890  1.00
assign  ( resid  26    and name N  ) ( resid  26    and name CA  )
           ( resid  26    and name CB ) ( resid  26    and name HB*  )  4.319  1.00
assign  ( resid  27    and name N  ) ( resid  27    and name CA  )
           ( resid  27    and name CB ) ( resid  27    and name HB*  )  4.991  1.00
assign  ( resid  27    and name N  ) ( resid  27    and name CA  )
           ( resid  27    and name CB ) ( resid  27    and name HB*  )  3.147  1.00
assign  ( resid  28    and name N  ) ( resid  28    and name CA  )
           ( resid  28    and name CB ) ( resid  28    and name HB*  )  5.536  1.00
assign  ( resid  29    and name N  ) ( resid  29    and name CA  )
           ( resid  29    and name CB ) ( resid  29    and name HB*  )  6.216  1.00
assign  ( resid  32    and name N  ) ( resid  32    and name CA  )
           ( resid  32    and name CB ) ( resid  32    and name HB*  )  4.959  1.00
assign  ( resid  33    and name N  ) ( resid  33    and name CA  )
           ( resid  33    and name CB ) ( resid  33    and name HB*  )  4.111  1.00
assign  ( resid  36    and name N  ) ( resid  36    and name CA  )
           ( resid  36    and name CB ) ( resid  36    and name HB*  )  6.529  1.00
assign  ( resid  39    and name N  ) ( resid  39    and name CA  )
           ( resid  39    and name CB ) ( resid  39    and name HB*  )  3.068  1.00
assign  ( resid  41    and name N  ) ( resid  41    and name CA  )
           ( resid  41    and name CB ) ( resid  41    and name HB*  )  5.563  1.00
assign  ( resid  43    and name N  ) ( resid  43    and name CA  )
           ( resid  43    and name CB ) ( resid  43    and name HB*  )  2.905  1.00
assign  ( resid  44    and name N  ) ( resid  44    and name CA  )
           ( resid  44    and name CB ) ( resid  44    and name HB*  )  5.741  1.00
assign  ( resid  46    and name N  ) ( resid  46    and name CA  )
           ( resid  46    and name CB ) ( resid  46    and name HB*  )  2.749  1.00
assign  ( resid  46    and name N  ) ( resid  46    and name CA  )
           ( resid  46    and name CB ) ( resid  46    and name HB*  )  4.713  1.00
assign  ( resid  47    and name N  ) ( resid  47    and name CA  )
           ( resid  47    and name CB ) ( resid  47    and name HB*  )  3.294  1.00
assign  ( resid  51    and name N  ) ( resid  51    and name CA  )
           ( resid  51    and name CB ) ( resid  51    and name HB*  )  6.120  1.00
assign  ( resid  53    and name N  ) ( resid  53    and name CA  )
           ( resid  53    and name CB ) ( resid  53    and name HB*  )  5.761  1.00
assign  ( resid  57    and name N  ) ( resid  57    and name CA  )
           ( resid  57    and name CB ) ( resid  57    and name HB*  )  4.370  1.00
assign  ( resid  61    and name N  ) ( resid  61    and name CA  )
           ( resid  61    and name CB ) ( resid  61    and name HB*  )  3.827  1.00
assign  ( resid  65    and name N  ) ( resid  65    and name CA  )
           ( resid  65    and name CB ) ( resid  65    and name HB*  )  3.291  1.00
assign  ( resid  65    and name N  ) ( resid  65    and name CA  )
           ( resid  65    and name CB ) ( resid  65    and name HB*  )  4.241  1.00
assign  ( resid  70    and name N  ) ( resid  70    and name CA  )
```

```
          ( resid  70    and name CB ) ( resid  70    and name HB*  )  5.607  1.00
assign  ( resid  71    and name N  ) ( resid  71    and name CA  )
          ( resid  71    and name CB ) ( resid  71    and name HB*  )  3.765  1.00
assign  ( resid  73    and name N  ) ( resid  73    and name CA  )
          ( resid  73    and name CB ) ( resid  73    and name HB*  )  6.594  1.00
assign  ( resid  74    and name N  ) ( resid  74    and name CA  )
          ( resid  74    and name CB ) ( resid  74    and name HB*  )  2.314  1.00
assign  ( resid  76    and name N  ) ( resid  76    and name CA  )
          ( resid  76    and name CB ) ( resid  76    and name HB*  )  6.303  1.00
assign  ( resid  76    and name N  ) ( resid  76    and name CA  )
          ( resid  76    and name CB ) ( resid  76    and name HB*  )  3.702  1.00
assign  ( resid  77    and name N  ) ( resid  77    and name CA  )
          ( resid  77    and name CB ) ( resid  77    and name HB*  )  3.125  1.00
assign  ( resid  77    and name N  ) ( resid  77    and name CA  )
          ( resid  77    and name CB ) ( resid  77    and name HB*  )  5.760  1.00
assign  ( resid  79    and name N  ) ( resid  79    and name CA  )
          ( resid  79    and name CB ) ( resid  79    and name HB*  )  7.046  1.00
assign  ( resid  81    and name N  ) ( resid  81    and name CA  )
          ( resid  81    and name CB ) ( resid  81    and name HB*  )  4.141
assign  ( resid  81    and name N  ) ( resid  81    and name CA  )
          ( resid  81    and name CB ) ( resid  81    and name HB*  )  5.936
assign  ( resid  85    and name N  ) ( resid  85    and name CA  )
          ( resid  85    and name CB ) ( resid  85    and name HB*  )  6.4185
assign  ( resid  86    and name N  ) ( resid  86    and name CA  )
          ( resid  86    and name CB ) ( resid  86    and name HB*  )  3.615
assign  ( resid  86    and name N  ) ( resid  86    and name CA  )
          ( resid  86    and name CB ) ( resid  86    and name HB*  )  6.291
assign  ( resid  96    and name N  ) ( resid  96    and name CA  )
          ( resid  96    and name CB ) ( resid  96    and name HB*  )  3.938
 assign  ( resid  96    and name N  ) ( resid  96    and name CA  )
          ( resid  96    and name CB ) ( resid  96    and name HB*  )  6.3185
assign  ( resid  99    and name N  ) ( resid  99    and name CA  )
          ( resid  99    and name CB ) ( resid  99    and name HB*  )  4.8052
assign  ( resid  100    and name N  ) ( resid  100    and name CA  )
          ( resid  100    and name CB ) ( resid  100    and name HB*  )  7.0714
assign  ( resid  100    and name N  ) ( resid  100    and name CA  )
          ( resid  100    and name CB ) ( resid  100    and name HB*  )  4.345
assign  ( resid  101    and name N  ) ( resid  101    and name CA  )
          ( resid  101    and name CB ) ( resid  101    and name HB*  )  4.397
assign  ( resid  101    and name N  ) ( resid  101    and name CA  )
          ( resid  101    and name CB ) ( resid  101    and name HB*  )  1.974
assign  ( resid  102    and name N  ) ( resid  102    and name CA  )
          ( resid  102    and name CB ) ( resid  102    and name HB*  )  6.993
assign  ( resid  105    and name N  ) ( resid  105    and name CA  )
          ( resid  105    and name CB ) ( resid  105    and name HB*  )  3.325
assign  ( resid  107    and name N  ) ( resid  107    and name CA  )
          ( resid  107    and name CB ) ( resid  107    and name HB*  )  5.061
```

```
assign  ( resid  108    and name N  ) ( resid  108    and name CA  )
          ( resid  108    and name CB ) ( resid  108    and name HB*  )  4.26
assign  ( resid  110    and name N  ) ( resid  110    and name CA  )
          ( resid  110    and name CB ) ( resid  110    and name HB*  )  5.106
assign  ( resid  111    and name N  ) ( resid  111    and name CA  )
          ( resid  111    and name CB ) ( resid  111    and name HB*  )  4.566
assign  ( resid  120    and name N  ) ( resid  120    and name CA  )
          ( resid  120    and name CB ) ( resid  120    and name HB*  )  2.140
assign  ( resid  127    and name N  ) ( resid  127    and name CA  )
          ( resid  127    and name CB ) ( resid  127    and name HB*  )  5.782
assign  ( resid  130    and name N  ) ( resid  130    and name CA  )
          ( resid  130    and name CB ) ( resid  130    and name HB*  )  5.469
assign  ( resid  132    and name N  ) ( resid  132    and name CA  )
          ( resid  132    and name CB ) ( resid  132    and name HB*  )  4.400
assign  ( resid  133    and name N  ) ( resid  133    and name CA  )
          ( resid  133    and name CB ) ( resid  133    and name HB*  )  6.696
assign  ( resid  136    and name N  ) ( resid  136    and name CA  )
          ( resid  136    and name CB ) ( resid  136    and name HB*  )  2.776
assign  ( resid  142    and name N  ) ( resid  142    and name CA  )
          ( resid  142    and name CB ) ( resid  142    and name HB*  )  5.053
assign  ( resid  147    and name N  ) ( resid  147    and name CA  )
          ( resid  147    and name CB ) ( resid  147    and name HB*  )  5.102
assign  ( resid  148    and name N  ) ( resid  148    and name CA  )
          ( resid  148    and name CB ) ( resid  148    and name HB*  )  4.643
assign  ( resid  158    and name N  ) ( resid  158    and name CA  )
          ( resid  158    and name CB ) ( resid  158    and name HB*  )  6.981
assign  ( resid  158    and name N  ) ( resid  158    and name CA  )
          ( resid  158    and name CB ) ( resid  158    and name HB*  )  4.748
assign  ( resid  186    and name N  ) ( resid  186    and name CA  )
          ( resid  186    and name CB ) ( resid  186    and name HB*  )  8.022
assign  ( resid  190    and name N  ) ( resid  190    and name CA  )
          ( resid  190    and name CB ) ( resid  190    and name HB*  )  8.906
assign  ( resid  197    and name N  ) ( resid  197    and name CA  )
          ( resid  197    and name CB ) ( resid  197    and name HB*  )  6.916
assign  ( resid  197    and name N  ) ( resid  197    and name CA  )
          ( resid  197    and name CB ) ( resid  197    and name HB*  )  3.983
```

APPENDIX C. MODEL-FREE ANALYSIS DATA

C.1. W_1 AT 16.4 T FOR 153 RESIDUES

Table 19 | W_1 at 16.4 T results for model 1 with an average S^2 of 0.84

Residue	S^2	Error
44	0.922	0.003
46	0.925	0.010
53	0.867	0.009
54	0.929	0.005
88	0.893	0.013
95	0.933	0.015
96	0.870	0.007
100	0.850	0.012
117	0.991	0.020

Table 20 | W_1 at 16.4 T results for model 2 with averages of 0.79 for S^2 and 45 s for τ_e

Residue	S^2	Error	τ_e (ps)	Error
2	0.000	0.000	80.64	2.72
15	0.889	0.016	586.86	71.34
17	0.874	0.008	32.97	10.82
18	0.867	0.007	23.57	10.75
24	0.860	0.006	21.30	8.93
27	0.798	0.008	23.67	7.12
36	0.818	0.007	65.94	7.17
37	0.822	0.009	60.88	8.21
47	0.911	0.008	31.73	17.64
52	0.909	0.008	31.68	14.14
55	0.934	0.008	40.10	20.27
57	0.874	0.008	28.73	10.70
58	0.890	0.008	42.52	13.13
59	0.885	0.009	26.13	12.71
60	0.828	0.010	24.72	7.86
61	0.771	0.006	43.82	5.10
62	0.762	0.008	46.71	5.52
68	0.900	0.009	35.77	16.45
69	0.891	0.008	20.86	12.39
71	0.902	0.012	27.53	13.75
75	0.924	0.007	56.02	19.35
76	0.872	0.010	51.86	11.19
91	0.831	0.013	15.79	7.08
101	0.859	0.008	28.66	10.61
102	0.880	0.009	33.80	12.05
104	0.878	0.008	28.97	11.64
106	0.855	0.007	35.10	9.12
122	0.931	0.010	111.93	53.76

128	0.860	0.007	32.52	10.67
132	0.822	0.007	111.14	8.31
135	0.844	0.009	65.64	10.23
136	0.844	0.011	35.26	9.36
147	0.910	0.011	141.30	37.02
199	0.055	0.003	91.72	0.60

Table 21| W$_1$ at 16.4 T results for model 3 with an average of 0.83 for S^2.

Residue	S^2	Error	R$_{ex}$ (s^{-1})	Error
23	0.834	0.005	1.789	0.388
29	0.928	0.006	1.094	0.345
30	0.951	0.021	2.558	0.756
32	0.859	0.007	1.500	0.404
33	0.874	0.008	1.163	0.363
41	0.958	0.008	0.719	0.328
83	0.869	0.005	0.504	0.363
85	0.919	0.004	1.867	0.449
86	0.878	0.014	0.826	0.374
92	0.877	0.007	0.573	0.345
98	0.894	0.005	1.366	0.306
110	0.859	0.002	1.819	0.357
115	0.860	0.004	1.313	0.283
116	0.883	0.008	1.229	0.408
118	0.856	0.003	0.918	0.419
120	0.879	0.004	2.176	0.499

Table 22 | W$_1$ at 16.4 T results for model 4 with averages of 0.80 for S^2 and 42 s for τ_e

Residue	S^2	Error	τ_e (ps)	Error	R$_{ex}$ (s^{-1})	Error
16	0.895	0.027	742.85	217.75	1.887	0.492
19	0.881	0.009	41.52	10.99	0.852	0.384
20	0.877	0.009	17.78	9.91	1.354	0.418
21	0.852	0.014	14.34	9.51	1.550	0.466
25	0.849	0.115	17.39	67.64	1.515	1.604
26	0.865	0.008	15.83	8.57	1.041	0.362
38	0.890	0.221	87.02	763.57	2.413	3.192
39	0.848	0.202	41.46	117.40	1.374	3.575
40	0.840	0.139	39.22	77.44	1.159	1.641
79	0.873	0.008	42.46	10.19	0.843	0.403
81	0.796	0.008	28.86	7.01	0.769	0.286
109	0.861	0.176	15.42	314.48	1.907	5.771
111	0.828	0.215	23.38	116.14	2.941	7.633
112	0.864	0.018	15.69	10.55	1.290	0.547
113	0.887	0.120	19.54	70.67	1.297	3.822
119	0.833	0.008	24.48	9.17	1.275	0.693
121	0.822	0.010	52.28	8.74	1.029	0.469
125	0.824	0.009	14.33	7.51	1.605	0.355
126	0.847	0.085	17.00	61.40	2.917	1.005

127	0.782	0.007	13.18	5.38	0.924	0.402
133	0.851	0.181	63.08	117.53	0.933	2.935
138	0.878	0.010	54.46	13.96	0.713	0.413
139	0.865	0.011	39.42	10.38	1.892	0.335
140	0.876	0.150	43.90	99.36	0.713	1.262
142	0.876	0.302	68.16	169.63	0.933	3.666
180	0.845	0.167	60.09	118.36	2.077	2.173

Table 23 | W$_1$ at 16.4 T results for model 5 with averages of 0.74 for S_f^2, 0.51 for S_s^2, and τ_s of 2.5 ns

Residue	S_f^2	Error	S_s^2	Error	τ_s (ns)	Error
7	0.783	0.040	0.204	0.048	0.648	0.051
8	0.845	0.042	0.192	0.046	0.641	0.032
10	0.836	0.020	0.268	0.030	0.738	0.023
13	0.951	0.016	0.530	0.018	0.741	0.034
14	0.876	0.020	0.532	0.021	0.965	0.054
28	0.766	0.027	0.851	0.025	0.508	0.123
42	0.954	0.017	0.932	0.018	1.112	1.564
43	0.862	0.019	0.942	0.028	1.676	2.812
48	0.522	0.013	0.284	0.033	1.243	0.053
49	0.661	0.017	0.304	0.030	1.269	0.068
50	0.907	0.027	0.948	0.022	1.233	2.473
51	0.899	0.018	0.913	0.026	2.212	2.011
56	0.871	0.021	0.918	0.023	0.916	0.363
63	0.908	0.019	0.853	0.019	0.482	0.119
64	0.870	0.018	0.883	0.020	1.040	0.306
65	0.891	0.018	0.904	0.020	0.945	0.270
66	0.856	0.019	0.871	0.022	1.128	0.293
67	0.906	0.021	0.949	0.019	1.216	2.153
70	0.918	0.020	0.952	0.019	1.147	2.315
72	0.880	0.016	0.889	0.021	0.850	0.187
73	0.890	0.020	0.919	0.027	1.474	1.915
77	0.847	0.018	0.911	0.018	1.087	1.026
80	0.853	0.015	0.808	0.017	0.857	0.119
87	0.867	0.018	0.921	0.030	7.923	2.993
90	0.893	0.022	0.946	0.017	1.283	1.933
93	0.847	0.020	0.778	0.026	0.893	0.116
99	0.905	0.026	0.958	0.020	1.287	2.682
103	0.783	0.022	0.890	0.026	1.295	0.563
105	0.855	0.026	0.921	0.025	0.987	0.887
107	0.791	0.015	0.899	0.020	0.996	0.239
108	0.847	0.023	0.917	0.026	1.025	0.914
129	0.840	0.029	0.876	0.037	1.416	0.790
130	0.854	0.046	0.893	0.030	0.446	0.193
134	0.921	0.019	0.852	0.021	0.814	0.145
144	0.859	0.055	0.165	0.052	0.575	0.104
146	0.933	0.021	0.920	0.016	0.544	0.202
149	0.901	0.018	0.643	0.023	0.863	0.060
153	0.858	0.017	0.354	0.023	0.840	0.034

155	0.850	0.014	0.251	0.016	0.754	0.022
156	0.824	0.016	0.186	0.031	0.863	0.029
158	0.871	0.020	0.189	0.028	0.816	0.023
160	0.770	0.010	0.200	0.016	0.756	0.018
162	0.753	0.031	0.007	0.064	0.796	0.039
164	0.884	0.064	0.214	0.092	0.627	0.060
165	0.855	0.027	0.206	0.033	0.720	0.024
169	0.784	0.008	0.109	0.013	0.720	0.015
171	0.771	0.027	0.095	0.026	0.680	0.048
173	0.795	0.043	0.025	0.058	0.685	0.053
174	0.593	0.046	0.004	0.059	0.737	0.035
176	0.797	0.015	0.187	0.046	0.666	0.023
178	0.709	0.033	0.075	0.030	0.718	0.042
181	0.837	0.031	0.161	0.041	0.697	0.034
182	0.759	0.022	0.101	0.022	0.824	0.037
184	0.745	0.011	0.090	0.025	0.682	0.017
186	0.812	0.042	0.098	0.021	0.592	0.061
187	0.793	0.023	0.094	0.046	0.617	0.032
188	0.843	0.021	0.247	0.042	0.867	0.037
192	0.798	0.044	0.085	0.031	0.554	0.040
193	0.784	0.041	0.093	0.036	0.704	0.059
195	0.856	0.054	0.154	0.053	0.446	0.058
196	0.821	0.058	0.159	0.057	0.442	0.061
197	0.756	0.067	0.000	0.040	0.504	0.062
198	0.777	0.082	0.113	0.053	0.278	0.078

C.2. W_{2-1} AT 16.4 T FOR 154 RESIDUES

Here are results for W_{2-1}:

Table 24 | W_{2-1} at 16.4 T results for model 1 with an average S^2 of 0.94.

Residue	S^2	Error
20	0.972	0.008
21	0.963	0.014
24	0.953	0.011
29	0.967	0.017
32	0.977	0.014
33	0.950	0.006
44	0.937	0.009
46	0.944	0.009
52	0.953	0.007
54	0.987	0.008
59	0.940	0.006
68	0.911	0.011
69	0.954	0.007
71	0.936	0.008
88	0.982	0.011
92	0.938	0.007
95	0.913	0.014
96	0.894	0.008
99	0.898	0.006
100	0.904	0.010
109	1.000	0.003
119	0.900	0.011
143	0.937	0.015

Table 25 | W_{2-1} at 16.4 T results for model 2 with averages of 0.80 for S^2 and 320 s for τ_e

Residue	S^2	Error	τ_e (ps)	Error
13	0.493	0.010	770.02	25.26
27	0.834	0.010	25.76	10.74
36	0.840	0.014	67.72	30.86
37	0.893	0.014	82.28	68.71
40	0.897	0.018	50.80	26.26
42	0.914	0.013	689.41	283.12
61	0.837	0.005	43.16	11.32
70	0.955	0.016	1145.71	3336.97
90	0.939	0.015	342.66	162.25
101	0.894	0.009	32.76	15.31
104	0.915	0.011	56.17	22.14
122	0.934	0.013	736.90	178.00
124	0.949	0.013	73.29	65.40
128	0.855	0.010	39.16	13.45
132	0.848	0.011	324.90	53.37
135	0.930	0.012	159.93	64.33

Residue	S^2	Error	τ_e (ps)	Error
147	0.921	0.016	428.33	172.75
165	0.099	0.008	664.44	13.17
196	0.134	0.009	589.15	8.27

Table 26 | W_{2-1} at 16.4 T results for model 3 with an average of 0.92 for S^2

Residue	S^2	Error	R_{ex} (s^{-1})	Error
18	0.926	0.011	0.971	0.407
23	0.902	0.013	1.024	0.294
25	0.902	0.015	0.872	0.291
26	0.931	0.017	0.633	0.346
48	0.929	0.009	1.470	0.375
49	0.944	0.009	2.810	0.420
53	0.930	0.014	0.769	0.442
83	0.938	0.015	1.249	0.281
84	0.942	0.018	0.652	0.442
85	0.968	0.006	1.396	0.291
86	0.920	0.008	0.423	0.265
94	0.896	0.010	1.814	0.203
98	0.896	0.020	1.880	0.390
106	0.935	0.015	1.164	0.353
110	0.930	0.008	1.398	0.244
111	0.909	0.014	2.530	0.296
112	0.881	0.015	2.110	0.521
113	0.934	0.011	0.997	0.218
115	0.911	0.015	1.909	0.402
118	0.902	0.013	1.345	0.432
120	0.888	0.017	3.189	0.359
121	0.854	0.015	2.259	0.936
123	0.985	0.008	4.658	0.331
125	0.897	0.011	1.462	0.242
126	0.931	0.012	2.542	0.377
127	0.827	0.013	1.078	0.337

Table 27 | W_{2-1} at 16.4 T results for model 4 with averages of 0.81 for S^2 and 335 s for τ_e

Residue	S^2	Error	τ_e (ps)	Error	R_{ex} (s^{-1})	Error
15	0.811	0.118	1069.11	169.98	0.976	1.838
16	0.845	0.032	1011.05	375.99	2.671	0.558
19	0.871	0.150	41.62	1477.27	2.542	2.119
38	0.907	0.136	410.61	1724.30	3.106	1.423
39	0.915	0.032	93.72	1266.21	1.735	0.546
41	0.949	0.024	1046.05	2737.06	1.088	0.383
62	0.807	0.171	64.10	84.47	0.543	2.049
79	0.926	0.311	93.05	1048.61	0.892	5.448
81	0.864	0.218	54.52	112.60	0.524	3.088
116	0.877	0.026	36.32	707.70	2.259	0.507
138	0.926	0.221	57.89	1420.73	1.220	3.077

Residue	S^2	Error	τ_e (ps)	Error	R_{ex} (s^{-1})	Error
139	0.947	0.164	462.67	1732.08	0.838	2.336
142	0.927	0.278	202.72	3034.78	1.681	4.221
144	0.899	0.100	55.76	1244.43	1.780	1.235

Table 28 | W$_{2-1}$ at 16.4 T results for model 5 with averages of 0.86 for S_f^2, 0.51 for S_s^2, and τ_s of 1.3 ns

Residue	S_f^2	Error	S_s^2	Error	τ_S (ns)	Error
5	0.838	0.053	0.110	0.032	0.583	0.072
8	0.843	0.031	0.124	0.021	0.749	0.039
9	0.863	0.016	0.310	0.009	0.698	0.032
11	0.914	0.059	0.080	0.032	0.689	0.093
14	0.892	0.022	0.493	0.016	1.000	0.076
17	0.955	0.016	0.927	0.014	0.713	0.281
28	0.818	0.009	0.857	0.014	0.515	0.099
43	0.930	0.019	0.920	0.020	1.176	2.273
47	0.972	0.011	0.942	0.017	2.113	3.671
50	0.878	0.015	0.900	0.027	9.025	1.360
51	0.917	0.012	0.882	0.025	9.025	3.132
55	0.974	0.011	0.956	0.013	1.366	3.490
56	0.927	0.011	0.961	0.011	1.609	3.712
57	0.907	0.015	0.960	0.014	1.257	3.478
58	0.904	0.022	0.961	0.018	1.244	3.722
60	0.896	0.021	0.920	0.019	0.787	0.459
63	0.915	0.015	0.864	0.015	0.570	0.124
64	0.903	0.013	0.932	0.014	1.313	2.534
65	0.902	0.015	0.915	0.016	1.219	2.114
66	0.906	0.014	0.933	0.014	0.799	1.236
67	0.940	0.015	0.937	0.017	1.246	3.162
72	0.919	0.014	0.912	0.020	9.025	3.514
73	0.893	0.013	0.943	0.021	2.078	3.663
74	0.836	0.043	0.122	0.028	0.738	0.048
75	0.978	0.014	0.921	0.014	0.480	0.321
76	0.956	0.015	0.954	0.015	0.575	1.398
77	0.878	0.014	0.926	0.021	1.411	3.025
80	0.850	0.015	0.809	0.016	0.843	0.143
93	0.824	0.020	0.694	0.018	1.339	0.299
102	0.898	0.013	0.941	0.018	1.409	3.200
105	0.875	0.023	0.947	0.021	0.815	2.422
107	0.856	0.026	0.854	0.033	1.509	1.920
108	0.901	0.026	0.925	0.025	0.903	1.852
129	0.869	0.030	0.924	0.037	1.679	3.444
130	0.901	0.023	0.908	0.018	0.483	0.180
134	0.945	0.018	0.829	0.016	0.901	0.185
136	0.946	0.017	0.928	0.016	0.708	0.510
146	0.951	0.015	0.902	0.013	0.884	0.371
148	0.929	0.015	0.852	0.013	0.722	0.167
149	0.902	0.024	0.555	0.020	1.070	0.092
153	0.867	0.023	0.274	0.011	0.946	0.056

Residue	S_f^2	Error	S_s^2	Error	τ_S (ns)	Error
155	0.850	0.024	0.219	0.015	0.813	0.051
156	0.836	0.018	0.144	0.004	0.900	0.045
158	0.819	0.020	0.144	0.018	0.892	0.044
160	0.765	0.019	0.133	0.021	0.837	0.036
161	0.883	0.022	0.287	0.034	0.694	0.031
162	0.786	0.027	0.131	0.038	0.792	0.039
163	0.888	0.020	0.221	0.013	0.783	0.030
164	0.849	0.030	0.155	0.042	0.762	0.068
166	0.864	0.070	0.098	0.041	0.654	0.081
167	0.905	0.018	0.348	0.011	0.725	0.035
169	0.757	0.060	0.069	0.035	0.762	0.120
171	0.714	0.011	0.105	0.005	0.730	0.024
173	0.809	0.039	0.116	0.036	0.696	0.079
174	0.737	0.026	0.080	0.015	0.721	0.041
176	0.811	0.030	0.142	0.028	0.739	0.058
178	0.728	0.023	0.078	0.019	0.810	0.042
180	0.799	0.022	0.111	0.010	0.765	0.050
181	0.684	0.029	0.149	0.036	0.770	0.048
182	0.762	0.020	0.099	0.009	1.013	0.049
184	0.703	0.029	0.000	0.046	0.877	0.047
186	0.813	0.010	0.101	0.031	0.727	0.017
188	0.848	0.022	0.202	0.016	0.969	0.053
190	0.828	0.013	0.318	0.015	0.639	0.023
192	0.790	0.017	0.160	0.005	0.853	0.035
193	0.806	0.018	0.134	0.005	0.913	0.043
195	0.869	0.018	0.147	0.006	0.773	0.033
197	0.815	0.016	0.132	0.005	0.845	0.032
198	0.836	0.030	0.102	0.018	0.758	0.056
199	0.735	0.019	0.103	0.025	0.845	0.036

C.3. $W_{2\text{-}2}$ AT 16.4 T FOR 133 RESIDUES

Table 29 | $W_{2\text{-}2}$ at 16.4 T results for model 1 with an average S^2 of 0.99.

Residue	S^2	Error
18	1.000	0.006
40	0.960	0.013
44	1.000	0.004
46	1.000	0.009
53	1.000	0.004
59	1.000	0.003
84	0.997	0.010
95	0.976	0.014
100	0.959	0.011
118	0.963	0.013
144	0.994	0.009

Table 30 | $W_{2\text{-}2}$ at 16.4 T results for model 2 with averages of 0.82 for S^2 and 1584 s for τ_e

Residue	S^2	Error	τ_e (ps)	Error
27	0.895	0.012	46.49	17.64
28	0.793	0.011	87.87	12.98
29	0.939	0.025	9746.84	3380.52
32	0.944	0.022	9746.84	4155.74
36	0.889	0.011	126.65	122.46
42	0.943	0.016	3210.05	2013.84
47	0.916	0.018	1898.40	1878.03
48	0.952	0.026	172.31	2563.86
49	0.975	0.017	814.59	2846.64
50	0.728	0.018	349.27	22.76
51	0.900	0.020	9746.84	3943.31
55	0.919	0.018	1360.25	3157.12
56	0.975	0.009	2936.69	2503.55
60	0.928	0.013	424.84	175.88
61	0.923	0.016	60.82	29.11
63	0.874	0.009	427.52	51.52
64	0.936	0.014	536.21	3399.19
65	0.922	0.014	614.01	147.68
67	0.961	0.014	2912.26	1865.90
73	0.966	0.012	508.45	2762.98
77	0.955	0.016	429.25	3199.20
81	0.925	0.015	104.98	32.68
88	0.950	0.017	5139.79	2700.14
102	0.953	0.010	449.81	330.96
105	0.923	0.010	68.95	109.23
108	0.953	0.015	390.11	361.54
109	0.921	0.029	1431.36	3196.88
127	0.876	0.011	23.03	14.53
132	0.859	0.015	633.73	60.76
134	0.835	0.007	873.23	89.15
135	0.901	0.015	678.51	113.65
136	0.926	0.013	798.49	1265.98
146	0.886	0.012	883.03	2675.75
148	0.847	0.011	632.83	913.61
178	0.058	0.003	477.31	5.52
186	0.057	0.006	493.98	4.84
195	0.056	0.003	425.73	6.13
197	0.028	0.004	546.48	10.88

Table 31 | $W_{2\text{-}2}$ at 16.4 T results for model 3 with an average of 0.98 for S^2

Residue	S^2	Error	R_{ex} (s^{-1})	Error
21	1.000	0.015	2.236	0.636
25	0.966	0.015	1.017	0.370
71	1.000	0.004	0.536	0.242
83	1.000	0.009	2.417	0.209
86	0.988	0.009	0.777	0.316
92	0.995	0.006	0.765	0.259
98	0.957	0.018	1.386	0.424
99	0.999	0.009	0.547	0.299

Residue	S^2	Error	R_{ex} (s^{-1})	Error
101	0.974	0.009	0.593	0.252
110	0.999	0.005	1.546	0.230
111	0.976	0.018	1.943	0.579
112	0.947	0.017	2.680	0.584
113	1.000	0.008	0.905	0.597
115	0.976	0.017	1.864	0.541
120	0.955	0.018	2.410	0.492
125	0.963	0.012	1.053	0.278

Table 32 | $W_{2\text{-}2}$ at 16.4 T results for model 4 with averages of 0.88 for S^2 and 2305 s for τ_e

Residue	S^2	Error	τ_e (ps)	Error	R_{ex} (s^{-1})	Error
15	0.780	0.208	1268.94	515.27	3.311	2.799
16	0.817	0.083	1294.69	915.87	3.113	1.184
17	0.901	0.242	755.85	913.75	1.144	3.796
19	0.929	0.329	101.05	123.07	1.278	8.071
20	0.964	0.244	3739.52	4116.48	1.814	4.216
24	0.970	0.132	1337.56	3884.44	0.936	1.731
33	0.942	0.163	434.87	1712.49	0.830	2.574
38	0.838	0.183	729.72	75.41	4.697	2.131
39	0.942	0.097	649.69	1165.27	1.352	1.455
41	0.910	0.178	1637.25	2803.04	2.688	2.505
52	0.959	0.332	9746.84	4209.42	0.506	4.656
54	0.911	0.244	914.38	753.44	3.506	7.541
69	0.962	0.249	9746.84	2508.08	2.272	3.943
70	0.860	0.094	9746.84	3648.07	0.932	1.158
75	0.887	0.244	741.23	426.39	2.585	5.795
76	0.927	0.096	608.53	951.24	1.153	1.310
79	0.926	0.382	642.16	1642.81	0.710	8.451
85	0.944	0.185	9746.84	3971.15	3.353	5.145
90	0.899	0.093	944.91	1793.84	0.905	1.211
122	0.906	0.195	1516.91	2248.25	1.154	3.011
123	0.953	0.280	3825.62	2754.87	5.386	8.097
126	0.965	0.022	299.31	2966.39	2.711	1.150
138	0.927	0.270	483.41	2341.32	1.610	3.854
139	0.891	0.024	943.58	1048.99	2.070	0.480
140	0.954	0.211	638.00	2826.68	1.359	3.666
142	0.878	0.190	660.07	397.80	2.635	2.560
147	0.878	0.207	814.88	190.03	0.974	3.208
196	0.103	0.051	583.12	104.32	1.439	1.085

Table 33 | $W_{2\text{-}2}$ at 16.4 T results for model 5 with averages of 0.74 for S_f^2, 0.51 for S_s^2, and τ_s of 2.5 ns

Residue	S_f^2	Error	S_s^2	Error	τ_S (ns)	Error
6	0.773	0.036	0.127	0.039	0.811	0.062
8	0.800	0.018	0.171	0.006	0.877	0.041
9	0.871	0.020	0.348	0.011	0.834	0.047
10	0.825	0.018	0.199	0.006	0.855	0.037
37	0.930	0.027	0.903	0.022	1.158	1.939
43	0.958	0.018	0.906	0.018	1.508	2.353
57	0.985	0.007	0.976	0.007	1.104	3.563

Residue	S_f^2	Error	S_s^2	Error	τ_S (ns)	Error
58	0.967	0.019	0.939	0.018	0.861	1.676
66	0.951	0.016	0.917	0.017	0.929	1.423
72	0.782	0.022	0.730	0.088	3.934	3.328
80	0.882	0.016	0.784	0.014	0.843	0.115
104	0.979	0.016	0.940	0.014	0.756	1.026
107	0.894	0.022	0.859	0.049	1.774	2.665
117	0.858	0.038	0.849	0.040	1.840	1.932
128	0.939	0.022	0.966	0.019	0.612	2.541
129	0.877	0.017	0.885	0.034	2.195	3.021
130	0.926	0.017	0.877	0.015	0.654	0.268
149	0.812	0.010	0.259	0.036	9.747	0.660
156	0.820	0.024	0.135	0.009	0.935	0.062
158	0.803	0.018	0.146	0.012	0.939	0.044
159	0.844	0.019	0.192	0.013	0.900	0.036
160	0.780	0.017	0.150	0.007	0.821	0.034
162	0.768	0.057	0.107	0.018	0.820	0.123
163	0.887	0.021	0.203	0.009	0.787	0.031
164	0.856	0.029	0.143	0.033	0.750	0.044
165	0.839	0.030	0.104	0.020	0.808	0.068
166	0.873	0.096	0.069	0.028	0.633	0.119
167	0.839	0.016	0.216	0.007	0.716	0.026
168	0.763	0.051	0.129	0.042	0.649	0.077
169	0.747	0.032	0.027	0.031	0.814	0.051
171	0.717	0.057	0.096	0.041	0.723	0.084
173	0.818	0.069	0.085	0.039	0.672	0.076
174	0.727	0.061	0.048	0.023	0.727	0.133
176	0.807	0.037	0.111	0.023	0.736	0.068
184	0.676	0.035	0.104	0.031	0.978	0.067
188	0.836	0.024	0.219	0.011	1.035	0.063
189	0.719	0.043	0.061	0.021	0.757	0.094
192	0.707	0.080	0.022	0.014	0.895	0.132
193	0.860	0.057	0.049	0.013	0.751	0.079
194	0.904	0.069	0.010	0.008	0.708	0.116

APPENDIX D. RESIDUAL DIPOLAR COUPLINGS

D.1. INTRODUCTION

Protein structure determination by NMR spectroscopy utilises many different experimental restraints. NOEs, J-couplings, dihedral angles, and the radius of gyration have been discussed in chapter 3. Residual dipolar couplings (RDCs) are another type of experimental restraint that can be used to determine protein structures. RDCs are unique because they provide relative orientation between nuclei irrespective of distance separation (Chen and Tjandra, 2012).

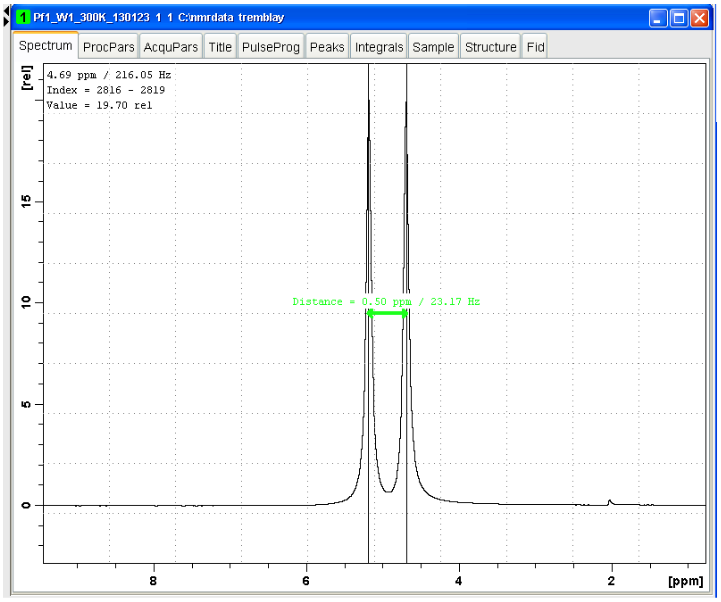
Molecular alignment may be quantified by the alignment tensor. RDCs arise when proteins in solution are partly oriented relative to the magnetic field in an anisotropic sample, hence the term *residual* dipolar coupling. In isotropic solution, the dipolar couplings average to zero over all molecules. Under anisotropic conditions, the dipolar coupling term (D.1) doesn't average to 0.

A dipolar coupling occurs when the magnetic field generated by one nuclear dipole affects the magnetic field at another nucleus (Rule and Hitchens 2006). The magnitude of the dipolar coupling depends on the strength of the magnetic field generated by one spin and the size of the magnetic dipole (μ_0) of the recipient spin. In isotropic solution, the ^{1}H-^{15}N coupling constant for an amide bond is ~92 Hz. In samples with complete ordering, the dipolar coupling is large, and greatly complicates the acquisition and analysis of NMR spectra through extensive broadening. When the alignment is kept sufficiently weak, the NMR spectra retain the simplicity normally observed in regular isotropic solution, while allowing quantitative measurement of a wide variety of residual dipolar couplings (RDCs), even in macromolecules (Zweckstetter and Bax 2002). The dipolar coupling term is described by equation D.1:

$$D_{ij} = -\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_i\gamma_j\hbar}{2\pi^2 r_{ij}^3}\left\langle \tfrac{1}{2}\left(3\cos^2\theta - 1\right)\right\rangle \tag{D.1}$$

where D_{ij} is the dipolar coupling in Hz, μ_o is the magnetic moment, γ_i and γ_j are the gyromagnetic ratios of nuclei *i* and *j*, respectively, $\hbar$ is the reduced Planck constant, r_{ij} is the distance between the two nuclei, and θ is the angle of the inter-spin vector relative to the external magnetic field.

The quadrupolar splitting of the HDO signal observed in a 1D ^{2}H-NMR spectrum may be used to determine the degree of sample orientation. When oriented, the deuterium peak splits with a *J*-coupling value. Ideally, the quadrupolar coupling should not to be too small (<10 Hz), as this indicates poor level of orientation, and it should not be too large (> 30-40 Hz), or else the peaks broaden(Hansen et al. 1998) (Figure 70).



```
4.69 ppm / 216.05 Hz
Index = 2816 - 2819
Value = 19.70 rel
```

```
Distance = 0.50 ppm / 23.17 Hz
```

Figure 70 | A spectrum of a sample deuterium splitting as a reflection of the quality of the splitting obtained with the protein. That sample was acquired at 300 MHz at the NMR-3 facility with 15 mg/mL of phage at ~pH 6.5. The distance in Hz between the deuterium peaks will reflect the expected amount of splitting for the protein of interest.

D.1.1. NMR Experiments Used

Observing the HDO splitting at the NMR-3. This protocol is for the Bruker 7.04 T instrument at the NMR-3 facility. The sample is locked and shim as would any other sample but *not* tuned (no wobb). In a 1D zg template, load the pulse program zg_2h or 1d_2h. On the console, turn the lock and the sweep off. In the command bar, set the locnuc to off (this is a parameter in the pulse program). Then, the lock power has to be set to -60. The magnet is now ready to acquire the quadrupolar HDO peak. Continue with the typical commands rga and zg.

I also wrote this protocol in the black manual book that is on the operating desk for the 300 MHz magnet.

IPAP. The in-phase anti-phase (IPAP) (Nolis and Parella 2007) NMR experiment is a common experiment for collecting RDCs. The data is collected in an interleaved fashion based on a semi-selective τ-π-τ pulse and later split inside Topspin with the "split ipap" command. The resulting two spectra will each have one of the two doublet components from the coupled peak (the α or β) , making each spectrum much less overlapped than if both components were acquired simultaneously. The RDCs are obtained as the difference in splitting observed in the IPAP experiments under anisotropic (J ± RDC) and isotropic (J only) conditions.

IDIS-RDC-TROSY. RDC measurement using isotopically discriminated (IDIS) NMR experiments (Bermel et al. 2009) allow the measurement of RDCs on proteins or protein complexes that are differentially labeled; one portion is uniformly ^{15}N-enriched, while the other portion is uniformly ^{13}C/^{15}N-enriched. The IDIS portion of the pulse sequence (Golovanov et al. 2007) record the spectra based on the presence or absence of ^{15}N - ^{13}C' coupling while selecting for the α or β peaks in an interleaved manner. The approach ensures that RDCs measured for both protein components have exactly the same alignment tensor, which makes measurements much more accurate (Bermel et al. 2009).

Table 34 | RDC experiments for measurements performed.

Experiments	Bruker pulse sequence	Relaxation delay	# of scans	Increments	Spectral width (ppm)	Center position (ppm)	^{1}H frequency	Facility
				F1|F2	F1|F2	F1|F2		
W_1								
1D	zg2h	2	64/128	-	-	-	300	NMR-3
IPAP HSQC	Varian	1.5?	?	4096|256	11.6|23	4.7|115.5	750	Oxford
W_2								
IDIS-IPAP HSQC	hsqcf3gpidphwg	1.5	64	2048|192	15|24	4.7|115.25	700	McMaster
^{13}C IPAP HSQC	hsqcetgpiajcsp	1.5	164	4096|256	14|85	4.7|40.0	600	PEI (NRC)

D.2. METHODS: RDC MEDIA PROTOCOLS

D.2.1. Overview Of Types Of Media For RDC Sample Preparation

Many types of alignment media exist and can be used to align proteins (Prestegard et al. 2004; Higman et al. 2011). *Lipid bicelles* were the first alignment medium to be used (Bax and Tjandra 1997). They form liquid crystals in solution and, depending on the lipid composition, can be used across most pHs (Losonczi and Prestegard 1998;

Cavagnero et al. 1999). *Filamentous phage Pf1* is a 7,349-nucleotide DNA-phage surrounded by a net negatively charged protein coat at a 1:1 nucleotide:protein ratio (Hansen et al. 2000) that spontaneously aligns in a magnetic field (Hansen et al. 1998). Stretched *polyacrylamide/polyelectrolyte gels* (Meier et al. 2002; Luy et al. 2004; Gebel and Shortle 2007) provide anisotropy due to the pores in the gel. *Collagen gels* also provide a gelatinous porous matrix in which to measure RDCs (Ma et al. 2008). *Liquid crystalline phases* spontaneously form in solution and can be made of (but not restricted to) ethylene glycol/glucopone and n-alkyl alcohol mixtures (Rückert and Otting 2000) or PEG/alcohol mixture (Barrientos et al. 2000). They can be used over a wide variety of temperature and pH values. *Purple membranes* are highly negatively charged bacterial membranes of *Halobacterium salinarum* containing bacteriorhodopsin as the sole protein and have a crystalline arrangement in a magnetic field (Sass et al. 1999; Koenig et al. 1999). *DNA nanotubes* and crystalline phase G-tetrad DNA (Lorieau et al. 2008) have recently been reported as a medium for the measurement of membrane bound proteins. The combinations of the filamentous phage Pf1 and stressed polyacrylamide gels (Trempe et al. 2002; Ruan and Tolman 2005; Park et al. 2009) and purple membrane and polyacrylamide gels (Sass et al. 2000) have also been reported to be useful for measuring RDCs of membrane proteins.

In this next section I'll introduce the different methods used to measure RDCs available to me at the time. The optimal conditions would produce a splitting of 15-25 Hz (Barrientos et al. 2000).

D.2.2. NMR Experiments

W_1 experiments. ^{15}N uniformly labeled W_1 in 50 mM potassium phosphate buffer, pH 7.5, 300mM NaCl was mixed with Pf1 at 15 and 18 mg/mL. The RDC data for W_1 was acquired at the University of Oxford, UK, by by Jan Rainey with Dr. Christina Redfield on April 22 2013 (Spectra title in analysis:Oxford_1/0_ipap1/0_pf1). IPAP HSQCs were acquired for W_1 with an without Pf1 using the standard HSQC parameters presented in the main text chapters (3/4/5) with exception of increments; this experiment requires double the increments since the data is interleaved and split. This data set was acquired at 17.6T on a homebuilt instrument and processed using NMRPipe (Delaglio et al. 1995).

W_2 experiments. W_2 was segmentally labeled, with one W being only ^{15}N isotopically labeled and the other $^{13}C/^{15}N$, through the use of split inteins (see chapter 4). IPAP versions of the IDIS-HSQCs were acquired, acquiring both labeled portions of W_2 interleaved in one IDIS experiment, but the α and β IPAP peaks as completely separate experiments. Jan Rainey acquired the data at McMaster University on their Bruker 16.4 T magnet. Here, I'm assuming that the concentration of W_2 was ~0.1 mM and that the concentration of phage was ~15-20 mg/mL.

^{13}C RDC were acquired on the same sample on a Bruker 14.0 T magnet using the Bruker pulse program hsqcetgpiajcsp (or CLIP/CLAP HSQC) (Enthart et al. 2008). The parameters were as follows: spectral width 85ppm, center position 40ppm, relaxation delay 1.5s.

D.2.3. Phage

The first medium attempted was long filamentous phage (Pf1), recommended to us by Dr. Gary Shaw. Pf1 phage strain LP11-92 is isolated from wild type *Pseudomonas aeruginosa* and propagated in the phage free strain LA23-99 (Hansen et al. 1998; Hansen et al. 2000). Obtaining RDCs with phage is rather straightforward as long as the pH is held above 6 (Figure 71). Phage particles have a tendency to aggregate below pH6. Also, if the protein interacts with phage, than NaCl can be added to balance out the charges. A good description of the use of Pf1-phage for RDCs is located in this reference (Zweckstetter and Bax 2001).

The protocol and phage were both obtained from Dr. Gary Shaw at Western University. Pf1 phage was cultured, purified, and produced by Bruce Stewart. Dr. Lingling Xu then dialyzed the phage into the appropriate NMR buffer (10%D2O, 90%H2O, 1mM NaN3, 20mM AcOH, pH6.5-7.5) without the protein to make a general stock of phage. The phage can then be mixed with the NMR sample. To obtain a quadrupolar splitting of 20 Hz (Figure 70), the phage concentration needs to be ~15mg/ml. The HDO quadrupolar splitting linearly increases with phage concentration above 20mg/mL (Zweckstetter and Bax 2001). Since Pf1-phage is only stable between the pH 6-8, the NMR sample has to be analyzed at pH 6.5, which was the closest usable pH to the W_1 structure pH of 5 (Chapter 3). About 300 mM of NaCl must be added to the sample to neutralize the highly negative charge from Pf1 or else W_1 interacts with Pf1 and unfolds.

Mixing the Pf1-phage with the protein sample causes bubble formation. I found that these bubbles don't really affect the resulting NMR spectra. The bubbles also seem to disappear during the acquisition.

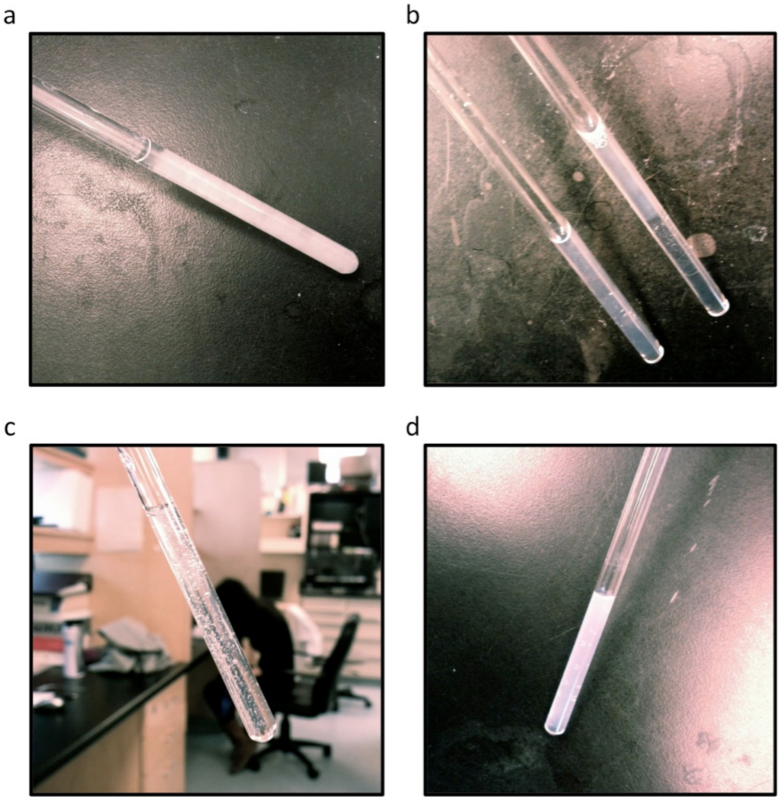


Figure 71 | Experimenting with different RDC media. A) First and only attempt at a collagen gel. The sample was at 500MHz at 37˚C overnight but only aggregates formed. B) PEG/hexanol failed liquid crystal mixtures. C) Phage preparation at 15 mg/mL when the sample is first placed in the tube and D) after an overnight in the magnet, the bubbles disappear and the phage aggregates when the pH is not right.

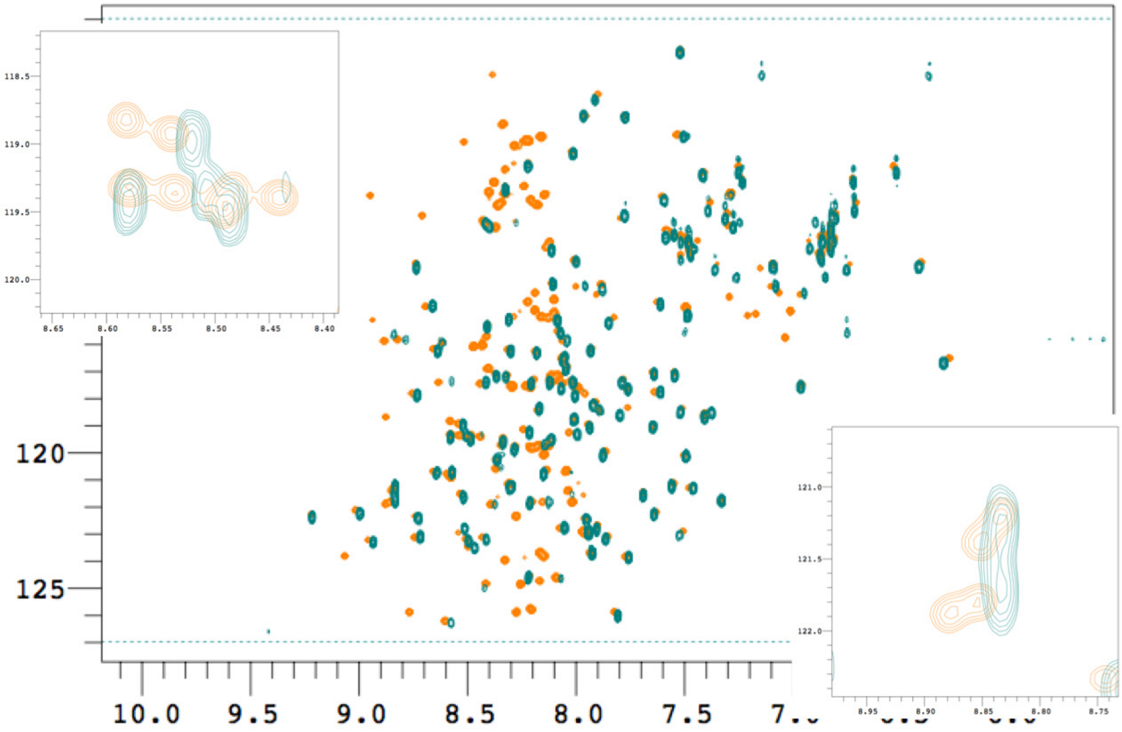


Figure 72 | W_1 spectra acquired at pH 5 (orange) and at pH 7.5 (green). Two insets are included for close-up of the chemical shift changes. The C-terminal tail, being flexible, is highly exchangeable with at pH 7.5 therefore many of the peak do not appear.

D.2.4. Collagen Gels

The original protocol is located in Ma *et al* (Ma et al. 2008). Below is a description of my protocol for making a collagen gel (without W_1) in an NMR tube. Making a collagen gel seems to be a fine equilibrium between making fibres and a gel matrix. A very slow increase in temperature seems to be the driving force for making a gel or else fibres form instead.

A collagen solution was made by concentrating the initially diluted 3 mg/mL of PureCol type 1 bovine collagen (Advanced BioMatrix, 100 mL bottle, 4yrs old, stored at pH ~3) using a spin filter (Millipore, 15 mL tube) to ~25mg/mL, a concentration estimated by the volume decrease. The concentrated collagen has a consistency thicker than corn syrup. Keeping the collagen at 4˚C, the collagen gel was mixed with potassium phosphate buffer (80 mM potassium phosphate buffer solution, 90%H_2O / 10% D_2O, pH 7.3, filtered with a 0.2 μm syringe filter to remove impurities), without removing the acetic acid, to a concentration of 3 mg/mL and 600μL volume. The pH was adjusted to

7.3. The quadrupolar splitting expected for the collagen gel increases with increasing amounts of collagen.

The NMR tube was kept on ice, until placed in the NMR on the 500 MHz using the 5mm BBO Smartprobe at the NMR-3 facility. The collagen was let at 4˚C in the magnet for 3h. Then, the temperature was ramped from 4-22˚C (room temp) at 1˚C/10min and from 22-37 1˚C/15min. I modified the protocol for ramping the temperature during the night and chose to ramp the temperature from 4˚C to 37˚C at 1˚C/15min, which is 495 min or 8.5 hours on the magnet. At the time I tried this experiment, the ramping wasn't calibrated properly within Topspin and therefore the temperature ramped up too fast and the collagen aggregated into fibres (Figure 71). Since the experiment hasn't been repeated since, I'm including the fix, which may or may not have been implemented since the winter 2013.

Robin from Bruker suggested:

> In this case, the AU program multi_zgvt will do what you want (and it will acquire a spectrum at each temperature, so you can track the formation of the . It's pretty simple-minded. Start by creating a vt list by typing edlist vt and going to File / New; make a list of each temperature you want to pause at (in K). Save this. (Or, just save vt_277_310, attached, into <topspin home>/exp/stan/nmr/lists/vt and use vt_277_310 as the VT list name.) Now, work in a new proton (or other experiment) dataset and type multi_zgvt. Type in the name of your vt list, check that the highest temperature, lowest temperature, and total number of experiments are correct, then enter the equilibration time at each temperature (perhaps 870 s, because it will take some time to change the temperature by 1 C). The AU program will then calculate a total time for the experiment, ask you to OK it, and then set the first temperature, equilibrate at it, run the proton experiment, go into the next experiment and repeat. The program uses IEXPNO to go to or create the subsequent experiments, so if those experiments don't exist, the initial proton (or other experiment) will just be copied and run for each subsequent temperature.
>
> I've attached a slightly nicer version of this AU program, au_multivtcf.dal. Save this and vtu_heater2 to <topspin home>/exp/stan/nmr/au/src/user. In your starting dataset, set the parameter vtlist to your VT list (eg vt_277_310). Then type au_multivtcf.dal. This will pretty much do the same as multi_zgvt, except that it takes the VT list from the parameter set. Also, the equilibration time is set within the AU program (type edau vt_277_310 and look for the line that says #define EQUILIBTIME to change it) and it doesn't check with you about the maximum and minimum pesrmitted temperatures. However, it will run topshim on the sample so it will be shimmed. (Let me know if you need help setting things up to run Topshim on an unlocked sample with a strong proton signal.)

D.2.5. Liquid Crystals: Christina Redfield's Protocol

The procedure sent to me was adapted from Rückert and Otting (Rückert and Otting 2000). I obtained this protocol from Dr. Christina Redfield *via* email:

> "Here is an extract from a recent student's thesis describing sample preparation. A 1 ml stock solution of 15% **(v/v) in deionized water (MLT)** C12E6 (Dodecylhexaglycol, Sigma) was prepared by gently melting C12E6 in a warm water bath (~60 °C) and transferring 150 µg (~150 µl) into an eppendorf tube containing 835 µl pre-warmed H_2O. At a molar ratio between C12E6 and hexanol of 0.64, the system adopts a lyotropic liquid crystalline phase (Lα). This phase is clear but shows a slight opalescence. The exact balance between C12E6 and hexanol is crucial; at a hexanol concentration below the optimum, the lamellar phase is unstable and the solution cloudy, whereas above the optimum droplets of hexanol are forming because of its insolubility in H_2O. However, the measurement of small quantities of both C12E6 and hexanol is not straightforward because of the high viscosity of C12E6 and the stickiness of hexanol. Therefore, hexanol was added step by step to the solution until the formation of the lamellar phase was initiated (~60 µl). The minimum amount of C12E6 to form a lamellar phase in the presence of protein solution is around ~2%. After the addition of the C12E6 stock solution to the protein solution, a further addition of around 0.5 µl of hexanol was required. The strength of alignment was measured using residual quadrupolar coupling of the D2O signal and this was measured before and after the acquisition of the anisotropic dataset.
> Best wishes,
>
> Christina
> Prof. Christina Redfield
> Dept. of Biochemistry, University of Oxford
> South Parks Road, Oxford, OX1 3QU, U.K.
> Telephone: +44 1865 275330 Fax: +44 1865 613201

A stock of 5% C12E6 (also tried 3%) in 90% deionized water / 10% D_2O was made in a 60˚C water bath. 1-Hexanol (old Sigma bottle from Dr. Wallace) was added in 0.2 µL increments using a 5 µL glass syringe, vigorously shaking the samples between each addition. When the hexanol is initially added, the solution was clear with a top layer of hexanol. But as the amount of hexanol increased, the solution becomes cloudy. Then, at the critical point where the mixture becomes a liquid crystal, the solution becomes spontaneously clear.

Making liquid crystals using this protocol was difficult to get exactly right (Figure 73). I obtained a very good quadrupolar HDO splitting the first time I made the liquid crystals, but every time afterwards I ended up overshooting the mixture and obtaining something between what you see in panels A and C in figure 73. With more experience, I believe that this protocol would be as easy to implement as Pf1-phage since I did obtain an alignment.

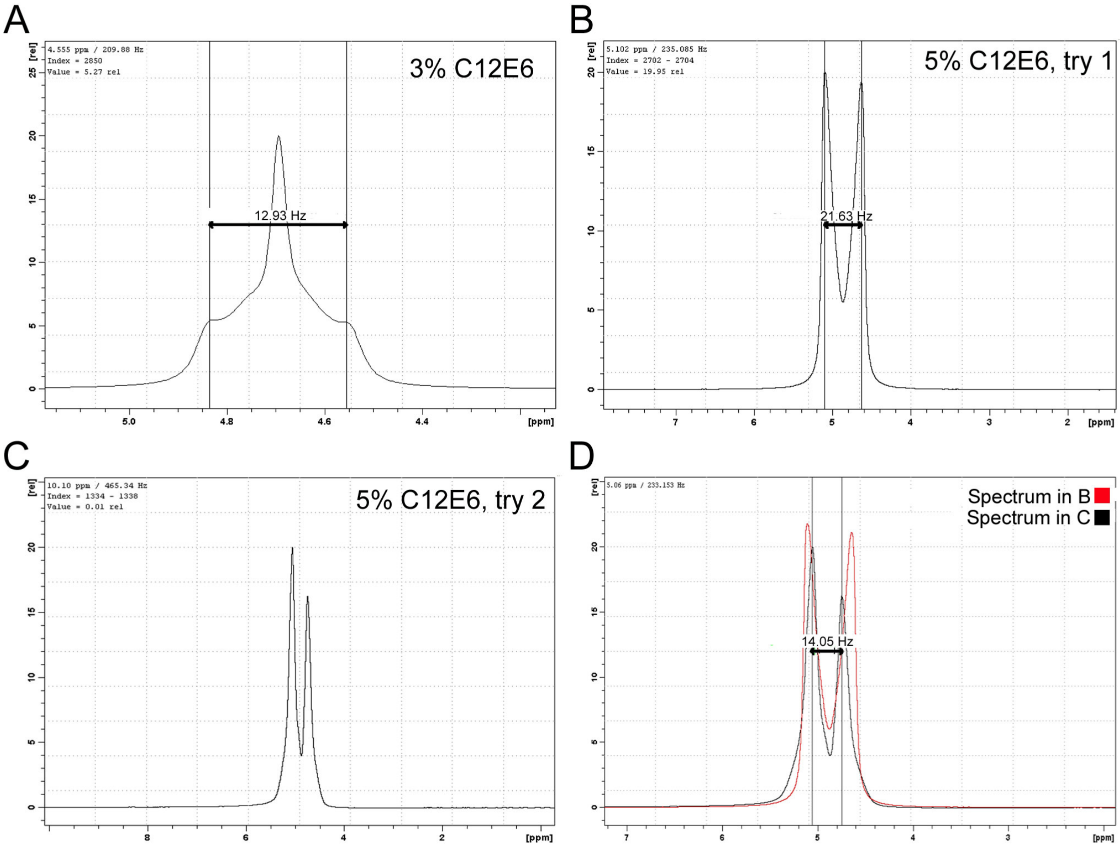


Figure 73 | My attempts at making liquid crystals using the Rückert and Otting method (Rückert and Otting 2000). A) Made with C12E6 at a 3% concentration. There is no real HDO splitting yet. B) First and most successful attempt resulting in a splitting of 21.63 Hz using a composition at 5% C12E6. C) Second attempt at the 5% C12E6 composition. The solution wasn't clear like the solution in panel B. D) Overlay between panel B and C.

D.3. W_1 RDC RESULTS

The medium of choice employed for ^{15}N-enriched W_1 (and W_2) RDCs collection was 15 or 18 mg/mL of Pf1-phage with the addition of 300 mM of NaCl to neutralize the charges on the Pf1 filaments, at pH 7.5. (Without the salt, W_1 completely denatures.) The IPAP experiment described in section D.1.1 was employed for data collection with parameters described in Table 34. Figure 74 shows the non-assigned IPAP spectra overlaid with the normal ^{1}H-^{15}N W_1 HSQC.

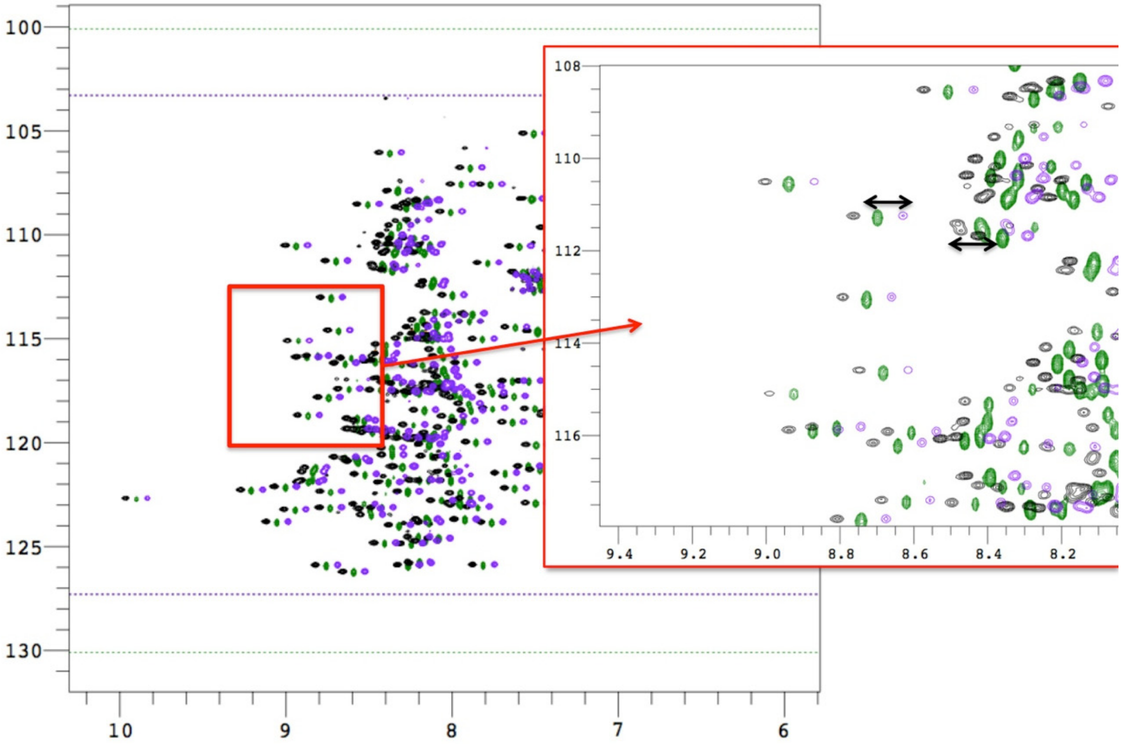


Figure 74 | The ^{15}N-^{1}H HSQC of W$_1$ at pH 6 in ~15mg/mL of Pf1. The green spectrum is a standard HSQC and in black and purple are the α and β components, respectively, of the IPAP HSQC.

IPAP HSQC-based RDCs were acquired at Oxford University in the U.K. Although some peaks change position from pH5 to pH 7.5 (Figure 72), most of them retain the same chemical shifts between the two pHs. Peaks were only assigned when there was minimal or no overlap with other residues and when the peaks could be traced from pH 5 to pH 7.5. Exactly 100 RDCs were extracted from the data at 15 mg/mL of Pf1 with couplings ranging in value from -7.80 to 7.38 Hz (Figure 75). 96 RDCs were extracted from the spectra at 18 mg/mL with values ranging from 16.03 to -16.23 Hz. RDCs in the C-terminal tail were not included in the analysis for the simple reason that they would not be useful due to their high dynamic.

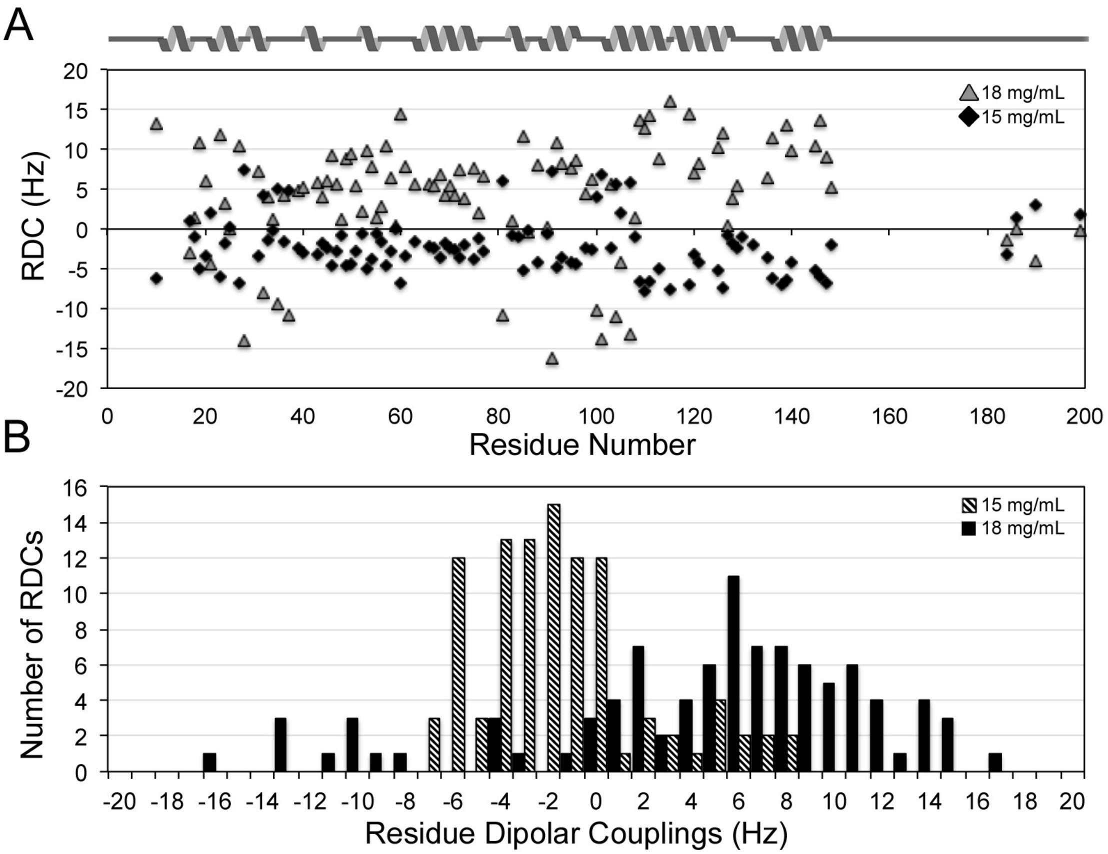


Figure 75 | W_1 RDCs collected in Oxford with Pf1-phage. A) RDC values as a function of residue number. The cartoon structure of W_1 is above the graph. B) The number of RDCs separated into 1 Hz bins for both conditions.

W_2 was selectively labeled with one W repeat unit being ^{15}N-enriched and the other ^{13}C-^{15}N enriched. This selective labeling scheme allowed the determination of RDC of two identical domains within one sample, removing error from acquiring RDCs on two different samples for one protein. 95 RDCs were selected for analysis for W_{2-1} and 94 for W_{2-2}. Although both domains in W_2 have very identical structures, their RDCs are different. The W_{2-1} RDCs range from -4.41 to 12.30 Hz and the W_{2-2} RDCs range from -3.96 to 5.83 Hz (Figure 76). This implies that each W unit may have its own alignment tensor and may display different rhombicity.

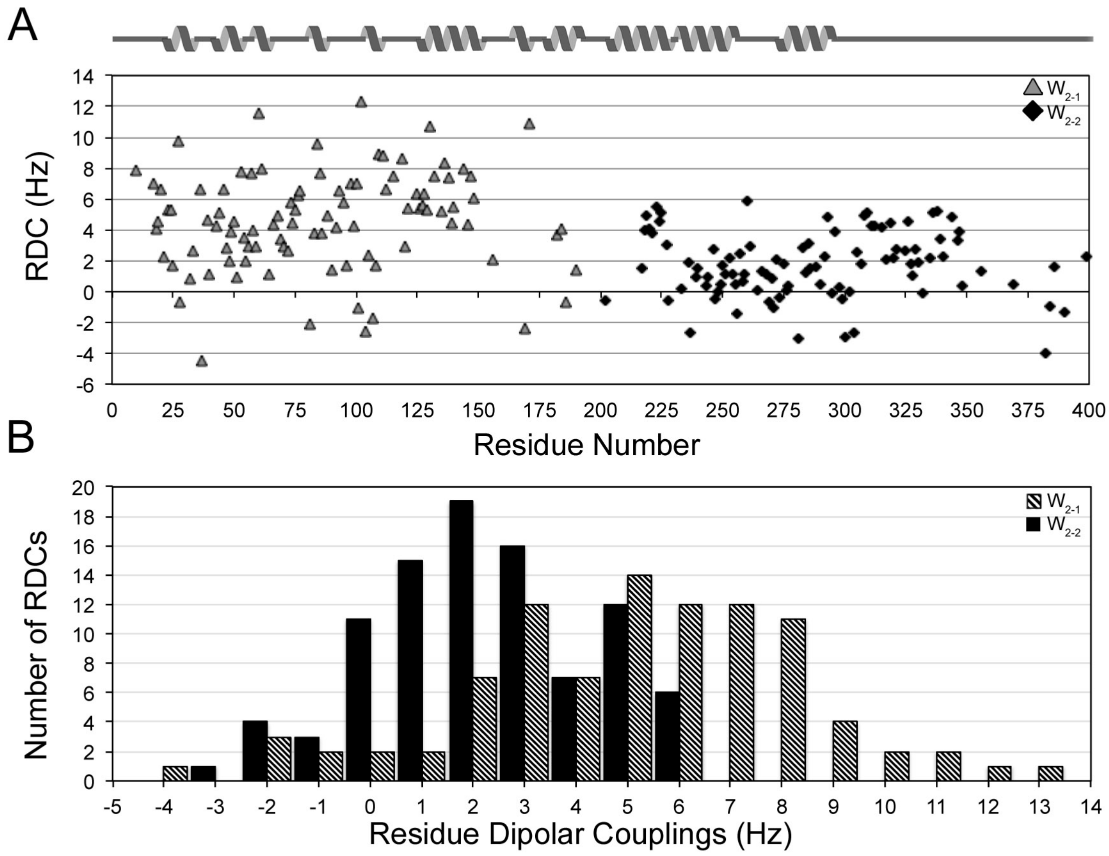


Figure 76 | W_2 RDCs collected at McMaster University with Pf1-phage. A) RDC values as a function of residue number. The cartoon structure of W_1 is above the graph. B) The number of RDCs separated into 1 Hz bins for both conditions.

D.4. INTERPRETING RDCS

To use RDCs in structure calculations or for structure refinement/validation, the alignment tensor (D_a) of all RDCs and the rhombicity (R) of the protein must first be determined. This section describes three methods for determining these parameters. Notably, all RDCs used to determine the alignment tensor and the rhombicity have an associated S^2 value of > 0.7. (See chapter 5 for a detailed description.)

*D.4.1. Method 1: REDCAT (**Re**sidual **D**ipolar **C**oupling **A**nalysis **T**ool)*

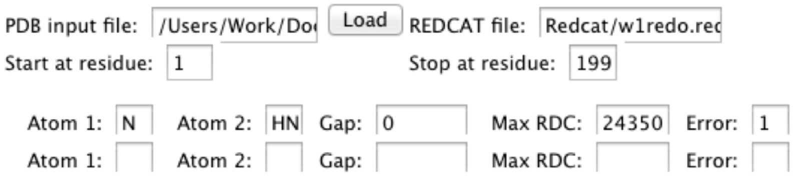
REDCAT (Valafar and Prestegard 2004) is a generally user-friendly program that helps with the analysis of RDCs. As a plus, it includes a GUI, useful for the less *techy* people. At this step, a structure is necessary to obtain validation for fitting. Two other programs that allow the same utility are PALES (Zweckstetter and Bax 2000) and MODULE (Dosset et al. 2001). At the ENC conference in 2014, I received the recommendation that I use MODULES by a postdoc from Martin Blackledge's group at the Institut de Biologie Structurale (IBS, Grenoble, France), but REDCAT performs similarly and was easier to install.

To determine D$_a$ and *R* (or D$_r$), the following equations were used:

$$D_a^{AB} = RDC\max \bullet \frac{S_{zz}}{2}$$

$$D_r = 0.5(S_{xx} - S_{yy})RDC\max$$

RDC$_{max}$ is 24350 (according to the REDCAT software documentation: http://ifestos.cse.sc.edu/REDCAT/documentation/tutorials.php)(Figure 77). The values of *S* are obtained from REDCAT once the RDCs are entered and run through the program (Table 35). The value of S$_{xx}$+S$_{yy}$+S$_{zz}$ should equal 0 (Valafar and Prestegard 2004).



PDB input file: /Users/Work/Do(Load REDCAT file: Redcat/w1redo.rec
Start at residue: 1 Stop at residue: 199
Atom 1: N Atom 2: HN Gap: 0 Max RDC: 24350 Error: 1
Atom 1: Atom 2: Gap: Max RDC: Error:

Figure 77 | REDCAT prepare input file. The website states an RDC from H→ N but it's actually H→ HN. The 'REDCAT file' is the output name that will be given.

Table 35 | Alignment tensor and rhombicity calculated using REDCAT. Here, R=D$_r$/D$_a$

Protein	S$_{xx}$	S$_{yy}$	S$_{zz}$	D$_a$	D$_r$	R
W$_1$	2.00E-04	5.00E-04	-7.00E-04	-8.52	-3.65	0.43
W$_{2-1}$	3.00E-04	7.00E-04	1.00E-03	12.18	-4.87	-0.40
W$_{2-2}$	-3.00E-04	-5.00E-04	8.00E-04	9.74	2.44	0.25

D.4.2. Method 2: The Ad Bax Method

Bax's theory (Clore et al. 1998) states that the dipolar couplings, when plotted in a histogram (bins of 1 Hz, Figure 75 and Figure 76), will resemble a solid-state NMR powder spectrum demonstrating chemical shift anisotropy (Figure 78). The maximum of the peak and both edges are the different CSA tensor components. Therefore the alignment tensor components can be represented by the different areas of the binned RDC histogram. The data actually looks more like a CSA spectrum when there are multiple kinds of RDC present (all normalized to NH of course, the conversion formula is given in (Clore et al. 1998)). Since I only have NH RDCs, it looks more like a weirdly shaped gaussian curve. Either way, this method seems to work well. D_{zz} is the highest RDC value of the protein and D_{yy} is the lowest one. Interestingly, W_1 displays a bimodal curve with a maximum at 16Hz and a minimum at -16Hz (Figure 75). Since D_{xx} is hard to locate, I just used the D_{zz} and D_{yy} 'tensors' to calculate the Da and the R. $R=D_r/D_a$ and D_r and D_a are the alignment and rhombic components of the traceless second rank diagonal tensor D:

$$D_{zz}^{AB} = 2D_a^{AB}$$

$$D_{yy}^{AB} = -D_a^{AB}(1+1.5R)$$

$$D_{xx}^{AB} = -D_a^{AB}(1+1.5R)$$

Plugging and chugging the values from the histogram into the equations, we get a D_a of -8 and an *R* of 0.66. D_{xx} is really hard to determine with only one type of RDC, therefore I only used D_{yy} to determine *R*. The values are presented in Table 36. Of note, I tried combining the RDCs for $W_{2\text{-}1}$ and $W_{2\text{-}2}$. There are two independent sets of RDCs, that theoretically should be treated as different proteins with their own alignment tensor and rhombicity. But as Bermel *et al* (Bermel et al. 2009) stated, the IDIS-RDC-TROSY is explicitly advantageous because all RDCs are measured in the same alignment conditions, therefore we only have the need for one overall tensor. This should also be explicitly true since $W_{2\text{-}1}$ and $W_{2\text{-}2}$ are two components of one protein.

Table 36 | Alignment tensor and rhombicity values determined the Bax method (Clore et al. 1998).

Protein	D_{xx}	D_{yy}	D_{zz}	D_a	*R*
W_1		16	-16	-8	-0.67
W_2	-3	2	-13	-6.5	0.46
$W_{2\text{-}1}$	-3	2	-13	-6.5	0.46
$W_{2\text{-}2}$		2	-6	-3	0.22

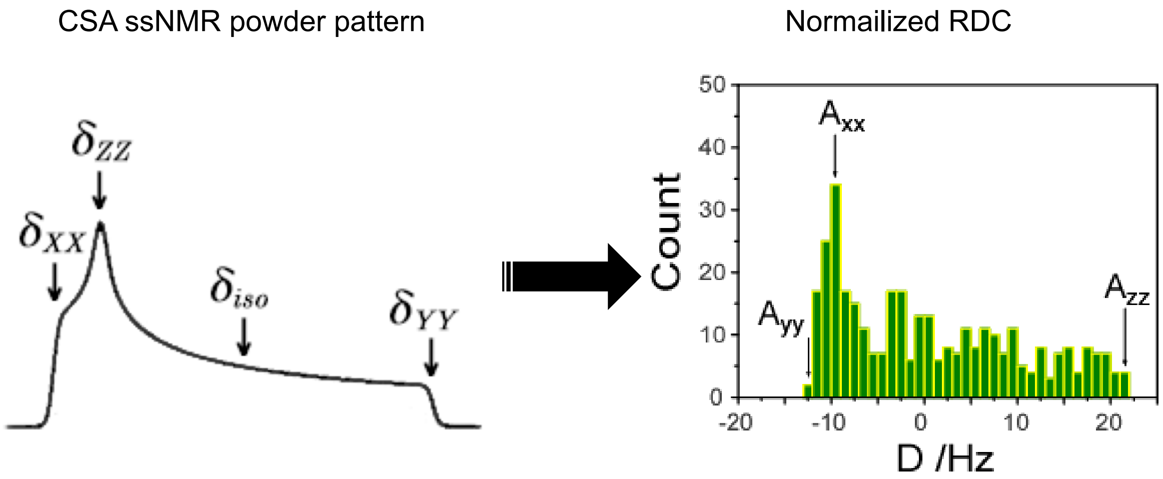


Figure 78 | Comparing a powder solid-state NMR spectrum to the 1 Hz count of the RDCs (Clore et al. 1998).

D.4.3. Method 3: Xplor-NIH

Xplor-NIH provides an alternative method for the calculation of the D_a and *R* tensors using the command calcTensor and a structure: http://t6510.science-biology-xplor-nih-general.biotalk.us/initial-da-and-rh-value-sensitive-t6510.html. At the command line, enter % calcTensor rdc.tbl structure.pdb. When refining a structure, it makes sense to obtain an initial guess from the starting structure, which can be done with W_1 but not W_2. The refine.py script apparently only expects a reasonable initial guess for D_a and R. The simulated structures of W_2, based solely on NOEs and dihedral angles, were used to estimate D_a and R. The following table summarizes the values obtained from this method.

Table 37 | Alignment tensor and rhombicity values determined Xplor-NIH.

Protein	D_a	R
W_1	1.46	0.286
W_2	-2.61	0.312
$W_{2\text{-}1}$	-4.08	0.220
$W_{2\text{-}2}$	-2.80	0.140

D.4.4. Conclusions

The values obtained from Xplor-NIH are drastically different from the other two methods used (Table 38), which led me to believe that it is not a reliable method to use for the calculation of the alignment tensor and the rhombicity. Method 1 and 2 have very similar result for R but not D_a. According to literature, every method seems to be valid.

Since Da and R can be varied within Xplor-NIH implies that the value don't have to be accurate. But this raises the question as to the amount of deviation allowed.

The other trend that I noticed (and expected) was that Da and R were different for W_{2-1} and W_{2-2}.

Table 38 | Summary of the alignment tensor values and rhombicity form all three methods described in this section. 1 represents the values obtained from REDCAT (Valafar and Prestegard 2004), 2 represents the values obtained from Bax's method (Clore et al. 1998), and 3 represents the values obtained from calcTensor within Xplor-NIH .

	D_a			R		
Method	1	2	3	1	2	3
W_1	-8.52	-8.00	1.46	0.43	-0.67	0.29
W_2	n.d.	-6.50	-2.61	n.d.	0.46	0.31
W_{21}	12.18	-6.50	-4.08	-0.40	0.46	0.22
W_{22}	9.74	-3.00	-2.80	0.25	0.22	0.14

D.5. CALCULATING STRUCTURES USING RDCS WITH XPLOR-NIH

The RDC file format for use in structure calculations using Xplor-NIH is:

```
assign ( resid 500 and name OO ) [resid 500 is the residue number for the tensor axis system]
    ( resid 500 and name Z )
    ( resid 500 and name X )
    ( resid 500 and name Y )
    ( resid 1 and name HN ) defines the first atom of the pair
    ( resid 1 and name N ) -8.1 0.5 0.5 [defines the second atom of the pair, RDC_value, Error_value]
```

The coordinate system is represented by four pseudo atoms: OO(origin), X ,Y and Z. The coordinate system has to be positioned far away to prevent any interaction with the protein. For more than one aligned media, define separate axes for each medium. These axes can either be defined at the end of the pdb file or as a separate .pdb file.

W_1 with RDC. Refining the W_1 structure using RDCs failed. Most of the RDCs didn't fit the structure (Figure 79). The residues displayed below all contain an S^2 of >0.7.

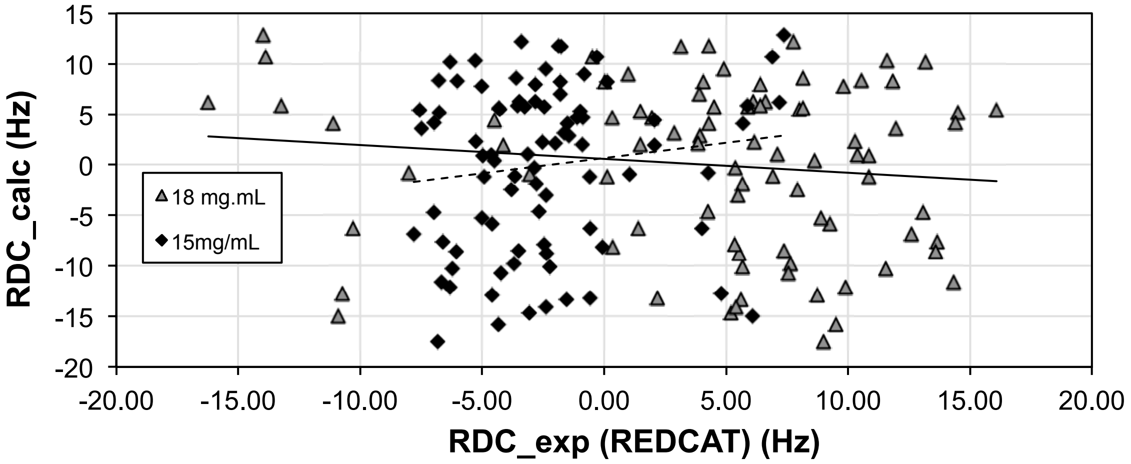


Figure 79 | Fitting of experimental RDCs versus calculated RDCs using REDCAT. The dotted line represents the fitting for the RDCs at 18mg/mL and the solid line is for the fitting at 15 mg/mL of Pf1. Both thread lines have an R^2 value of ~0.01.

W_2 with RDCs. The purpose of the RDCs for W2 was to orient one domain to the others; therefore the RDCs would be included within the structure calculations in the anneal.py script rather than the refine.py. Within the anneal.py script, there is a whole section dedicated to RDCs that require the file input, D_a, R, whether or not to vary the D_a and R or to hold them fixed during structure calculations. RDC violations can also be exported; this can be flagged at the end of the script by adding [rdc] in the potlist for violations. This next section describes the line to be modified. In this case, I used the same D_a and R for both W_2 units:

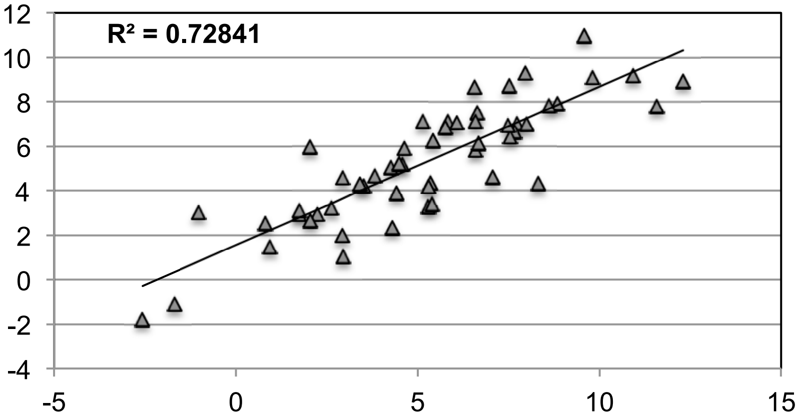
```
#                      medium  Da   rhombicity
for (medium,Da,Rh) in [ ('p1',   -9.294, 0.246),
                       ('p2',   -9.294, 0.342)]:
    oTensor = create_VarTensor(medium)
    oTensor.setDa(Da)
    oTensor.setRh(Rh)
    media[medium] = oTensor
    pass
```

P1 and P2 are the 2 individual sets of RDCs with their own alignment tensor D_a and rhombicity R. If Da and R are the same for both units, than only one file needs to be used. The D_a and the R were varied during all calculations I executed. This is especially useful when the true tensor is unknown or if the true structure of the protein is unknown.

Using the same Da and R for both W_2 unit made the RDC fit the structure a lot better than have two separate D_a and R for each unit. This is sensible considering the RDCs were collected from the same sample in the same orientation. Here again, only RDCs located in rigid portions of W_2 are used for structural refinement with S^2 of >0.7 (Chapter 5 for values). The violation file includes the following lines:

```
RDCPot1  R-fac            R-infinite        Chi^2             Da                Rhombicity
p1_NH    10.740( 2.504)  17.031( 2.437)  0.851( 0.234) -10.093( 1.343)  0.220( 0.059)
p2_NH    23.253( 6.461)  38.037( 3.268)  0.966( 0.164)  -5.063( 0.991)  0.306( 0.106)
```

and these are indicative of the fitting of the RDCs to the structure. 'R-fac' has to be kept low.



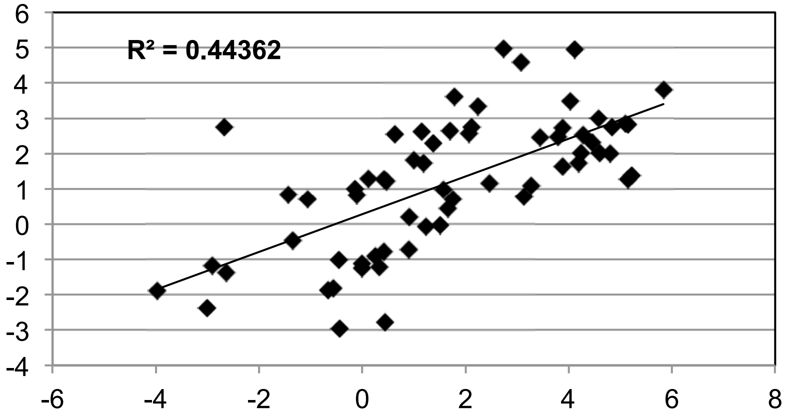


Figure 80 | RDC fitting of experimental and calculated $W_{2\text{-}1}$ (top) and $W_{2\text{-}2}$ (bottom) using the same D_a and for each domain.

D.6. DISCUSSION/CONCLUSIONS

The acquired W_1 RDCs did not match the structure during refinement. I also tried doing structure calculations with RDCs in the anneal.py script, but to no avail. I believe that the greatest problem with the W_1 RDCs matching the structure is that these were collected at a different pH values (5 vs 7.5). As demonstrated in Figure 74, the HSQC's don't exactly line up and some chemical shifts do move significantly. This indicates that the packing of W_1 at pH5 is not the same as at pH 7.5, even though CD implies that the global structuring is extremely similar. Interestingly, the pH within the dragline silk gland also changes during fibrillogenesis - whether this happens in the acinform gland, or not, remains to be determined. A subtle pH-induced structural change might therefore be an indication of minor structural changes that happen within the gland during fibre formation.

W_2 RDCs show more promise simply because I'm modelling the structure based on W_1. Here, RDCs have the potential to orient the folded domains. Since, within the same sample, W_{2-1} and W_{2-2} RDCs differ, this implies that the orientations are different, which shows promise. From chapter 4, I described the structure of W_2 using a radius of gyration (R_g). With this restraint, I was able to demonstrate that the domains are close in proximity, but not so close as to violate NOE restraints. The RDCs might be able to aid the R_g restraint in packing the domains relative to each other. Another possibility would be to include these restraints within a molecular dynamics simulation.

In all, I have heard mixed feelings about using RDCs in structure calculations. Some researchers seem to think that they are not very useful since they rarely fit structures and others seem to apply them with success. I have also been told that I needed more than one set of RDCs or else they will never fit the structure. In the case of W_1, RDCs were not necessary and it would have been better to collect them at the same pH and conditions as the calculated structure. For W_2, they have more use since they are restraints unique to W_2, like $R_{g,}$ whereas NOEs and dihedral restraints were not.

D.7. Lists Of RDCs For W_1 And W_2

D.7.1. W_1 RDCs at 15mg/mL of Pf1 with NaCl

Assign F1	Beta (Hz)	Alpha (Hz)	Coupling	RDC (Hz)
10ThrH	8610.20	8527.12	83.077	-6.296
17LeuH	9375.19	9285.55	89.639	1.055
18IleH	9045.73	8957.12	88.609	-0.962
19SerH	8831.20	8746.33	84.872	-4.966
20ArgH	9383.68	9297.05	86.634	-3.451
21ValH	9054.70	8963.47	91.227	2.083
23AsnH	8878.73	8794.66	84.067	-6.017
24AlaH	9243.04	9154.50	88.541	-1.765
25LeuH	8807.41	8717.47	89.945	0.129
27AsnH	8650.31	8568.14	82.172	-6.764
28ThrH	8408.90	8311.94	96.968	7.386
31LeuH	9043.67	8958.92	84.748	-3.411
32ArgH	8757.12	8662.43	94.688	4.278
33ThrH	8519.71	8432.60	87.108	-1.449
34ValH	9056.65	8965.84	90.815	-0.107
35LeuH	9118.47	9025.15	93.318	4.962
36ArgH	8911.38	8824.70	86.682	-1.500
37ThrH	8949.20	8854.72	94.487	4.808
39ValH	9354.10	9266.56	87.538	-2.383
40SerH	9384.89	9296.05	88.838	-3.068
43IleH	9247.05	9161.31	85.738	-3.248
44AlaH	9271.63	9182.98	88.648	-1.782
45SerH	8448.16	8361.47	86.695	-2.301
46SerH	8940.70	8855.72	84.981	-4.591
47ValH	9206.74	9119.83	86.913	-2.779
48ValH	8886.93	8798.29	88.642	-0.703
49GlnH	9146.56	9060.95	85.603	-4.594
50ArgH	8962.24	8877.19	85.053	-4.317
51AlaH	9315.59	9227.92	87.669	-2.860
52AlaH	9114.72	9025.12	89.599	-0.577
53GlnH	8878.65	8793.92	84.734	-5.000
54SerH	8879.67	8794.03	85.643	-3.801
55LeuH	9391.04	9301.62	89.414	-0.588
56AlaH	9289.19	9200.66	88.527	-1.637
57SerH	8638.24	8553.50	84.741	-4.614
58ThrH	8981.20	8894.54	86.659	-2.808
59LeuH	9025.36	8936.67	88.696	-0.084
60GlyH	8346.93	8263.57	83.355	-6.755
61ValH	8573.53	8488.79	84.744	-3.396
63GlyH	8834.26	8746.51	87.751	-1.538
66LeuH	9087.44	8999.77	87.668	-2.225
67AlaH	9080.14	8992.31	87.827	-2.391
68ArgH	8981.65	8895.55	86.103	-3.642
69PheH	8928.35	8839.72	88.634	-1.866

70AlaH	9380.03	9291.92	88.111	-2.381
71ValH	8999.03	8911.93	87.102	-2.687
72GlnH	9209.24	9122.85	86.393	-3.507
73AlaH	9161.11	9072.70	88.413	-1.983
75SerH	8948.23	8862.83	85.399	-3.699
76ArgH	8974.71	8887.24	87.471	-1.090
77LeuH	9281.76	9193.83	87.932	-2.815
81SerH	8946.90	8850.74	96.156	6.087
83ThrH	8795.76	8708.34	87.423	-0.820
84SerH	8859.82	8770.87	88.942	-1.042
85AlaH	9588.94	9504.40	84.546	-5.266
86TyrH	8990.39	8900.41	89.979	-0.276
88GlnH	9098.93	9013.82	85.108	-4.289
90PheH	8965.05	8875.18	89.868	-0.583
91SerH	8541.20	8443.49	97.713	7.194
92SerH	8653.69	8569.13	84.555	-4.896
93AlaH	9422.65	9335.92	86.733	-3.601
95PheH	8964.72	8879.34	85.381	-4.212
96AsnH	9323.92	9238.41	85.510	-4.487
98GlyH	8077.93	7990.41	87.523	-2.431
99ValH	9082.31	8996.28	86.026	-2.526
100LeuH	8651.39	8560.22	91.172	4.026
101AsnH	8760.77	8663.84	96.928	6.890
103SerH	8548.37	8461.51	86.859	-2.359
104AsnH	9065.04	8969.50	95.540	5.691
105IleH	9267.79	9176.97	90.814	2.084
107ThrH	8477.09	8382.57	94.525	5.872
108LeuH	9330.25	9242.81	87.442	-0.895
109GlyH	8248.33	8164.17	84.166	-6.620
110SerH	8850.93	8765.68	85.246	-7.805
111ArgH	9434.40	9351.77	82.629	-6.673
113LeuH	9108.70	9023.30	85.401	-5.000
115AlaH	9357.20	9274.44	82.762	-7.556
119GlyH	8379.63	8296.04	83.594	-6.965
120ValH	9393.17	9306.71	86.464	-3.131
121SerH	8699.96	8615.16	84.802	-4.296
125GlnH	9035.51	8950.89	84.616	-5.254
126GlyH	8206.20	8122.69	83.517	-7.466
127LeuH	9162.88	9074.91	87.964	-0.860
128GlyH	8253.31	8165.26	88.052	-1.773
129IleH	9329.73	9243.50	86.225	-2.461
130AsnH	9515.72	9428.07	87.649	-0.906
132AspH	9609.60	9522.01	87.584	-2.019
135SerH	9037.86	8952.33	85.532	-3.539
136ValH	9317.77	9234.09	83.684	-6.222
138SerH	8850.49	8767.66	82.836	-6.992
139AspH	9410.22	9326.22	84.001	-6.328
140IleH	9357.08	9271.82	85.252	-4.243
145SerH	9000.51	8916.65	83.854	-5.215
146PheH	9283.62	9200.59	83.854	-6.058

147LeuH	9104.51	9018.35	83.022	-6.816
148SerH	8710.61	8623.53	86.153	-1.953
184GlyH	8303.87	8212.81	87.079	-3.109
186LeuH	9290.62	9200.43	91.052	1.398
190AlaH	9494.33	9403.09	90.196	3.101
199GlyH	8785.56	8696.06	91.243	1.875

D.7.2. W_1 RDCs at 18mg/mL of Pf1 With NaCl

Assign F1	Beta (Hz)	Alpha (Hz)	Coupling	RDC (Hz)
10ThrH	8612.96	8536.75	76.21	13.17
17LeuH	9379.30	9287.67	91.62709	-3.04248
18IleH	9050.28	8962.20	88.08012	1.49088
19SerH	8833.04	8754.05	78.98354	10.85374
20ArgH	9384.17	9300.19	83.97968	6.10571
21ValH	9059.09	8965.46	93.62984	-4.48532
23AsnH	8879.93	8801.70	78.23631	11.84731
24AlaH	9244.51	9157.35	87.15598	3.14961
25LeuH	8812.35	8722.56	89.78674	0.02959
27AsnH	8651.38	8572.98	78.39458	10.5415
28ThrH	8413.43	8309.86	103.56551	-13.98355
31LeuH	9044.85	8963.99	80.86508	7.29348
32ArgH	8761.88	8663.47	98.4112	-8.00164
33ThrH	8523.22	8438.61	84.61918	3.93831
34ValH	9057.46	8967.66	89.79458	1.12803
35LeuH	9121.25	9023.53	97.72359	-9.36786
36ArgH	8912.94	8829.06	83.87509	4.30679
37ThrH	8952.98	8852.56	100.42132	-10.74218
39ValH	9357.13	9272.11	85.02122	4.89971
40SerH	9385.47	9298.75	86.72194	5.18417
43IleH	9249.80	9166.73	83.06906	5.91685
44AlaH	9274.08	9187.71	86.37366	4.05629
45SerH	8451.56	8368.63	82.93879	6.05759
46SerH	8942.18	8861.88	80.30592	9.26609
47ValH	9208.72	9124.71	84.01852	5.67338
48ValH	8890.07	8801.94	88.12476	1.22067
49GlnH	9149.88	9068.43	81.45034	8.74676
50ArgH	8962.95	8883.08	79.86548	9.50423
51AlaH	9317.43	9232.27	85.15983	5.36927
52AlaH	9118.76	9030.76	87.99998	2.17586
53GlnH	8880.71	8800.80	79.91357	9.82048
54SerH	8880.82	8799.30	81.5179	7.92581
55LeuH	9394.01	9305.41	88.59883	1.4041
56AlaH	9293.86	9206.57	87.28441	2.88008
57SerH	8639.90	8560.95	78.95078	10.40436
58ThrH	8983.02	8899.95	83.06702	6.40066
59LeuH	9029.37	8940.93	88.43925	0.34018
60GlyH	8348.70	8273.10	75.59282	14.51694

61ValH	8574.31	8493.94	80.36057	7.77944
63GlyH	8840.36	8756.69	83.66574	5.6229
66LeuH	9090.67	9006.44	84.22418	5.6689
67AlaH	9082.35	8997.53	84.81909	5.3988
68ArgH	8984.19	8901.38	82.81088	6.93421
69PheH	8930.75	8844.56	86.18674	4.31346
70AlaH	9382.60	9297.60	85.00373	5.48842
71ValH	9002.11	8916.58	85.52847	4.2599
72GlnH	9211.33	9128.81	82.52026	7.37924
73AlaH	9164.03	9077.50	86.53298	3.8622
75SerH	8951.13	8869.66	81.46768	7.63125
76ArgH	8976.43	8889.85	86.58828	1.97217
77LeuH	9284.86	9200.72	84.13497	6.6121
81SerH	8950.32	8849.37	100.95348	-10.88422
83ThrH	8799.93	8712.70	87.23465	1.00873
85AlaH	9590.29	9512.09	78.20108	11.61128
86TyrH	8993.41	8902.68	90.72832	-0.47334
88GlnH	9101.10	9019.72	81.3839	8.01294
90PheH	8968.34	8878.01	90.32836	0.12237
91SerH	8546.56	8439.81	106.74813	-16.22957
92SerH	8655.08	8576.50	78.57937	10.87148
93AlaH	9423.83	9341.66	82.17123	8.16301
95PheH	8967.04	8885.00	82.0409	7.55271
96AsnH	9324.92	9243.55	81.36943	8.6281
98GlyH	8081.04	7995.59	85.44824	4.50635
99ValH	9083.58	9001.19	82.38984	6.1615
100LeuH	8655.63	8558.18	97.4446	-10.29807
101AsnH	8765.80	8661.89	103.91019	-13.87159
103SerH	8552.00	8468.32	83.67836	5.53956
104AsnH	9069.09	8968.13	100.95913	-11.1096
105IleH	9271.91	9179.05	92.85651	-4.12588
107ThrH	8482.30	8380.41	101.89144	-13.23842
108LeuH	9334.73	9247.89	86.84252	1.49402
109GlyH	8250.24	8173.09	77.14917	13.63647
110SerH	8855.39	8774.92	80.46469	12.58627
111ArgH	9436.39	9361.41	74.98006	14.32238
113LeuH	9112.43	9030.93	81.5018	8.89905
115AlaH	9358.75	9284.52	74.22823	16.0897
119GlyH	8379.66	8303.52	76.13979	14.41934
120ValH	9395.52	9313.03	82.48564	7.10965
121SerH	8703.73	8622.79	80.9391	8.15882
125GlnH	9038.46	8958.88	79.57614	10.29397
126GlyH	8207.52	8128.51	79.00615	11.977
127LeuH	9165.46	9076.99	88.47133	0.35233
128GlyH	8257.00	8171.09	85.90875	3.91661
129IleH	9331.50	9248.16	83.3394	5.34698
135SerH	9040.70	8958.03	82.66531	6.40564
136ValH	9320.47	9242.11	78.36779	11.53858
139AspH	9412.37	9335.09	77.27161	13.05766
140IleH	9359.62	9280.01	79.61325	9.88148

145SerH	9001.40	8922.71	78.68644	10.38197
146PheH	9286.11	9209.78	76.32996	13.58144
147LeuH	9106.71	9025.88	80.8295	9.00812
148SerH	8712.52	8629.70	82.82123	5.28422
184GlyH	8307.24	8215.75	91.49473	-1.30728
186LeuH	9293.43	9203.76	89.66943	-0.01503
190AlaH	9496.51	9405.40	91.10381	-4.0086
199GlyH	8788.77	8699.22	89.54742	-0.17923

D.7.3. W_{2-1} RDCs at 18 mg/mL Pf1 With NaCl

Assign F1	Beta (Hz)	Alpha (Hz)	Coupling	RDC (Hz)
10ThrN	5731.83	5647.60	84.24	7.884
17LeuN	5316.26	5228.26	88.01	6.984
18IleN	5318.06	5229.23	88.83	4.100
19SerN	5706.99	5621.64	85.34	4.580
20ArgN	5555.71	5468.83	86.88	6.594
21ValN	5509.15	5418.68	90.47	2.245
23AsnN	6101.55	6016.41	85.13	5.353
24AlaN	5343.02	5255.10	87.92	5.286
25LeuN	5544.33	5455.00	89.33	1.747
27AsnN	5012.31	4928.64	83.67	9.811
28ThrN	5245.04	5150.80	94.24	-0.692
32ArgN	5385.23	5292.08	93.15	0.816
33ThrN	5495.28	5407.89	87.39	2.621
36ArgN	4455.63	4372.15	83.48	6.682
37ThrN	5892.05	5795.47	96.58	-4.412
39ValN	5691.83	5604.06	87.77	4.630
40SerN	5951.40	5861.95	89.44	1.180
43IleN	5277.98	5189.42	88.56	4.257
44AlaN	6019.88	5933.04	86.84	5.153
46SerN	5400.02	5314.10	85.92	6.648
47ValN	6106.89	6019.75	87.14	2.824
48ValN	5778.82	5689.54	89.28	2.038
49GlnN	5853.29	5765.02	88.27	3.883
50ArgN	5663.31	5575.89	87.42	4.581
51AlaN	6359.08	6269.98	89.10	0.940
53GlnN	5602.49	5517.25	85.24	7.744
54SerN	5866.50	5779.33	87.17	3.531
55LeuN	6005.80	5917.54	88.25	2.039
56AlaN	5801.88	5712.23	89.65	2.938
57SerN	5650.65	5565.19	85.46	7.672
58ThrN	5706.40	5618.88	87.52	3.976
59LeuN	5595.30	5506.99	88.31	2.941
60GlyN	5661.44	5577.59	83.86	11.568
61ValN	5363.12	5280.91	82.21	7.995
64AsnN	5899.86	5805.79	94.07	1.168

66LeuN	5615.62	5526.30	89.31	4.330
68ArgN	5483.83	5395.93	87.90	4.940
69PheN	5687.87	5599.20	88.68	3.416
70AlaN	6160.20	6071.87	88.33	2.923
71ValN	6165.48	6077.28	88.19	2.632
72GlnN	5757.10	5670.45	86.65	5.829
73AlaN	5568.52	5479.84	88.68	4.423
75SerN	5911.73	5825.48	86.25	5.302
76ArgN	4921.44	4835.61	85.83	6.273
77LeuN	5180.59	5093.26	87.33	6.574
81SerN	5337.94	5242.36	95.58	-2.098
83ThrN	5870.11	5782.82	87.29	3.830
84SerN	6087.75	6001.81	85.94	9.581
85AlaN	5513.91	5428.84	85.07	7.686
86TyrN	5381.33	5292.63	88.70	3.826
88GlnN	5800.27	5713.61	86.66	4.887
90PheN	5946.98	5856.51	90.47	1.387
92SerN	6169.79	6083.22	86.58	4.213
93AlaN	5596.83	5510.67	86.16	6.587
95PheN	5496.71	5411.49	85.22	5.770
96AsnN	6510.40	6422.60	87.80	1.747
98GlyN	5316.10	5228.70	87.39	7.058
99ValN	5405.24	5318.77	86.47	4.308
100LeuN	4539.45	4451.56	87.89	6.988
101AsnN	6121.18	6029.26	91.91	-1.015
102AlaN	6239.82	6154.83	84.99	12.297
104AsnN	5664.54	5572.20	92.35	-2.566
105IleN	5436.80	5346.89	89.91	2.377
107ThrN	5373.13	5280.61	92.52	-1.696
108LeuN	6166.89	6080.59	86.30	1.697
109GlyN	5621.63	5536.28	85.35	8.870
111ArgN	5482.21	5398.66	83.55	8.849
112ValN	5907.10	5822.10	85.00	6.686
115AlaN	5583.03	5499.17	83.85	7.535
119GlyN	5801.16	5715.96	85.19	8.609
120ValN	6312.83	6225.51	87.32	2.946
121SerN	5728.10	5641.34	86.76	5.388
125GlnN	5771.42	5686.55	84.87	6.376
126GlyN	5589.91	5503.44	86.47	5.428
127LeuN	5295.85	5208.66	87.20	5.615
128GlyN	5491.66	5403.63	88.03	6.352
129IleN	5618.31	5532.93	85.38	5.356
130AsnN	5947.79	5865.60	82.19	10.696
132AspN	6067.57	5980.59	86.98	7.481
135SerN	5573.03	5486.63	86.39	5.249
136ValN	5400.34	5315.15	85.19	8.310
138SerN	5683.67	5598.62	85.05	7.415
139AspN	5983.98	5897.07	86.92	4.494
140IleN	6015.68	5930.08	85.60	5.478
144SerN	5940.23	5856.37	83.86	7.960

146PheN	5744.34	5659.86	84.48	4.328
147LeuN	5654.56	5569.51	85.05	7.522
148SerN	5572.56	5485.90	86.66	6.076
156TyrN	5671.24	5579.24	92.00	2.090
169TyrN	5679.23	5586.77	92.46	-2.327
171GlyN	5144.60	5052.68	91.93	10.934
182TyrN	5639.80	5550.35	89.45	3.713
184GlyN	5308.17	5216.64	91.53	4.124
186LeuN	5922.76	5832.14	90.62	-0.694
190AlaN	5705.71	5614.79	90.92	1.390

D.7.4. W$_{2-2}$ RDCs at 18 mg/mL Pf1 With NaCl

Assign F1	Beta (Hz)	Alpha (Hz)	Coupling	RDC (Hz)
2GlyH	5692.62	5599.47	93.152	-0.591
10ThrH	5736.59	5649.42	87.173	3.719
17LeuH	5316.74	5227.14	89.604	1.514
18IleH	5318.05	5230.59	87.459	4.027
19SerH	5708.06	5622.16	85.892	4.903
20ArgH	5559.51	5472.57	86.947	4.116
21ValH	5510.03	5421.80	88.235	3.797
23AsnH	6102.86	6016.01	86.854	5.483
24AlaH	5344.53	5256.88	87.647	4.576
25LeuH	5545.17	5457.59	87.577	5.137
27AsnH	5016.68	4930.11	86.569	3.073
28ThrH	5245.49	5154.02	91.476	-0.555
33ThrH	5497.50	5407.66	89.844	0.167
36ArgH	4462.54	4376.72	85.819	1.860
37ThrH	5892.46	5796.83	95.622	-2.631
39ValH	5695.55	5605.11	90.441	0.909
40SerH	5953.03	5863.09	89.940	1.475
43IleH	5279.77	5189.03	90.740	0.416
44AlaH	6026.12	5934.25	91.871	0.907
46SerH	5403.15	5314.02	89.129	2.796
47ValH	6111.06	6018.79	92.268	-0.446
48ValH	5783.36	5690.88	92.484	-0.002
49GlnH	5854.77	5763.52	91.247	0.472
50ArgH	5667.09	5578.36	88.722	1.692
51AlaH	6362.45	6270.56	91.899	1.103
53GlnH	5607.57	5517.82	89.749	2.172
54SerH	5869.56	5779.04	90.512	1.156
55LeuH	6008.51	5916.90	91.609	0.453
56AlaH	5804.01	5711.08	92.924	-1.432
57SerH	5653.39	5564.90	88.498	2.464
58ThrH	5709.70	5618.77	90.931	0.627
59LeuH	5596.45	5506.57	89.888	1.182
60GlyH	5662.21	5574.63	87.578	5.835
61ValH	5363.04	5278.62	84.423	2.908

64AsnH	5902.93	5811.29	91.635	0.140
66LeuH	5622.00	5531.20	90.800	1.370
68ArgH	5489.22	5397.98	91.238	1.141
69PheH	5696.42	5604.81	91.606	-0.665
70AlaH	6162.19	6070.70	91.497	0.842
71ValH	6169.57	6077.21	92.364	-1.054
72GlnH	5759.88	5669.50	90.381	2.117
73AlaH	5573.75	5481.10	92.646	-0.382
75SerH	5910.13	5820.69	89.436	1.780
76ArgH	4928.80	4839.37	89.424	0.123
77LeuH	5182.52	5091.48	91.039	0.419
81SerH	5338.91	5243.91	95.001	-2.999
83ThrH	5872.66	5785.18	87.482	2.861
84SerH	6093.56	6004.93	88.632	1.237
85AlaH	5515.06	5426.60	88.460	3.130
86TyrH	5379.81	5290.01	89.801	1.530
88GlnH	5805.06	5714.96	90.103	1.572
90PheH	5949.52	5858.65	90.865	0.447
92SerH	6169.30	6079.62	89.688	2.261
93AlaH	5598.57	5510.33	88.242	4.801
95PheH	5499.97	5409.45	90.524	-0.099
96AsnH	6508.92	6419.44	89.486	3.874
98GlyH	5320.67	5228.66	92.017	0.255
99ValH	5405.91	5316.19	89.720	-0.438
100LeuH	4539.33	4451.81	87.518	-2.903
102AlaH	6242.33	6150.80	91.532	0.000
104AsnH	5663.93	5572.18	91.750	-2.667
105IleH	5435.66	5347.16	88.501	2.565
107ThrH	5372.99	5284.93	88.063	1.825
108LeuH	6165.09	6079.93	85.156	4.886
109GlyH	5625.28	5536.35	88.928	5.142
111ArgH	5487.38	5400.84	86.534	4.274
112ValH	5908.32	5820.50	87.821	4.236
115AlaH	5584.39	5496.70	87.694	4.194
101AsnH	6060.71	5970.42	90.295	2.064
119GlyH	5805.25	5715.79	89.461	4.451
120ValH	6313.13	6223.79	89.335	2.143
121SerH	5731.88	5642.05	89.835	2.795
125GlnH	5775.88	5686.29	89.589	2.609
126GlyH	5594.28	5506.17	88.111	4.595
127LeuH	5298.28	5210.51	87.765	1.756
128GlyH	5495.20	5404.01	91.193	1.007
129IleH	5621.70	5533.82	87.880	2.737
130AsnH	5956.61	5867.76	88.851	1.898
132AspH	6069.42	5976.88	92.533	-0.135
135SerH	5577.99	5487.93	90.061	2.223
136ValH	5405.43	5318.36	87.069	5.094
138SerH	5688.42	5601.14	87.284	5.221
139AspH	5985.82	5897.39	88.429	3.448
140IleH	6017.46	5928.11	89.357	2.243

144SerH	5943.82	5856.33	87.490	4.836
146PheH	5748.35	5661.09	87.265	3.275
147LeuH	5657.82	5570.53	87.287	3.879
148SerH	5574.70	5486.96	87.749	0.336
156TyrH	5672.52	5579.17	93.347	1.281
169TyrH	5678.57	5586.43	92.140	0.502
182TyrH	5640.41	5547.92	92.491	-3.967
184GlyH	5312.01	5218.92	93.094	-0.939
186LeuH	5926.13	5834.49	91.632	1.660
190AlaH	5709.64	5617.71	91.933	-1.345
199GlyH	5297.51	5205.81	91.692	2.297

APPENDIX E. SOLID-STATE NMR ON SELECTIVELY LABELED W_3

E.1. PROJECT INTRODUCTION

Although the solution structure of W1 is quite useful for investigating and characterizing fibrillogenesis, the major goal is to determine the structure of the AcSp1 protein in the fibre. For this reason, I moved from solution-state to solid-state NMR. Using magic angle spinning (MAS), it is possible to investigate, at the atomic level, the structure of solid crystalline proteins. But solid-state NMR doesn't have the resolution of solution-state NMR, and peaks may still be quite broad, depending on the sample. Since W_1 doesn't form fibres, I moved to larger construct. W_3 was chosen as the ideal candidate since it forms better fibres than W_2 (Figure 81), with better reproduction of the mechanical properties of native silk than those of W_2 (Xu et al. 2012a)

Our end goal was to contrast the structure of wrapping silk protein in solution to that in the fibre. We know that there is a transition from helix-rich to a mixed helix/sheet structure during fibre formation (evidence in Figure 67, (Lefèvre et al. 2011)). The intrinsically disordered C-terminal tail may act as a "seeding" site for β-strand formation, since the energetic and steric barriers to strand formation would be lowest in this region. Initially, we started by exploring different labeling schemes. The first was selective unlabeling (Atreya and Chary 2001), which would be amenable for ssNMR at our local facilities, but we quickly moved on to explore segmental intein-based labeling instead. We produced a W_3 protein with only the C-terminal 38 residues are enriched with ^{15}N and ^{13}C. With this construct, I wanted to compare the structure of this final C-terminal region in 1) *lyophilized powder* and 2) in a *fibre* (Figure 82).

I started by testing the feasibility locally. MAS solid-state NMR experiments on lyophilized W_3 powder were carried out at 9.4 T at NMR-3. Based on our local expertise, pulse sequence availability and probe availability, a ^{13}C CP-MAS was initially performed and, after some pulse programming, INADEQUATE experiments. Then, we moved the experiments over to NHMFL in Tallahassee, FL, where the equipment and personnel were more equipped to deal with proteins.

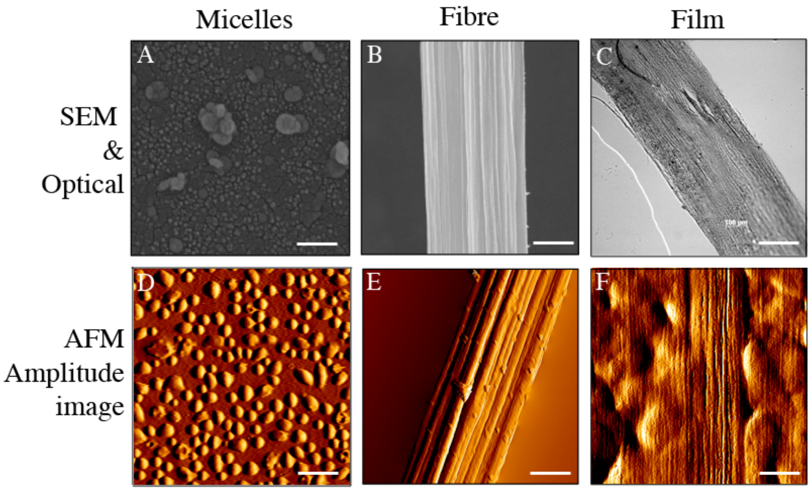


Figure 81 | Scanning electron microscopy (SEM) and atomic force microscopy (AFM) images of silk micelles (W1), a fibres (W3), and films (W3).

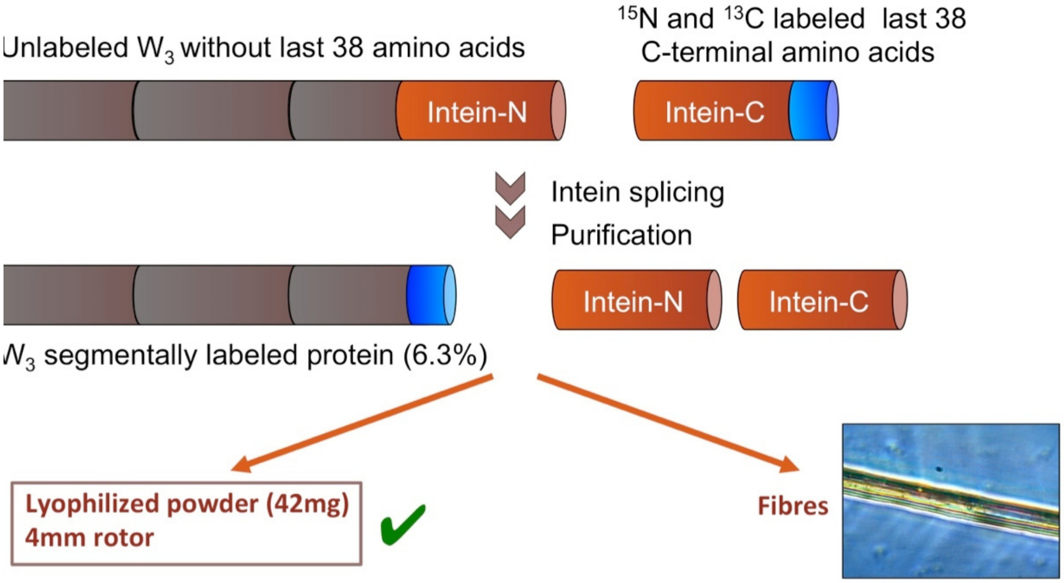

Unlabeled W3 without last 38 amino acids
15N and 13C labeled last 38 C-terminal amino acids
Intein-N
Intein-C
Intein splicing
Purification
W3 segmentally labeled protein (6.3%)
Intein-N
Intein-C
Lyophilized powder (42mg)
4mm rotor
✓
Fibres

Figure 82 | Protein construct used for the solid-state experiments. The inteins are in orange, the brown rectangles are W_1 at natural abundance and in blue in the $^{13}C/^{15}N$ labeled last 38 amino acids in W_3. The lyophilized W_3 powder was then used to pack the rotor. The alternative to the powder is solving the structure of the C-terminal labeled portion within fibres.

E.2. MATERIALS AND METHODS

E.2.1. Selective Un-Labeling Of W_1

W_1 was expressed by Lingling Xu according to our previously published ^{15}N labeling protocol (Xu et al. 2012b). To un-label, we followed a protocol by Atreya (Atreya and Chary 2001) in which natural abundance amino acids are introduced in the bacterial medium upon expression to be incorporated in the protein as a source of amino acids. The amino acids that were chosen for ^{15}N labeling were Q, R, P, and D. To test the efficiency of labeling, a ^{15}N-HSQC was acquired on the soluble W$_1$ sample in NMR buffer (20 mM Acetate, pH 5, 1 mM DSS)

E.2.2. Expression And ^{13}C/^{15}N Labeling Of W$_3$ For Solid-State NMR Experiments

Lingling Xu performed the expression and labeling of the ^{13}C/^{15}N W$_3$ NMR sample. We chose this specific spot within the protein because the inteins require a cysteine, serine, or a threonine to undergo the protein ligation reaction. The sequence of W$_3$ for the solid-state NMR experiments is as follows:

1 AGPQGGFGAT GGASAGLISR VANALANTST LRTVLRTGVS QQIASSVVQR
51 AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
101 NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSSSSFLSTS
151 SSSASYSQAS ASSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGGS
201 AGPQGGFGAT GGASAGLISR VANALANTST LRTVLRTGVS QQIASSVVQR
251 AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
301 NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSSSSFLSTS
351 SSSASYSQAS ASSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGGS
401 AGPQGGFGAT GGASAGLISR VANALANTST LRTVLRTGVS QQIASSVVQR
451 AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
501 NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSSSSFLSTS
551 SSSASYSQAS A**SSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGG**

Figure 83 | Labeled sequence for ssNMR construct brought to the NHMFL. In bold are the ^{13}C/^{15}N enriched residues. Blue, orange, and red denotes the individual W units.

and the bolded and underlined sequence is labeled and the remaining part of the protein is left t natural abundance.

E.2.3. NMR Spectroscopy

The ssNMR data was acquired on a Bruker Avance 9.4 T with a 4 mm ^{1}HX MAS probe at the Dalhousie University Nuclear Magnetic Resonance Research Resource (NMR3) and on a Bruker Avance 14.1 T equipped with a Low-E 3.2 mm MAS ^{1}HXY probe at the National High Magnetic Field Laboratory (NHMFL). The experiments collected on this sample, with accompanying parameters, are outlined in Table 39.

The data collected locally on the 9.4 T Bruker Avance instrument was acquired in a 4 mm rotor with a two-channel MAS-probe (using ^{13}C and ^{1}H) and spinning at 9 kHz. To obtain a reasonable level of signal-to-noise, the Incredible Natural Abundance DoublE QUAntum Transfer Experiment (INADEQUATE) (Lesage et al. 1999) spectrum was acquired with 14,400 scans over 38 increments, requiring two weeks of instrument time to obtain the spectrum.

The dipolar assisted rotational resonance (DARR) (Takegoshi et al. 2001) experiments were collect at 10 ms and 50 ms, where a short delay only displays peaks within residues and 50ms shows long range contacts. The 2D NCA and NCO (Baldus et al. 1998) took about 1 day each for acquisition and the 3D NCACO (Pauli et al. 2001) experiment took about 3 days for acquisition.

Table 39 | List of experiments conducted on W_3 labeled at the C-terminal 38 amino acids.

Experiment	Bruker pulse sequence	Spinning speed (kHz)	Relaxation Delay (s)	Number of scans	Number of points F2|F1	Spectral width (ppm)	Center position (ppm)	^{1}H Freq.	Facility
^{13}C CP-MAS	CPvarTPPM.uwz	9	5	96	1798	496	110	400	NMR3
INADEQUATE	cpinadrd.98.uwz	9	3	14400	994|128	496|268	70	400	NMR3
^{13}C CP-MAS	cp90.ih	10	2	360	3072	496	103	600	NHMFL
^{13}C-^{13}C DARR	cpdarr.ih	10	2	96	3072|200	248|274	103|126	600	NHMFL
NCA	NCA.ih	10	2	360	3072|16	82|82	52.8|119	600	NHMFL
NCO	NCO.ih	10	2	360	3072|16	82|82	175.8|119	600	NHMFL
NCACX	NCACO.ih	10	2	1296	3072|16|6	186|33.1|54.8	48.2|48.2|116.8	600	NHMFL

The data were processed in TopSpin v2.1 (Bruker) and imported into CcpNMR Analysis (Vranken et al. 2005) for data analysis.

E.3. RESULTS

E.3.1. Selective Un-Labeling

This method for labeling is mostly effective, but some of the residues expected to be un-labeled still appeared to be enriched in ^{15}N (Figure 84). In this method, the protein is expressed as per usual for ^{13}C and/or ^{15}N labeling using minimal media, but are unlabeled by the re-incorporation of natural abundance amino acids. This way, the only amino acids that ere labeled were the selected ones. We tried this labeling method on W_1 since it would be rather easy to identify its efficiency by acquiring a simple ^{15}N HSQC then move on to larger W proteins.

We tried different labeling schemes using different combinations of amino acids but every time there was isotope leakage. As identified in Figure 84 in green, there are a few more labeled amino acids than anticipated. In the end, we decided that this wasn't an optimal method and moved onto intein-based selective labeling instead.

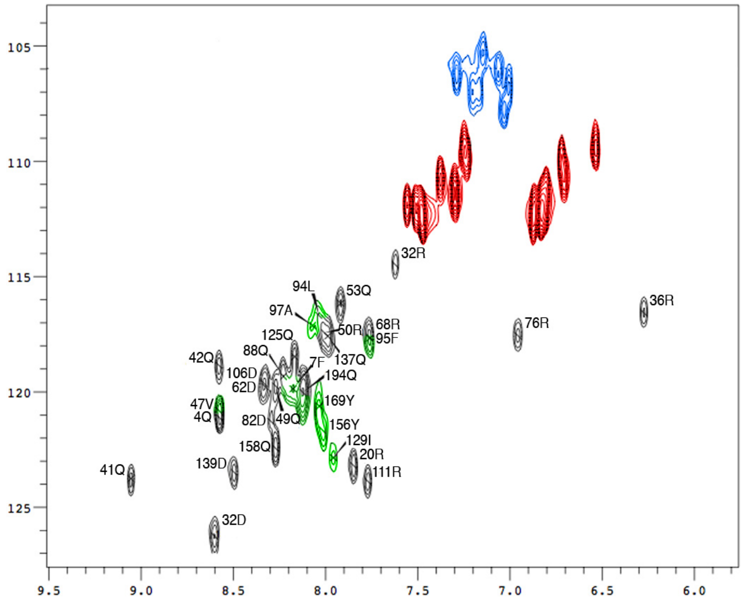


Figure 84 | ^{1}H-^{15}N HSQC of selectively unlabeled W$_1$. In red are some side chains of Asn and Gln. In blue are the Arg Ne and in green are peaks that are not supposed to be labeled. The only residues that were supposed to be labeled are Gln, Arg, Asp, and Pro.

E.3.2. W$_3$-38Cterm ssNMR

The last 38 amino acids of W$_3$ (termed W$_3$-38Cterm) were selectively enriched with ^{13}C/^{15}N isotopes using split-intein *trans* splicing. About 42 mg of lyophilized protein was produced for packing within a 4 mm rotor. Since it was the first time that a protein of this nature was analyzed at NMR-3, pulse sequences were tested out using a 4 mm rotor packed with ^{13}C-labeled valine. Then, to test inter-connectivity on a larger molecule, I over-expressed and labeled, with ^{13}C and ^{15}N isotopes in minimal medium, GB1. The INADEQUATE was tested with GB1 to verify the functionality of the pulse sequence.

When all conditions were tested, W$_3$-38Cterm was placed in the magnet. The resulting CP-MAS is presented in Figure 85. The appropriate spinning speed to use was tested by repeated CP-MAS spectra with the aim of positioning the spinning side bands such that they did not interfere with the protein peaks. Given the amount of labeled protein in the rotor, a high degree of isotope incorporation is apparent.

The signal-to-noise of the INADEQUATE experiment was very low. For this reason, we acquired the spectrum for 2 weeks. Although the signal-to-noise was high enough for analysis, the overlap between the peaks is overwhelming. For this reason, assignment was very difficult. Using random coil chemical shift tables (Wishart et al. 1995b; Tremblay et al. 2010), I predicted the INADEQUATE spectra in order to help with the assignment. Luckily, the amino sequence is not too repetitive and there are some amino acids that are unique, like Leu186, but others that are never going to be assigned, like the glycines. As we can see from Figure 86, there is a large peak for the two Ala Cβ with no distinction between either Cβ. There are also some unique chemical shifts for serine and threonine, both of which are abundant in the sequence.

The INADEQUATE was definitely not the best spectrum to acquire first but it was chosen based on the simplicity of the pulse sequence and the information that we are able to gather from it. Since the W_3-38Cterm was uniformly enriched, I had hoped that the signal would be much better than it was. We then moved on to seek the help of experts in the field at the NHMFL.

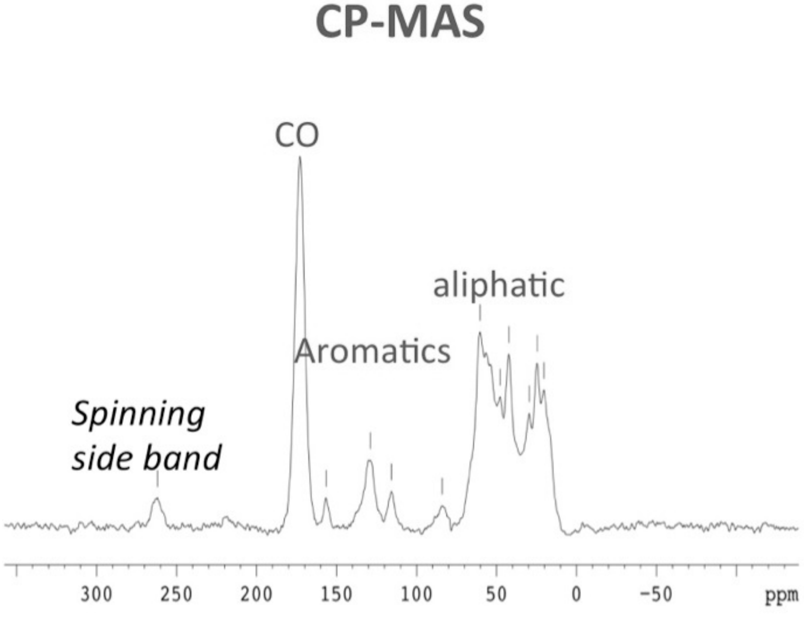


Figure 85 | 1D cross-polarization 13C spectrum of W_3-38Cterm, with 9 khz spin for 96 scans.

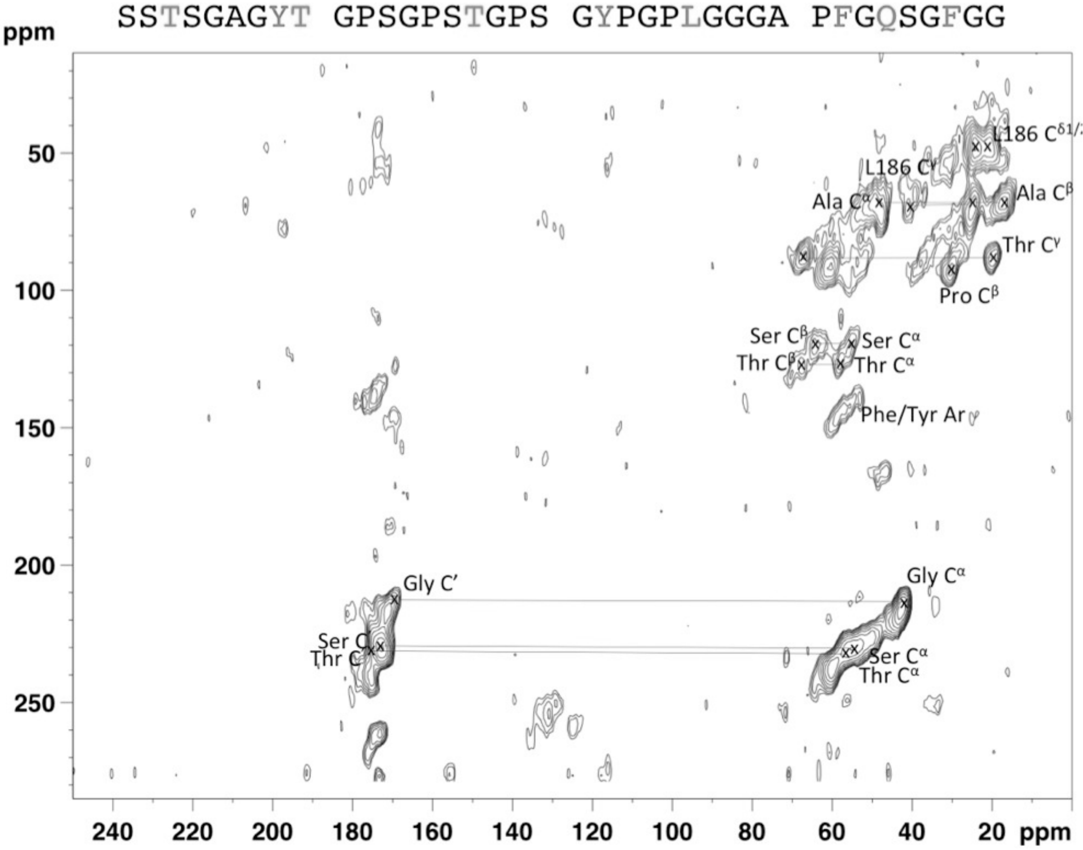


Figure 86 | INADEQUATE spectrum labeled with possible assignments.

E.3.3. NHMFL Data Acquisition

The NMR data were acquired, processed within TopSpin or using NMRpipe (Delaglio et al. 1995). The spectra were imported in Analysis 2.3. As expected, there is a difference between the DARR spectra (Bertini et al. 2010) at 10ms and 50 ms (Figure 87); there are more peaks in DARR at 50 ms, meaning that we acquired information on inter-residue contacts. In the DARR, the magnetization is transferred from ^{1}H to ^{13}C, then to other ^{13}C that are dipolar-coupled through proximity in space. Overlaying both spectra show the long distance peaks. Using random coil chemical shift tables (Wishart et al. 1995b; Tremblay et al. 2010), some chemical shifts were assigned but only per amino acid type rather than unique residues. The peaks were too broad to differentiate between unique frequencies (Figure 88), unfortunately.

The 2D NCA and NCO provided no information on W$_3$-38Cterm. The peaks are concentrated into one big huge peak, even more so with the NCO than the NCA (Figure

89). The glycine Cα peaks are visibly distinct from other Cα in the NCA but no individual peaks are discernable.

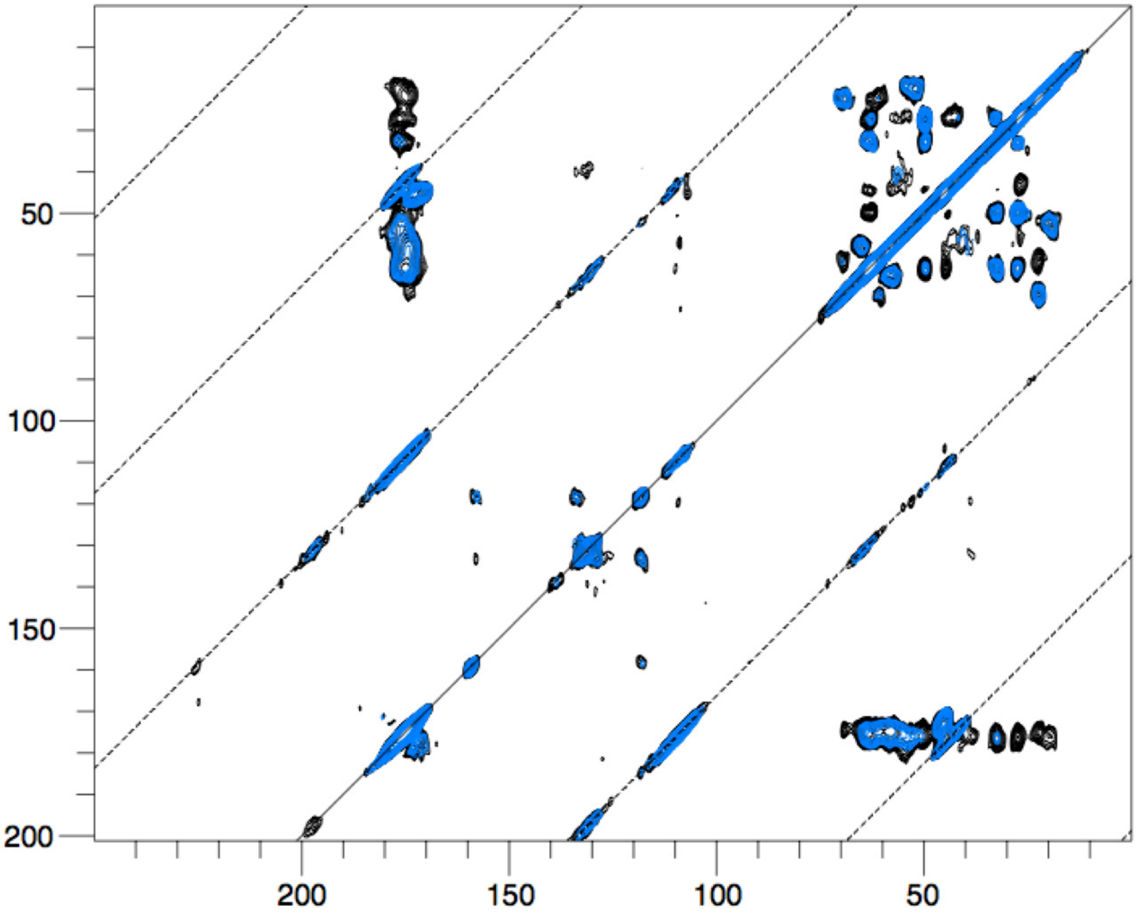


Figure 87 | W_3-38Cterm DARR spectra. Blue = 10ms and black = 50ms. Diagonal dotted lines represent the spinning side bands.

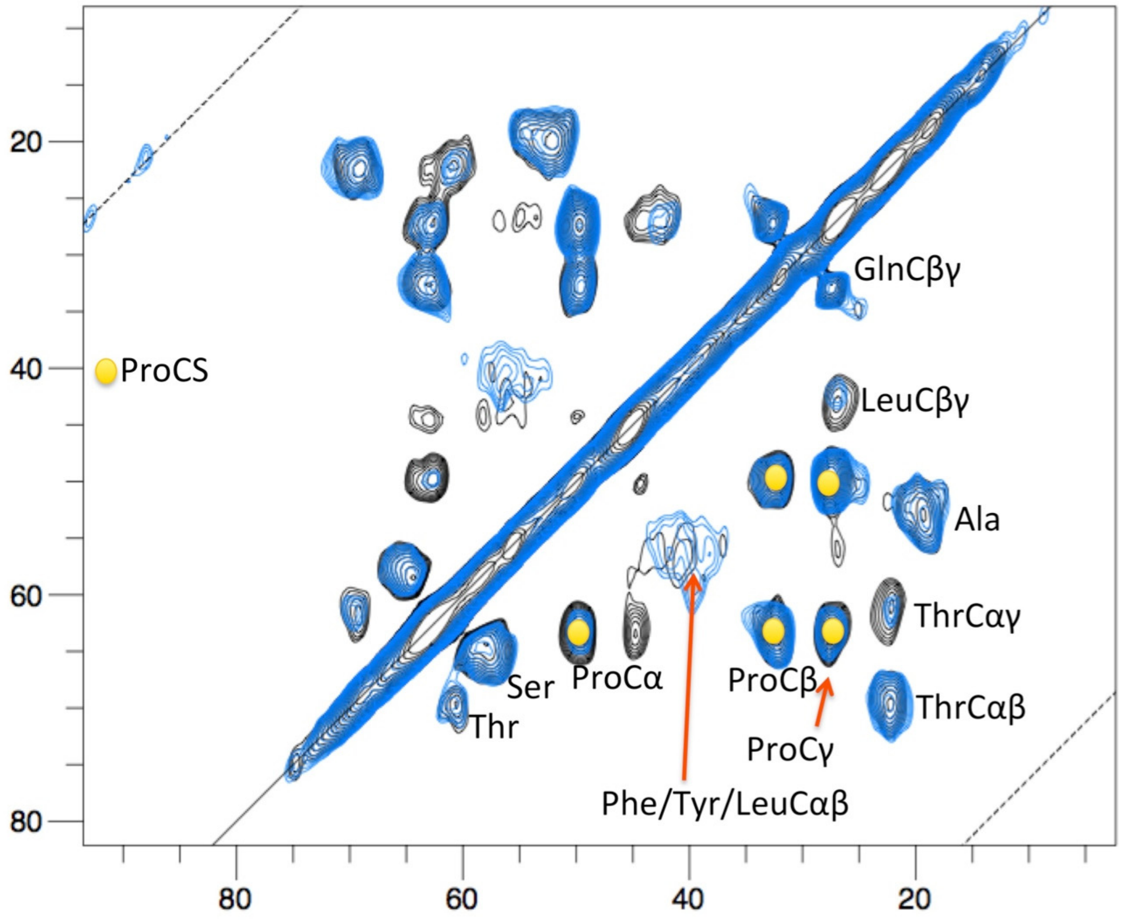


Figure 88 | W_3-38Cterm DARR close-up with a few assignments, blue = 10ms and black = 50ms. Diagonal dotted lines represent the spinning side bands. The yellow dots represent proline chemical shifts.

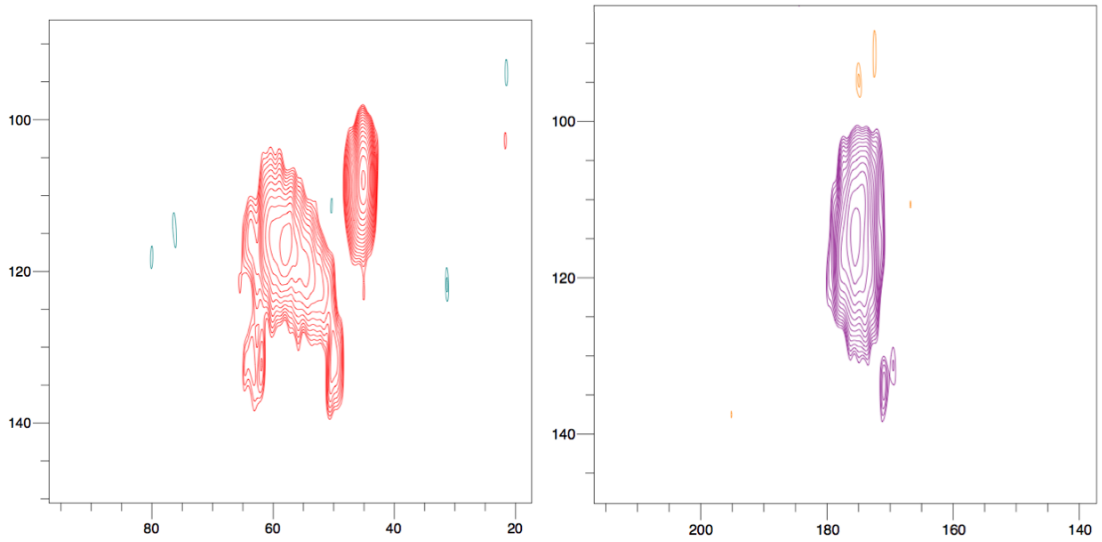


Figure 89 | W_3-38Cterm NCA (left) and NCO (right). No assignments were possible on these spectra.

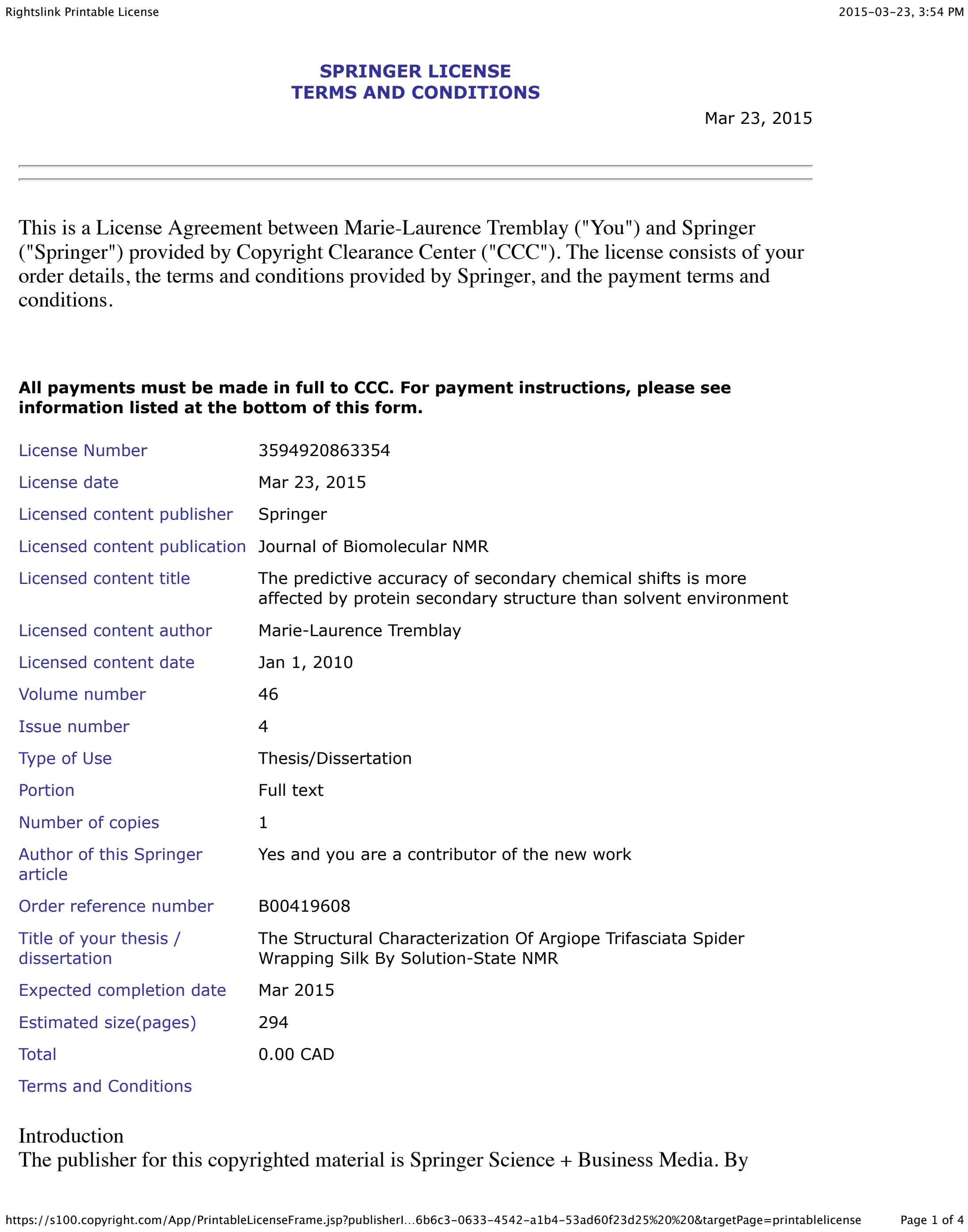
E.4. CONCLUSIONS

The data collected for this project were not assignable. The greatest problem appeared to be the fact that the protein was in powdered form. ssNMR requires some form of crystallization (or order) for clearer and narrower signals. Without some form of sample ordering, such as the β-sheet nanocrystals (Lewis 2006), it is very difficult to make any conclusive assignments from the peaks. The solution to this problem would be to generate fibres and perform ssNMR on that sample, which would likely be much more informative.

The next step would be to pull fibres from a labeled dope solution and pack them into a rotor. The problem in this endeavour was our very limited ability to form enough fibres to fill (or even partially fill) a rotor. We calculated that we would need approximately 10^6 fibres for a 1.7 mm rotor. Therefore, before this project can go further, we need to devise a better method for producing fibres in high quantity that will, preferably, eliminate human error during pulling.

Also, after acquiring data on the denaturation of W_1 and W_2 (see chapter 6), it was clear that perhaps the section that we labeled from residues 562-599 likely will not result in the formation of β-sheets within the fibres. Also, this part of the sequence contains 6/8 prolines, which are secondary structure breakers (Langelaan et al. 2010). My speculations is that the region from ~135-155 might have more potential for forming a β-sheet given that the helix from residue 135-150 is rather plastic and dynamic. With the labeling strategies that we have in place at the lab, maybe on day we shall shed light on the mystery of the location of those β-sheets in *A. trifasciata* aciniform silk.

APPENDIX F. COPYRIGHT AGREEMENT LETTERS



SPRINGER LICENSE
TERMS AND CONDITIONS

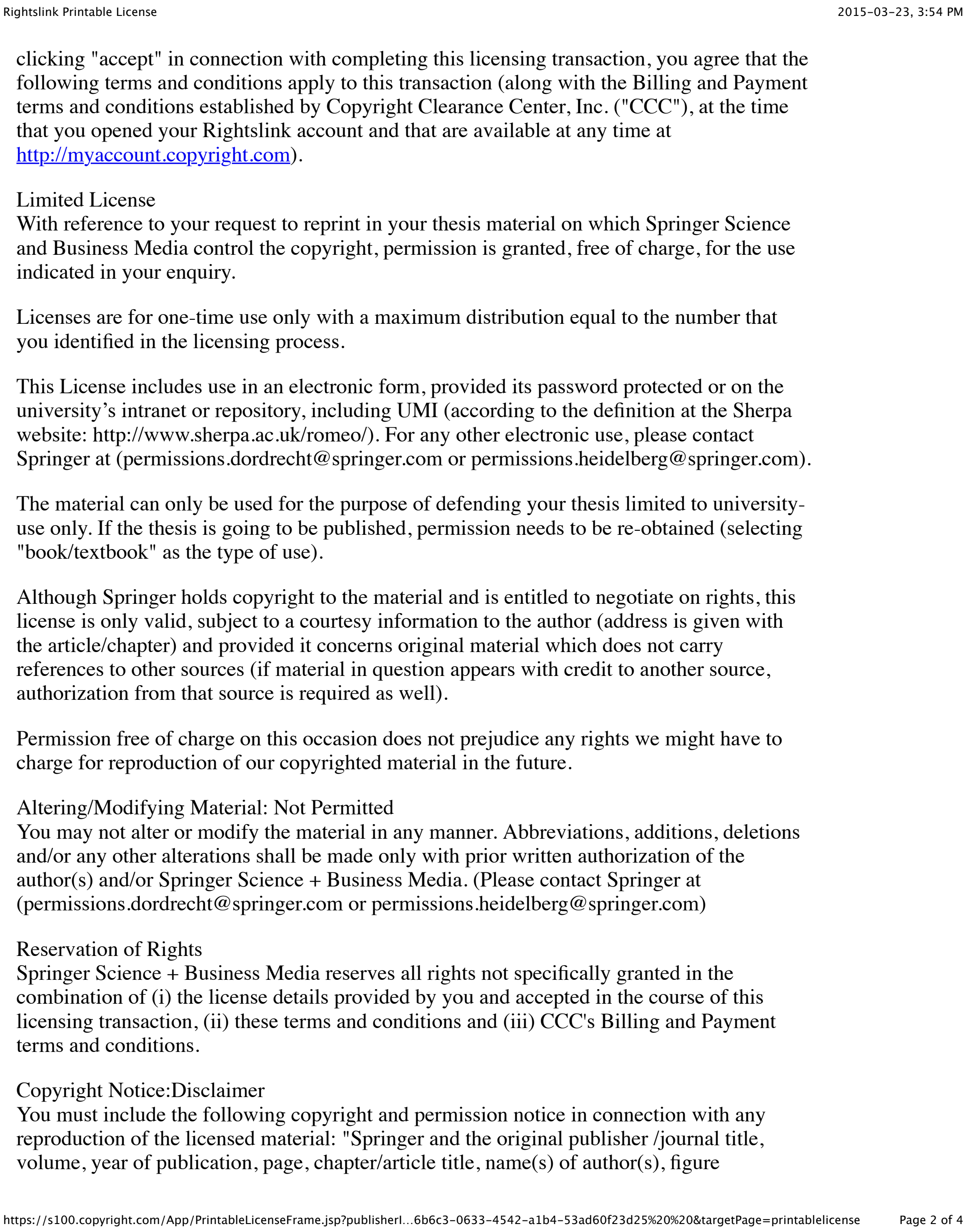
Mar 23, 2015

This is a License Agreement between Marie-Laurence Tremblay ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3594920863354
License date	Mar 23, 2015
Licensed content publisher	Springer
Licensed content publication	Journal of Biomolecular NMR
Licensed content title	The predictive accuracy of secondary chemical shifts is more affected by protein secondary structure than solvent environment
Licensed content author	Marie-Laurence Tremblay
Licensed content date	Jan 1, 2010
Volume number	46
Issue number	4
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	B00419608
Title of your thesis / dissertation	The Structural Characterization Of Argiope Trifasciata Spider Wrapping Silk By Solution-State NMR
Expected completion date	Mar 2015
Estimated size(pages)	294
Total	0.00 CAD
Terms and Conditions	

Introduction
The publisher for this copyrighted material is Springer Science + Business Media. By



clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License
With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: http://www.sherpa.ac.uk/romeo/). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis limited to university-use only. If the thesis is going to be published, permission needs to be re-obtained (selecting "book/textbook" as the type of use).

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted
You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights
Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:Disclaimer
You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure

number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

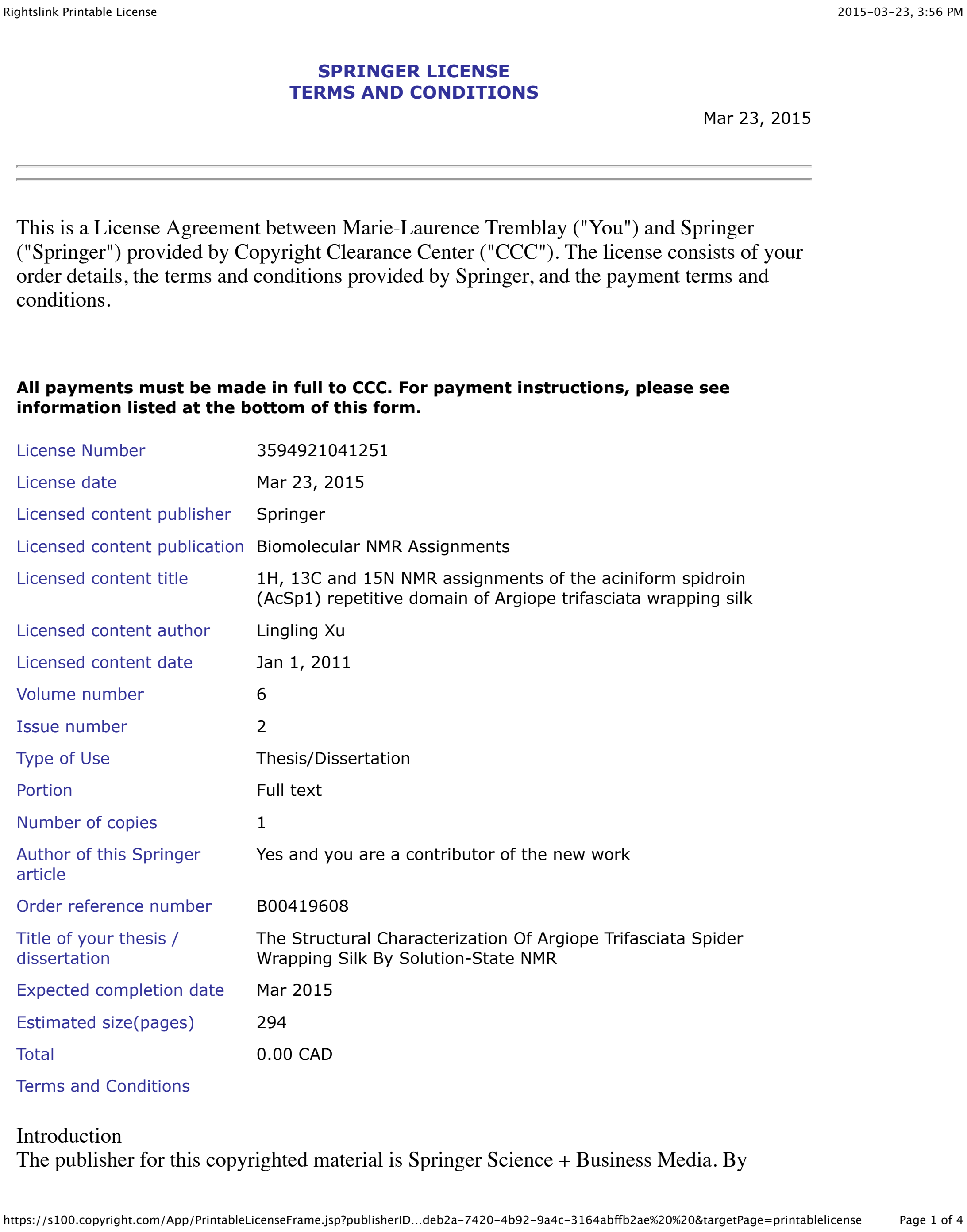
No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).***OR:***

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of



SPRINGER LICENSE
TERMS AND CONDITIONS

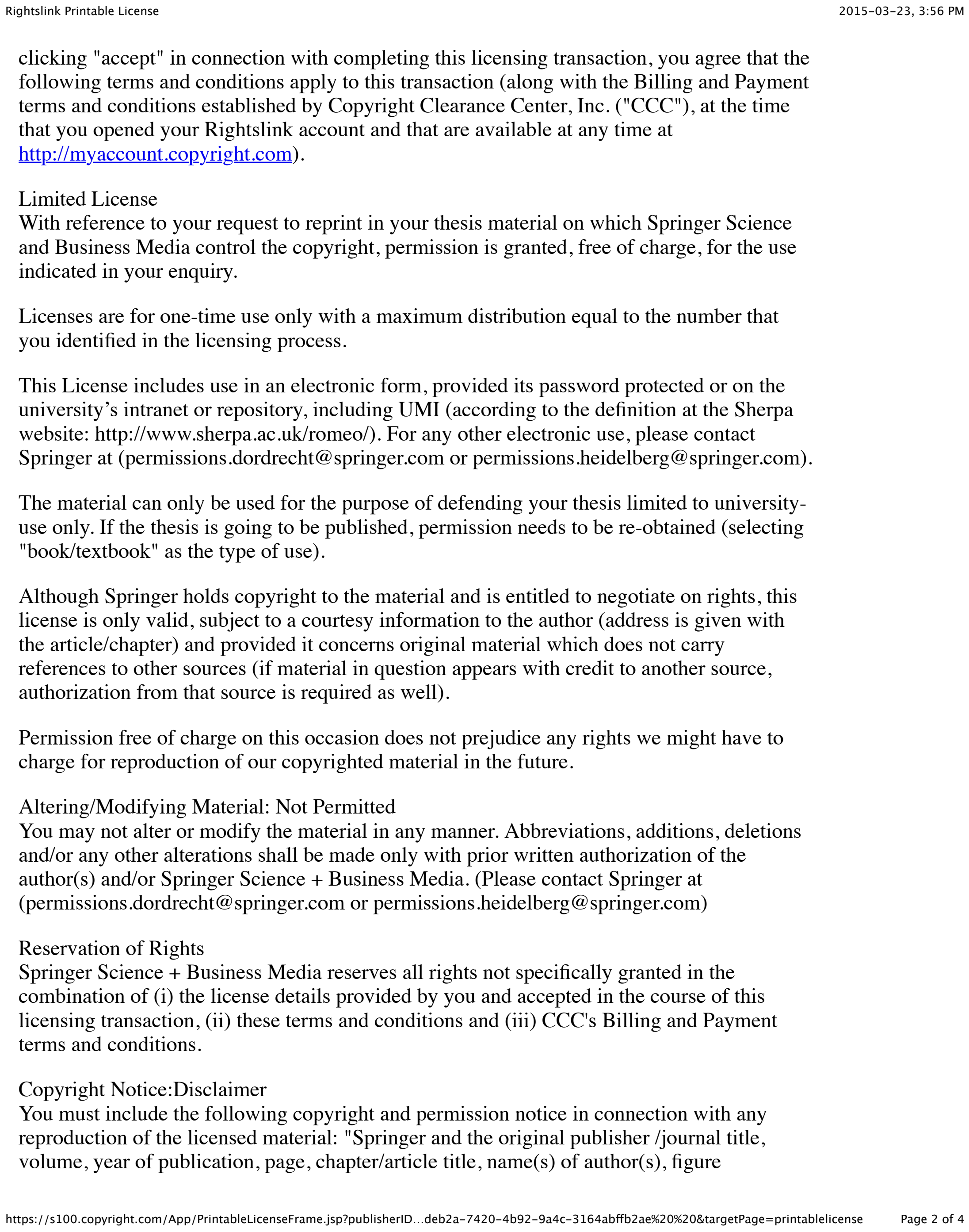
Mar 23, 2015

This is a License Agreement between Marie-Laurence Tremblay ("You") and Springer ("Springer") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Springer, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	3594921041251
License date	Mar 23, 2015
Licensed content publisher	Springer
Licensed content publication	Biomolecular NMR Assignments
Licensed content title	1H, 13C and 15N NMR assignments of the aciniform spidroin (AcSp1) repetitive domain of Argiope trifasciata wrapping silk
Licensed content author	Lingling Xu
Licensed content date	Jan 1, 2011
Volume number	6
Issue number	2
Type of Use	Thesis/Dissertation
Portion	Full text
Number of copies	1
Author of this Springer article	Yes and you are a contributor of the new work
Order reference number	B00419608
Title of your thesis / dissertation	The Structural Characterization Of Argiope Trifasciata Spider Wrapping Silk By Solution-State NMR
Expected completion date	Mar 2015
Estimated size(pages)	294
Total	0.00 CAD
Terms and Conditions	

Introduction
The publisher for this copyrighted material is Springer Science + Business Media. By



clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

Limited License

With reference to your request to reprint in your thesis material on which Springer Science and Business Media control the copyright, permission is granted, free of charge, for the use indicated in your enquiry.

Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process.

This License includes use in an electronic form, provided its password protected or on the university's intranet or repository, including UMI (according to the definition at the Sherpa website: http://www.sherpa.ac.uk/romeo/). For any other electronic use, please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com).

The material can only be used for the purpose of defending your thesis limited to university-use only. If the thesis is going to be published, permission needs to be re-obtained (selecting "book/textbook" as the type of use).

Although Springer holds copyright to the material and is entitled to negotiate on rights, this license is only valid, subject to a courtesy information to the author (address is given with the article/chapter) and provided it concerns original material which does not carry references to other sources (if material in question appears with credit to another source, authorization from that source is required as well).

Permission free of charge on this occasion does not prejudice any rights we might have to charge for reproduction of our copyrighted material in the future.

Altering/Modifying Material: Not Permitted

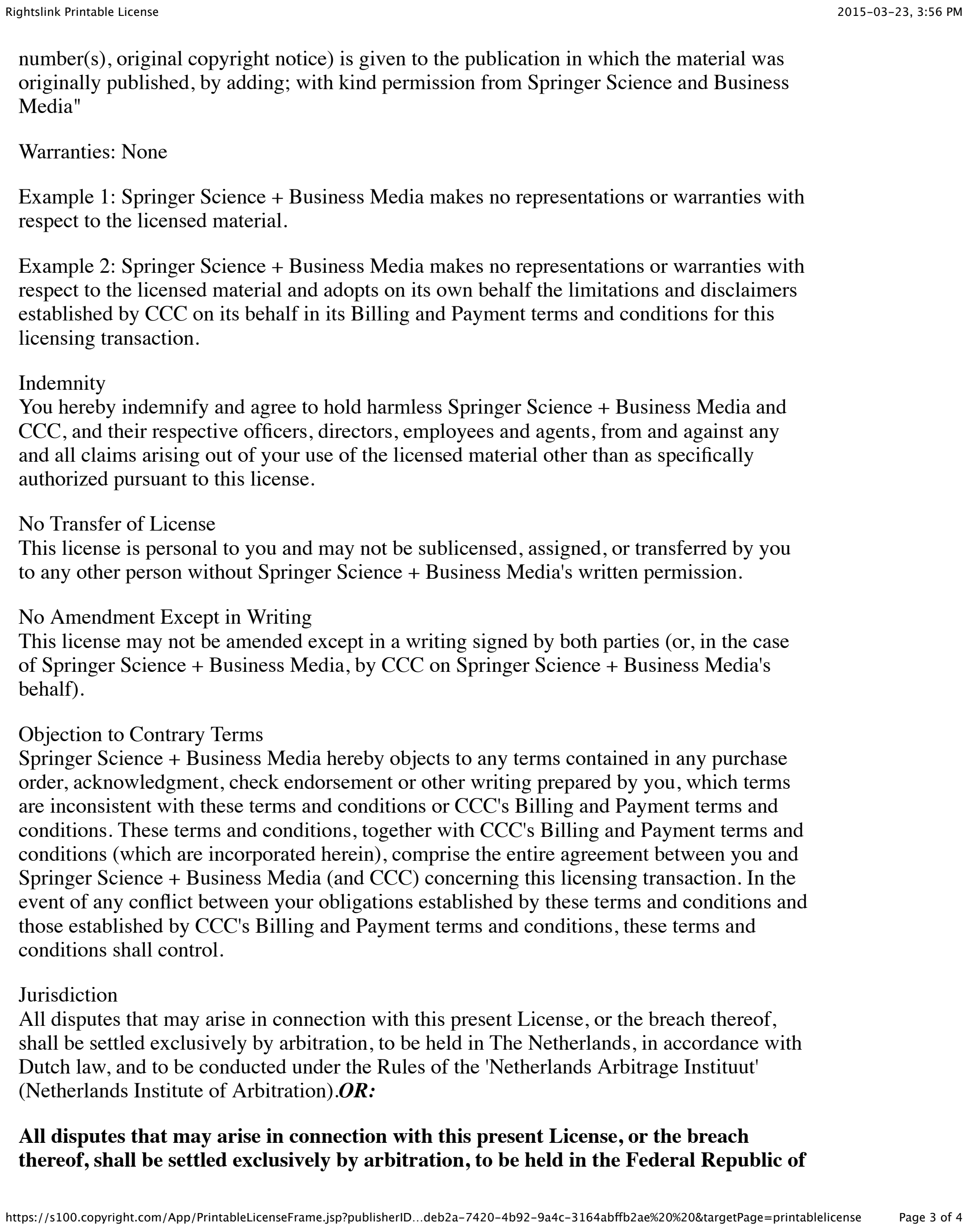
You may not alter or modify the material in any manner. Abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of the author(s) and/or Springer Science + Business Media. (Please contact Springer at (permissions.dordrecht@springer.com or permissions.heidelberg@springer.com)

Reservation of Rights

Springer Science + Business Media reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

Copyright Notice:Disclaimer

You must include the following copyright and permission notice in connection with any reproduction of the licensed material: "Springer and the original publisher /journal title, volume, year of publication, page, chapter/article title, name(s) of author(s), figure



number(s), original copyright notice) is given to the publication in which the material was originally published, by adding; with kind permission from Springer Science and Business Media"

Warranties: None

Example 1: Springer Science + Business Media makes no representations or warranties with respect to the licensed material.

Example 2: Springer Science + Business Media makes no representations or warranties with respect to the licensed material and adopts on its own behalf the limitations and disclaimers established by CCC on its behalf in its Billing and Payment terms and conditions for this licensing transaction.

Indemnity
You hereby indemnify and agree to hold harmless Springer Science + Business Media and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

No Transfer of License
This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without Springer Science + Business Media's written permission.

No Amendment Except in Writing
This license may not be amended except in a writing signed by both parties (or, in the case of Springer Science + Business Media, by CCC on Springer Science + Business Media's behalf).

Objection to Contrary Terms
Springer Science + Business Media hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and Springer Science + Business Media (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

Jurisdiction
All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in The Netherlands, in accordance with Dutch law, and to be conducted under the Rules of the 'Netherlands Arbitrage Instituut' (Netherlands Institute of Arbitration).***OR:***

All disputes that may arise in connection with this present License, or the breach thereof, shall be settled exclusively by arbitration, to be held in the Federal Republic of

BIBLIOGRAPHY

Abraham RJ, Cavalli L, Pachler KGR (1966) Rotational isomerism. Molecular Physics 11:471–494. doi: 10.1080/00268976600101281

Agnarsson I, Dhinojwala A, Sahni V, Blackledge TA (2009) Spider silk as a novel high performance biomimetic muscle driven by humidity. J Exp Biol 212:1990–1994. doi: 10.1242/jeb.028282

Altman GH, Diaz F, Jakuba C, et al. (2003) Silk-based biomaterials. Biomaterials 24:401–416. doi: 10.1016/S0142-9612(02)00353-8

Amirjahed AK, Blake MI (1975) Deviation of dielectric constant from ideality for certain binary solvent systems. J Pharm Sci 64:1569–1570.

Andersen SO (1970) Amino acid composition of spider silks. Comparative Biochemistry and Physiology 35:705–711. doi: 10.1016/0010-406X(70)90988-6

Anderson DE, Becktel WJ, Dahlquist FW (1990) pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. Biochemistry 29:2403–2408. doi: 10.1021/bi00461a025

Andersson M, Chen G, Otikovs M, et al. (2014) Carbonic anhydrase generates CO2 and H+ that drive spider silk formation via opposite effects on the terminal domains. PLoS Biol 12:e1001921. doi: 10.1371/journal.pbio.1001921

Appleby-Tagoe JH, Thiel IV, Wang Y, et al. (2011) Highly efficient and more general cis- and trans-splicing inteins through sequential directed evolution. J Biol Chem 286:34440–34447. doi: 10.1074/jbc.M111.277350

Arai M, Kuwajima K (2000) Role of the molten globule state in protein folding. Ad Prot Chem 53:209–282. doi: 10.1016/S0065-3233(00)53005-8

Archer SJ, Ikura M, Torchia DA, Bax A (1991) An alternative 3D NMR technique for correlating backbone ^{15}N with side chain Hβ resonances in larger proteins. J Mag Res 95:636–641. doi: 10.1016/0022-2364(91)90182-S

Asakura T, Suzuki Y, Nakazawa Y, et al. (2013a) Elucidating silk structure using solid-state NMR. Soft Matter 9:11440–11450. doi: 10.1039/c3sm52187g

Asakura T, Suzuki Y, Nakazawa Y, et al. (2013b) Silk structure studied with nuclear magnetic resonance. Prog Nucl Mag Res Sp 69:23–68. doi: 10.1016/j.pnmrs.2012.08.001

Askarieh G, Hedhammar M, Nordling K, et al. (2010) Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. Nature 465:236–238. doi: 10.1038/nature08962

Atreya HS, Chary KVR (2001) Selective `unlabeling' of amino acids in fractionally 13C labeled proteins: An approach for stereospecific NMR assignments of CH3 groups in Val and Leu residues. J Biomol NMR 19:267–272. doi: 10.1023/A:1011262916235

Avbelj F, Baldwin RL (2004) Origin of the neighboring residue effect on peptide backbone conformation. Proc Natl Acad Sci USA 101:10967. doi: pnas.0404050101

Ayoub NA, Garb JE, Kuelbs A, Hayashi CY (2013) Ancient properties of spider silks revealed by the complete gene sequence of the prey-wrapping silk protein (AcSp1). Mol Biol Evol 30:589–601. doi: 10.1093/molbev/mss254

Ayoub NA, Garb JE, Tinghitella RM, et al. (2007) Blueprint for a high-performance biomaterial: full-length spider dragline silk Genes. PLoS ONE 2:e514. doi: 10.1371/journal.pone.0000514

Baldus M, Petkova At, Herzfeld J, Griffin RG (1998) Cross polarization in the tilted frame: assignment and spectral simplification in heteronuclear spin systems. Molecular Physics 95:1197–1207. doi: 10.1080/00268979809483251

Baldwin RL (2007) Energetics of Protein Folding. J Mol Biol 371:283–301. doi: 10.1016/j.jmb.2007.05.078

Baldwin RL (2005) Weak Interactions in Protein Folding: Hydrophobic Free Energy, van der Waals Interactions, Peptide Hydrogen Bonds, and Peptide Solvation. In: Buchner J, Kiefhaber T (eds) Protein Folding Handbook. Wiley-VCH Verlag GmbH, Weinheim, Germany, pp 127–162

Barbato G, Ikura M, Kay LE, et al. (1992) Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. Biochemistry 31:5269–5278. doi: 10.1021/bi00135a017

Barghout J, Thiel BL, Viney C (1999) Spider (Araneus diadematus) cocoon silk: a case of non-periodic lattice crystals with a twist? Int J Biol Macromol 24:211–217. doi: 10.1016/S0141-8130(99)00007-0

Barnwal RP, Rout AK, Chary KVR, Atreya HS (2007) Rapid measurement of ^{3}J (H^{N}-H^{a}) and ^{3}J (N-H^{b}) coupling constants in polypeptides. J Biomol NMR 39:259–263. doi: 10.1007/s10858-007-9200-8

Barón M (2001) Definitions of basic terms relating to low-molar-mass and polymer liquid crystals. Pure Appl Chem 73:845–895.

Barrientos LG, Dolan C, Gronenborn AM (2000) Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings. J Biomol NMR 16:329–337.

Bassolino-Klimas D, Tejero R, Krystek SR, et al. (1996) Simulated annealing with restrained molecular dynamics using a flexible restraint potential: theory and evaluation with simulated NMR constraints. Protein Sci 5:593–603. doi: 10.1002/pro.5560050404

Bax A, Clore GM, Gronenborn AM (1990) ^{1}H- ^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. J Mag Res 88:425–431. doi: 10.1016/0022-2364(90)90202-K

Bax A, Ikura M (1991) An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α-carbon of the preceding residue in uniformly ^{15}N/^{13}C enriched proteins. J Biomol NMR 1:99–104. doi: 10.1007/BF01874573

Bax A, Tjandra N (1997) High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium. J Biomol NMR 10:289–292. doi: 10.1023/A:1018308717741

Beger RD, Bolton PH (1997) Protein phi and psi dihedral restraints determined from multidimensional hypersurface correlations of backbone chemical shifts and their use in the determination of protein tertiary structures. J Biomol NMR 10:129–142.

Bell FI, McEwen IJ, Viney C (2002) Fibre science: supercontraction stress in wet spider dragline. Nature 416:37. doi: 10.1038/416037a

Berjanskii MV, Neal S, Wishart DS (2006) PREDITOR: a web server for predicting protein torsion angle restraints. Nucleic Acids Res 34:W63–9. doi: 10.1093/nar/gkl341

Berjanskii MV, Wishart DS (2007) The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. Nucleic Acids Res 35:W531–7. doi: 10.1093/nar/gkm328

Berman HM, Westbrook J, Feng Z, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242. doi: 10.1093/nar/28.1.235

Bermel W, Tkach EN, Sobol AG, Golovanov AP (2009) Simultaneous measurement of residual dipolar couplings for proteins in complex using the isotopically discriminated NMR approach. J Am Chem Soc 131:8564–8570. doi: 10.1021/ja901602c

Bernadó P, García de la Torre J, Pons M (2002) Interpretation of ^{15}N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. J Biomol NMR 23:139–150. doi: 10.1023/A:1016359412284

Bini E, Knight DP, Kaplan DL (2004) Mapping domain structures in silks from insects and spiders related to protein assembly. J Mol Biol 335:27–40. doi: doi:10.1016/j.jmb.2003.10.043

Blaber M, Zhang XJ, Matthews BW (1993) Structural basis of amino acid alpha helix propensity. Science 260:1637–1640. doi: 10.1126/science.8503008

Blasingame E, Tuton-Blasingame T, Larkin L, et al. (2009) Pyriform spidroin 1, a novel member of the silk gene family that anchors dragline silk fibers in attachment discs of the black widow spider, Latrodectus hesperus. J Biol Chem 284:29097–29108. doi: 10.1074/jbc.M109.021378

Bordbar A, Saboury A, Housaindokht M, Moosavi-Movahedi A (1997) Statistical Effects of the Binding of Ionic Surfactant to Protein. J Colloid Interface Sci 192:415–419.

Breslauer DN, Muller SJ, Lee LP (2010) Generation of monodisperse silk microspheres prepared with microfluidics. Biomacromolecules 11:643–647. doi: 10.1021/bm901209u

Brockman H (1994) Dipole potential of lipid membranes. Chemistry and Physics of Lipids 73:57–79. doi: 10.1016/0009-3084(94)90174-0

Brooks BR, Bruccoleri RE, Olafson BD, et al. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4:187–217. doi: 10.1002/jcc.540040211

Brown CP, Rosei F, Traversa E, Licoccia S (2011) Spider silk as a load bearing biomaterial: tailoring mechanical properties via structural modifications. Nanoscale 3:870. doi: 10.1039/c0nr00752h

Brunger AT, Adams PD, Clore GM, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D 54:905–921.

Brünger AT, Kuriyan J, Karplus M (1987) Crystallographic R factor refinement by molecular dynamics. Science 235:458–460. doi: 10.1126/science.235.4787.458

Buchinger E, Aachmann FL, Aranko AS, et al. (2010) Use of protein trans-splicing to produce active and segmentally (2)H, (15)N labeled mannuronan C5-epimerase AlgE4. Protein Sci 19:1534–1543. doi: 10.1002/pro.432

Bundi A, Grathwohl C, Hochmann J, Keller RM (1975) Proton NMR of the protected tetrapeptides TFA-Gly-Gly-X-Ala-OCH_3, where X stands for One of the 20 common amino acids. J Mag Res. doi: 10.1016/0022-2364(75)90237-1

Burger R, Bigler P (1998) DEPTQ: distorsionless enhancement by polarization transfer including the detection of quaternary nuclei. J Mag Res 135:529–534. doi: 10.1006/jmre.1998.1595

Busche AEL, Aranko AS, Talebzadeh Farooji M, et al. (2009) Segmental isotopic labeling of a central domain in a multidomain protein by protein *trans*-splicing using only one robust DnaE intein. Angew Chem Int Ed Engl 48:6128–6131. doi: 10.1002/anie.200901488

Buskirk AR, Ong Y-C, Gartner ZJ, Liu DR (2004) Directed evolution of ligand dependence: small-molecule-activated protein splicing. Proc Natl Acad Sci USA 101:10505–10510. doi: 10.1073/pnas.0402762101

Cai YQ, Mao HK, Chow PC, et al. (2005) Ordering of hydrogen bonds in high-pressure low-temperature H 2 O. Phys Rev Lett 94:025502 1–4. doi: 10.1103/PhysRevLett.94.025502

Candelas GC, Cintron J (1981) A spider fibroin and its synthesis. J Exp Zool 216:1–6. doi: 10.1002/jez.1402160102

Carr HY, Purcell EM (1954) Effects of diffusion on free precession in nuclear magnetic resonance experiments. Phys Rev Lett 94:630–638. doi: 10.1103/PhysRev.94.630

Case DA (2000) Interpretation of chemical shifts and coupling constants in macromolecules. Curr Opin Struct Biol 10:197–203.

Case DA (1998) The use of chemical shifts and their anisotropies in biomolecular structure determination. Curr Opin Struct Biol 8:624–630.

Cavagnero S, Dyson HJ, Wright PE (1999) Improved low pH bicelle system for orienting macromolecules over a wide temperature range. J Biomol NMR 13:387–391.

Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. Proc Natl Acad Sci USA 104:9615–9620. doi: 10.1073/pnas.0610313104

Challis RJ, Goodacre SL, Hewitt GM (2006) Evolution of spider silks: conservation and diversification of the C-terminus. Insect Mol Biol 15:45–56. doi: 10.1111/j.1365-2583.2005.00606.x

Chaw RC, Zhao Y, Wei J, et al. (2014) Intragenic homogenization and multiple copies of prey-wrapping silk genes in Argiope garden spiders. BMC Evol Biol 14:1–12. doi: 10.1186/1471-2148-14-31

Chellgren BW, Creamer TP (2004) Short Sequences of Non-Proline Residues Can Adopt the Polyproline II Helical Conformation †. Biochemistry 43:5864–5869. doi: 10.1021/bi049922v

Chen J, Lu Z, Sakon J, Stites WE (2000) Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. J Mol Biol 303:125–130. doi: 10.1006/jmbi.2000.4140

Chen, K. & Tjandra, N. The use of residual dipolar coupling in studying proteins by NMR. Top Curr Chem 326, 47–68 (2012). doi: 10.1007/978-3-642-28917-0

Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. J Mag Res 202:223–233. doi: 10.1016/j.jmr.2009.11.008

Clore GM, Driscoll PC, Wingfield PT (1990a) Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear nitrogen-15-proton NMR spectroscopy. Biochemistry 29:7389–7401. doi: 10.1021/bi00484a006

Clore GM, Gronenborn AM, Bax A (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. J Mag Res 133:216–221. doi: 10.1006/jmre.1998.1419

Clore GM, Szabo A, Bax A, Kay LE (1990b) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. ChemBioChem 112:4989–4991. doi: 10.1021/ja00168a070

Cloutier I, Leclerc J, Lefèvre T, Auger M (2011) Solid-state nuclear magnetic resonance (NMR) spectroscopy reveals distinctive protein dynamics in closely related spider silks. Can J Chem 89:1047–1054. doi: 10.1139/v11-036

Coddington JA (1989) Spinneret silk spigot morphology: evidence for the monophyly of orbweaving spiders, *Cyrtophorinae* (*Araneidae*), and the group *Theridiidae* plus *Nesticidae*. J Arachnology 17:71–95.

Colgin MA, Lewis RV (1998) Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like “spacer regions.” Protein Sci 7:667–672. doi: 10.1002/pro.5560070315

Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 13:289–302. doi: 10.1023/A:1008392405740

Craig CL, Riekel C (2002) Comparative architecture of silks, fibrous proteins and their encoding genes in insects and spiders. Comp Biochem Physiol B 133:493–507.

Creager MS, Jenkins JE, Thagard-Yeaman LA, et al. (2010) Solid-state NMR comparison of various spiders' dragline silk fiber. Biomacromolecules 11:2039–2043. doi: 10.1021/bm100399x

Cui L, Liu F, Ou-Yang Z (2009) The study of the elasticity of spider dragline silk with liquid crystal model. Thin Solid Films. doi: 10.1016/j.tsf.2009.07.080

d'Auvergne EJ, Gooley PR (2008) Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. J Biomol NMR 40:107–119. doi: 10.1007/s10858-007-9214-2

d'Auvergne EJ, Gooley PR (2003) The use of model selection in the model-free analysis of protein dynamics. J Biomol NMR 25:25–39. doi: 10.1023/A:1021902006114

Dalgarno DC, Levine BA, Williams RJ (1983) Structural information from NMR secondary chemical shifts of peptide alpha C-H protons in proteins. Biosci Rep 3:443–452.

Das A, Mukhopadhyay C (2009) Urea-Mediated Protein Denaturation: A Consensus View. J Phys Chem B 113:12816–12824. doi: 10.1021/jp906350s

Davis IW, Leaver-Fay A, Chen VB, et al. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375–83. doi: 10.1093/nar/gkm216

De Luca G, Rey AD (2006) Dynamic interactions between nematic point defects in the spinning extrusion duct of spiders. J Chem Phys 124:144904. doi: 10.1063/1.2186640

De Rosa L, Russomanno A, Romanelli A, D'Andrea LD (2013) Semi-synthesis of labeled proteins for spectroscopic applications. Molecules 18:440–465. doi: 10.3390/molecules18010440

Delaglio F, Grzesiek S, Vuister GW, et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293. doi: 10.1007/BF00197809

Demarco A, LlináS M, Wüthrich K (1978) ^{1}H-^{15}N Spin–spin couplings in alumichrome. Biopolymers 17:2727–2742. doi: 10.1002/bip.1978.360171118

Dempsey CE, Piggot TJ, Mason PE (2005) Dissecting contributions to the denaturant sensitivities of proteins. Biochemistry 44:775–781. doi: 10.1021/bi048389g

Dicko C, Kenney JM, Knight D, Vollrath F (2004a) Transition to a beta-sheet-rich structure in spidroin in vitro: the effects of pH and cations. Biochemistry 43:14080–14087. doi: 10.1021/bi0483413

Dicko C, Knight D, Kenney JM, Vollrath F (2004b) Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. Temperature and concentration effects. Biomacromolecules 5:2105–2115. doi: 10.1021/bm034486y

Diercks T, Coles M, Kessler H (1999) An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. J Biomol NMR 15:177–180. doi: 10.1023/A:1008367912535

Dill KA (1990) Dominant forces in protein folding. Biochemistry 29:7133–7155.

Dill KA, Shortle D (1991) Denatured states of proteins. Annu Rev Biochem 60:795–825.

Ding J, Rainey JK, Xu C, et al. (2006) Structural and functional characterization of transmembrane segment VII of the Na^+/H^+ exchanger isoform 1. J Biol Chem 281:29817–29829. doi: 10.1074/jbc.M606152200

Donald AM, Windle AH, Hanna S (2006) Liquid Crystalline Polymers. Cambridge University Press

Dosset P, Hus J-C, Marion D, Blackledge M (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. J Biomol NMR 20:223–231. doi: 10.1023/A:1011206132740

Duggan BM, Legge GB, Dyson HJ, Wright PE (2001) SANE (Structure Assisted NOE Evaluation): An automated model-based approach for NOE assignment. J Biomol NMR 19:321–329. doi: 10.1023/A:1011227824104

Düx P, Whitehead B, Boelens R, et al. (1997) Measurement of ^{15}N-^{1}H coupling constants in uniformly ^{15}N-labeled proteins: Application to the photoactive yellow protein. J Biomol NMR 10:301–306. doi: 10.1023/A:1018393225804

Eisoldt L, Thamm C, Scheibel T (2012) Review the role of terminal domains during storage and assembly of spider silk proteins. Biopolymers 97:355–361. doi: 10.1002/bip.22006

Eles PT, Michal CA (2004a) A DECODER NMR study of backbone orientation in *Nephila clavipes* dragline silk under varying strain and draw rate. Biomacromolecules 5:661–665. doi: 10.1021/bm0342685

Eles PT, Michal CA (2004b) Strain Dependent Local Phase Transitions Observed during Controlled Supercontraction Reveal Mechanisms in Spider Silk. Macromolecules 37:1342–1345. doi: 10.1021/ma035567p

Eliezer D, Yao J, Dyson HJ, Wright PE (1998) Structural and dynamic characterization of partially folded states of apomyoglobin and implications for protein folding. Nat Struct Mol Biol 5:148–155. doi: 10.1038/nsb0298-148

Enthart A, Freudenberger JC, Furrer J, et al. (2008) The CLIP/CLAP-HSQC: pure absorptive spectra for the measurement of one-bond couplings. J Mag Res 192:314–322. doi: 10.1016/j.jmr.2008.03.009

Evans JNS (1995) Biomolecular Nmr Spectroscopy. Oxford University Press

Evans TC, Martin D, Kolly R, et al. (2000) Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of Synechocystis species PCC6803. J Biol Chem 275:9091–9094. doi: 10.1074/jbc.275.13.9091

Exler JH, Hümmerich D, Scheibel T (2007) The amphiphilic properties of spider silks are important for spinning. Angew Chem Int Ed 46:3559–3562. doi: 10.1002/anie.200604718

Farrow NA, Zhang O, Forman-Kay JD, Kay LE (1997) Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. Biochemistry 36:2390–2402. doi: 10.1021/bi962548h

Farrow NA, Zhang O, Forman-Kay JD, Kay LE (1994) A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chemical exchange rates of systems in slow equilibrium. J Biomol NMR 4:727–734. doi: 10.1007/BF00404280

Farrow NA, Zhang O, Szabo A, et al. (1995) Spectral density function mapping using 15N relaxation data exclusively. J Biomol NMR 6:153–162. doi: 10.1007/BF00211779

Fitch CA, Whitten ST, Hilser VJ, García-Moreno E B (2006) Molecular mechanisms of pH-driven conformational transitions of proteins: insights from continuum electrostatics calculations of acid unfolding. Proteins 63:113–126. doi: 10.1002/prot.20797

Florczak, A., Mackiewicz, A. & Dams-Kozlowska, H. (2014) Functionalized Spider Silk Spheres As Drug Carriers for Targeted Cancer Therapy. *Biomacromolecules* 15, 2971–2981. doi: 10.1021/bm500591p

Foo CWP, Bini E, Hensman J, et al. (2006) Role of pH and charge on silk protein assembly in insects and spiders. Appl Phys A 82:223–233. doi: 10.1007/s00339-005-3426-7

Fossi M, Oschkinat H, Nilges M, Ball LJ (2005) Quantitative study of the effects of chemical shift tolerances and rates of SA cooling on structure calculation from automatically assigned NOE data. J Mag Res 175:92–102. doi: 10.1016/j.jmr.2005.03.020

Friend SH, Gurd FRN (1979) Electrostatic stabilization in myoglobin. The pH dependence of summed electrostatic contributions. Biochemistry 18:4612–4619. doi: 10.1021/bi00588a023

Fuxreiter M, Tompa P (2012) Fuzziness. doi: 10.1007/978-1-4614-0659-4

Gaines WA, Sehorn MG, Marcotte WR (2010) Spidroin N-terminal domain promotes a pH-dependent association of silk proteins during self-assembly. J Biol Chem 285:40745–40753. doi: 10.1074/jbc.M110.163121

Gao Z, Lin Z, Huang W, et al. (2013) Structural characterization of minor ampullate spidroin domains and their distinct roles in fibroin solubility and fiber formation. PLoS ONE 8:e56142. doi: 10.1371/journal.pone.0056142

Garb JE, Hayashi CY (2005) Modular evolution of egg case silk genes across orb-weaving spider superfamilies. Proc Natl Acad Sci USA 102:11379–11384.

García de la Torre J, Huertas ML, Carrasco B (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. J Mag Res 147:138–146. doi: 10.1006/jmre.2000.2170

García-Moreno B, Dwyer JJ, Gittis AG, et al. (1997) Experimental measurement of the effective dielectric in the hydrophobic core of a protein. Biophys Chem 64:211–224. doi: 10.1016/S0301-4622(96)02238-7

Gardner KH, Konrat R, Rosen MK, Kay LE (1996) An (H)C(CO)NH-TOCSY pulse scheme for sequential assignment of protonated methyl groups in otherwise deuterated (15)N, (13)C-labeled proteins. J Biomol NMR 8:351–356. doi: 10.1007/BF00410333

Gebel EB, Shortle D (2007) Characterization of denatured proteins using residual dipolar couplings. Methods Mol Biol 350:39–48.

Geisler M, Pirzer T, Ackerschott C, et al. (2008) Hydrophobic and Hofmeister effects on the adhesion of spider silk proteins onto solid substrates: An AFM-based single-molecule study. Langmuir 24:1350–1355. doi: 10.1021/la702341j

Geurts P, Zhao L, Hsia Y, et al. (2010) Synthetic spider silk fibers spun from Pyriform Spidroin 2, a glue silk protein discovered in orb-weaving spider attachment discs. Biomacromolecules 11:3495–3503. doi: 10.1021/bm101002w

Giriat I, Muir TW (2003) Protein semi-synthesis in living cells. J Am Chem Soc 125:7180–7181. doi: 10.1021/ja034736i

Glišovic A, Vehoff T, Davies RJ, Salditt T (2008) Strain dependent structural changes of spider dragline silk. Macromolecules 41:390–398. doi: 10.1021/ma070528p

Gnesa E, Hsia Y, Yarger JL, et al. (2012) Conserved C-terminal domain of spider tubuliform spidroin 1 contributes to extensibility in synthetic fibers. Biomacromolecules 13:304–312. doi: 10.1021/bm201262n

Golovanov AP, Blankley RT, Avis JM, Bermel W (2007) Isotopically discriminated NMR spectroscopy: A tool for investigating complex protein interactions in vitro. J Am Chem Soc 129:6528–6535. doi: 10.1021/ja070505q

Gosline JM, DeMont ME, Denny MW (1986) The structure and properties of spider silk. Endeavour 10:37–43. doi: 10.1016/0160-9327(86)90049-9

Gosline JM, Denny MW, DeMont ME (1984) Spider silk as rubber. Nature 551–552. doi: 10.1038/309551a0

Gosline JM, Guerette PA, Ortlepp CS, Savage KN (1999) The mechanical design of spider silks: from fibroin sequence to mechanical function. J Exp Biol 202:3295–3303.

Goto NK, Kay LE (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. Curr Opin Struct Biol 10:585–592. doi: 10.1016/S0959-440X(00)00135-4

Granville V, Krivánek M, Rasson JP (1994) Simulated annealing: a proof of convergence. IEEE Trans Pattern Anal Mach Intell 16:652–656. doi: 10.1109/34.295910

Greenfield N, Fasman GD (1969) Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry 8:4108–4116.

Greenfield NJ, Huang YJ, Palm T, et al. (2001) Solution NMR structure and folding dynamics of the N terminus of a rat non-muscle α-tropomyosin in an engineered chimeric protein. J Mol Biol 312:833–847. doi: 10.1006/jmbi.2001.4982

Grigera JR, McCarthy AN (2010) The behavior of the hydrophobic effect under pressure and protein denaturation. Biophys J 98:1626–1631. doi: 10.1016/j.bpj.2009.12.4298

Gronwald W, Moussa S, Elsner R, et al. (2002) Automated assignment of NOESY NMR spectra using a knowledge based method (KNOWNOE). J Biomol NMR 23:271–287. doi: 10.1023/A:1020279503261

Guerry P, Herrmann T (2011) Advances in automated NMR protein structure determination. Q Rev Biophys 44:257–309. doi: 10.1017/S0033583510000326

Gust D, Moon RB, Roberts JD (1975) Applications of natural-abundance nitrogen-15 nuclear magnetic resonance to large biochemically important molecules. Proc Natl Acad Sci USA 72:4696–4700.

Guy CA, Fields GB (1997) Trifluoroacetic acid cleavage and deprotection of resin-bound peptides following synthesis by Fmoc chemistry. Meth Enzymol 289:67–83. doi: 10.1016/S0076-6879(97)89044-1

Güntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353–378. doi: 10.1385/1-59259-809-9:353

Güntert P, Berndt KD, Wüthrich K (1993) The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination. J Biomol NMR 3:601–606. doi: 10.1007/BF00174613

Hagn F (2012) A structural view on spider silk proteins and their role in fiber assembly. J Peptide Sci 18:357–365. doi: 10.1002/psc.2417

Hagn F, Eisoldt L, Hardy JG, et al. (2010a) A conserved spider silk domain acts as a molecular switch that controls fibre assembly. Nature 465:239–242. doi: 10.1038/nature08936

Hagn F, Thamm C, Scheibel T, Kessler H (2010b) pH-dependent dimerization and salt-dependent stabilization of the N-terminal domain of spider dragline silk-implications for fiber formation. Angew Chem Int Ed 50:310–313. doi: 10.1002/anie.201003795

Hajer J, Malý J, Hrubá L, Reháková D (2009) Egg sac silk of *Theridiosoma gemmosum* (*Araneae: Theridiosomatidae*). J Morphol 270:1269–1283. doi: 10.1002/jmor.10757

Hansen MR, Hanson P, Pardi A (2000) Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. Meth Enzymol 317:220–240.

Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Mol Biol 5:1065–1074. doi: 10.1038/4176

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

Hass MAS, Jensen MRO, Led JJ (2008) Probing electric fields in proteins in solution by NMR spectroscopy. Proteins 72:333–343. doi: 10.1002/prot.21929

Hayakawa I, Linko YY, Linko P (1996) Mechanism of high pressure denaturation of proteins. LWT-Food Science and Technology 29:756–762. doi: 10.1006/fstl.1996.0118

Hayashi CY, Blackledge TA, Lewis RV (2004) Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. Mol Biol Evol 21:1950–1959. doi: 10.1093/molbev/msh204

Hayashi CY, Lewis RV (1998) Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. J Mol Biol 275:773–784. doi: 10.1006/jmbi.1997.1478

Hayashi CY, Shipley NH, Lewis RV (1999) Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. Int J Biol Macromol 24:271–275.

Hedhammar M, Rising A, Grip S, et al. (2008) Structural properties of recombinant nonrepetitive and repetitive parts of major ampullate spidroin 1 from *Euprosthenops australis*: implications for fiber formation. Biochemistry 47:3407–3417. doi: 10.1021/bi702432y

Heidebrecht A, Scheibel T (2013) Recombinant production of spider silk proteins. Adv Appl Microbiol 82:115–153. doi: 10.1016/B978-0-12-407679-2.00004-1

Heim M, Keerl D, Scheibel T (2009) Spider silk: from soluble protein to extraordinary fiber. Angew Chem Int Ed 48:3584–3596. doi: 10.1002/anie.200803341

Heremans K, Smeller L (1998) Protein structure and dynamics at high pressure. Biochim Biophys Acta. doi: 10.1016/S0167-4838(98)00102-2

Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227. doi: 10.1016/s0022-2836(02)00241-3

Higman VA, Boyd J, Smith LJ, Redfield C (2011) Residual dipolar couplings: are multiple independent alignments always possible? J Biomol NMR 49:53–60. doi: 10.1007/s10858-010-9457-1

Hijirida DH, Do KG, Michal C, et al. (1996) 13C NMR of Nephila clavipes major ampullate silk gland. Biophys J 71:3442–3447. doi: 10.1016/S0006-3495(96)79539-5

Hofer M, Winter G, Myschik J (2012) Recombinant spider silk particles for controlled delivery of protein drugs. Biomaterials 33:1554–1562. doi: 10.1016/j.biomaterials.2011.10.053

Hoffman RE (2006) Standardization of chemical shifts of TMS and solvent signals in NMR solvents. Magn Reson Chem 44:606–616. doi: 10.1002/mrc.1801

Hoffman RE (2003) Variations on the chemical shift of TMS. J Mag Res 163:325–331. doi: 10.1016/S1090-7807(03)00142-3

Holland GP, Creager MS, Jenkins JE, et al. (2008a) Determining secondary structure in spider dragline silk by carbon−carbon correlation solid-state NMR spectroscopy. J Am Chem Soc 130:9871–9877. doi: 10.1021/ja8021208

Holland GP, Jenkins JE, Creager MS, et al. (2008b) Quantifying the fraction of glycine and alanine in beta-sheet and helical conformations in spider dragline silk using solid-state NMR. Chem Commun 5568–5570. doi: 10.1039/b812928b

Holland GP, Jenkins JE, Creager MS, et al. (2008c) Solid-state NMR investigation of major and minor ampullate spider silk in the native and hydrated states. Biomacromolecules 9:651–657. doi: 10.1021/bm700950u

Hsia Y, Gnesa E, Jeffery F, et al. (2011) Spider Silk Composites and Applications. In: Metal, Ceramic and Polymeric Composites for Various Uses, 1st ed. InTech, Shanghai, China, pp 303–324

Huang X, Powers R (2001) Validity of using the radius of gyration as a restraint in NMR protein structure determination. J Am Chem Soc 123:3834–3835. doi: 10.1021/ja005770p

Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38. doi: 10.1016/0263-7855(96)00018-5

Hutchinson EG, Thornton JM (1996) PROMOTIF—a program to identify and analyze structural motifs in proteins. Protein Sci 5:212-220. doi: 10.1002/pro.5560050204

Ishima R (2012) Recent Developments in ^{15}N NMR Relaxation Studies that Probe Protein Backbone Dynamics. In: Zhu G (ed) NMR of Proteins and Small Biomolecules. Springer Berlin Heidelberg, pp 99–122

Ittah S, Cohen S, Garty S, et al. (2006) An essential role for the C-terminal domain of a dragline spider silk protein in directing fiber formation. Biomacromolecules 7:1790–1795. doi: 10.1021/bm060120k

Iwadate M, Asakura T, Williamson MP (1999) Cα and Cβ carbon-13 chemical shifts in proteins from an empirical database. J Biomol NMR 13:199–211.

Iwaï H, Züger S, Jin J, Tam P-H (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. FEBS Letters 580:1853–1858. doi: 10.1016/j.febslet.2006.02.045

Izdebski T, Akhenblit P, Jenkins JE, et al. (2010) Structure and dynamics of aromatic residues in spider silk: 2D carbon correlation NMR of dragline fibers. Biomacromolecules 11:168–174. doi: 10.1021/bm901039e

Jarymowycz VA, Stone MJ (2006) Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. Chem Rev 106:1624–1671. doi: 10.1021/cr040421p

Jaudzems K, Askarieh G, Landreh M, et al. (2012) pH-dependent dimerization of spider silk N-terminal domain requires relocation of a wedged tryptophan side chain. J Mol Biol 422:477–487. doi: 10.1016/j.jmb.2012.06.004

Jiao D, Barfield M, Hruby VJ (1993) Ab initio IGLO study of the phi- and chi-angle dependence of the carbon-13 chemical shifts in the model peptide N-acetyl-N'-methylglycinamide. J Am Chem Soc. doi: 10.1021/ja00076a052

Jin H-J, Kaplan DL (2003) Mechanism of silk processing in insects and spiders. Nature 424:1057–1061. doi: 10.1038/nature01809

Jones JA, Wilkins DK, Smith LJ, Dobson CM (1997) Characterisation of protein unfolding by NMR diffusion measurements. J Biomol NMR 10:199–203. doi: 10.1023/A:1018304117895

Jones MN, Skinner HA, Tipping E (1975) The interaction between bovine serum albumin and surfactants. Biochem J 147:229–234.

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637. doi: 10.1002/bip.360221211

Kadeřávek P, Zapletal V, Rabatinová A, et al. (2014) Spectral density mapping protocols for analysis of molecular motions in disordered proteins. J Biomol NMR 58:193–207. doi: 10.1007/s10858-014-9816-4

Kamatari YO, Kitahara R, Yamada H, et al. (2004) High-pressure NMR spectroscopy for characterizing folding intermediates and denatured states of proteins. Methods 34:133–143. doi: 10.1016/j.ymeth.2004.03.010

Kanwar M, Wright RC, Date A, et al. (2013) Protein switch engineering by domain insertion. Meth Enzymol 523:369–388. doi: 10.1016/B978-0-12-394292-0.00017-5

Karplus M (1963) Vicinal Proton Coupling in Nuclear Magnetic Resonance. J Am Chem Soc 85:2870–2871. doi: 10.1021/ja00901a059

Karplus PA (1996) Experimentally observed conformation-dependent geometry and hidden strain in proteins. Protein Sci 5:1406–1420. doi: 10.1002/pro.5560050719

Kauzmann W (1959) Some factors in the interpretation of protein denaturation. Advances in protein chemistry 14:1–63.

Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Mag Res 89:496–514. doi: 10.1016/0022-2364(90)90333-5

Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28:8972–8979. doi: 10.1021/bi00449a003

Keeler J (2002) Understanding NMR Spectroscopy. Wiley & Sons, Cambridge, UK

Keten S, Xu Z, Ihle B, Buehler MJ (2010) Nanoconfinement controls stiffness, strength and mechanical toughness of β-sheet crystals in silk. Nat Mater 9:359–367. doi: 10.1038/nmat2704

Kim H, Gao J, Burgess DJ (2009) Evaluation of solvent effects on protonation using NMR spectroscopy: implication in salt formation. Int J Pharm 377:105–111. doi: 10.1016/j.ijpharm.2009.05.018

Knight D, Vollrath F (1999) Hexagonal columnar liquid crystal in the cells secreting spider silk. Tissue Cell 31:617–620. doi: 10.1054/tice.1999.0076

Ko FK, Jovicic J (2004) Modeling of mechanical properties and structural design of spider web. Biomacromolecules 5:780–785. doi: 10.1021/bm0345099

Koehorst RBM, Spruijt RB, Hemminga MA (2008) Site-directed fluorescence labeling of a membrane protein with BADAN: Probing protein topology and local environment. Biophys J 94:3945–3955. doi: 10.1529/biophysj.107.125807

Koenig BW, Ferretti JA, Gawrisch K (1999) Site-specific deuterium order parameters and membrane-bound behavior of a peptide fragment from the intracellular domain of HIV-1 gp41. Biochemistry 38:6327–6334. doi: 10.1021/bi982800g

Kozlowski LP, Bujnicki JM (2012) MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinf 13:111. doi: 10.1186/1471-2105-13-111

Kronqvist N, Otikovs M, Chmyrov V, et al. (2014) Sequential pH-driven dimerization and stabilization of the N-terminal domain enables rapid spider silk formation. Nat Commun 5:3254. doi: 10.1038/ncomms4254

Kundu B, Kurland NE, Bano S, Patra C (2014) Silk proteins for biomedical applications: bioengineering perspectives. Prog Polym Sci 39:251–267. doi: 10.1016/j.progpolymsci.2013.09.002

Kunugi S, Tanaka N (2002) Cold denaturation of proteins under high pressure. Biochim Biophys Acta 1595:329-344. doi: 10.1016/S0167-4838(01)00354-5

Kuszewski J, Gronenborn AM, Clore GM (1996a) A potential involving multiple proton chemical-shift restraints for nonstereospecifically assigned methyl and methylene protons. J Mag Res Ser B 112:79–81.

Kuszewski J, Gronenborn AM, Clore GM (1997) Improvements and extensions in the conformational database potential for the refinement of NMR and X-ray structures of proteins and nucleic acids. J Mag Res 125:171–177. doi: 10.1006/jmre.1997.1116

Kuszewski J, Gronenborn AM, Clore GM (1996b) Improving the quality of NMR and crystallographic protein structures by means of a conformational database potential derived from structure databases. Protein Sci 5:1067–1080. doi: 10.1002/pro.5560050609

Kuszewski J, Gronenborn AM, Clore GM (1999) Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. J Am Chem Soc 121:2337–2338. doi: 10.1021/ja9843730

Kümmerlen J, van Beek JD, Vollrath F (1996) Local structure in spider dragline silk investigated by two-dimensional spin-diffusion nuclear magnetic resonance. Macromolecules 29:2920–2928. doi: 10.1021/ma951098i

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132. doi: 10.1016/0022-2836(82)90515-0

La Mattina C, Reza R, Hu X, et al. (2008) Spider minor ampullate silk proteins are constituents of prey wrapping silk in the cob weaver *Latrodectus hesperus*. Biochemistry 47:4692–4700. doi: 10.1021/bi800140q

Lammel A, Schwab M, Hofer M, et al. (2011) Recombinant spider silk particles as drug delivery vehicles. Biomaterials 32:2233–2240. doi: 10.1016/j.biomaterials.2010.11.060

Lammel A, Schwab M, Slotta U, et al. (2008) Processing conditions for the formation of spider silk microspheres. ChemSusChem 1:413–416. doi: 10.1002/cssc.200800030

Langelaan DN, Bebbington EM, Reddy T, Rainey JK (2009) Structural Insight into G-Protein Coupled Receptor Binding by Apelin. Biochemistry 48:537–548. doi: 10.1021/bi801864b

Langelaan DN, Wieczorek M, Blouin C, Rainey JK (2010) Improved helix and kink characterization in membrane proteins allows evaluation of kink sequence predictors. J Chem Inf Model 50:2213–2220. doi: 10.1021/ci100324n

Laskowski RA, Rullmannn JA, MacArthur MW, et al. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486. doi: 10.1007/BF00228148

Le H, Oldfield E (1994) Correlation between 15N NMR chemical shifts in proteins and secondary structure. J Biomol NMR 4:341–348. doi: 10.1007/BF00179345

Lee D, Hilty C, Wider G, Wüthrich K (2006) Effective rotational correlation times of proteins from NMR relaxation interference. J Mag Res 178:72–76. doi: 10.1016/j.jmr.2005.08.014

Lee Y-Z, Lee Y-T, Lin Y-J, et al. (2014) A streamlined method for preparing split intein for NMR study. Protein Expr Purif 99:106–112. doi: 10.1016/j.pep.2014.04.005

Lefèvre T, Boudreault S, Cloutier C, Pézolet M (2011) Diversity of molecular transformations involved in the formation of spider silks. J Mol Biol 405:238–253. doi: 10.1016/j.jmb.2010.10.052

Lefèvre T, Rousseau M-E, Pézolet M (2006) Orientation-insensitive spectra for Raman microspectroscopy. Appl Spectrosc 60:841–846. doi: 10.1366/000370206778062039

Lefèvre T, Rousseau M-E, Pézolet M (2007) Protein secondary structure and orientation in silk as revealed by Raman spectromicroscopy. Biophys J 92:2885–2895. doi: 10.1529/biophysj.106.100339

Lesage A, Bardet M, Emsley L (1999) Through-Bond Carbon−Carbon Connectivities in Disordered Solids by NMR. J Am Chem Soc 121:10987–10993. doi: 10.1021/ja992272b

Lewis RV (2006) Spider Silk: Ancient Ideas for New Biomaterials. Chem Rev 106:3762–3774. doi: 10.1021/cr010194g

Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. Proteins 61:704–721. doi: 10.1002/prot.20660

Li SC, Deber CM (1994) A measure of helical propensity for amino acids in membrane environments. Nat Struct Mol Biol 1:558–373.

Li X-GG, Wu L-YY, Huang M-RR, et al. (2008) Conformational transition and liquid crystalline state of regenerated silk fibroin in water. Biopolymers 89:497–505. doi: 10.1002/bip.20905

Lide DR (2000) CRC Handbook of Chemistry and Physics, 81st Edition. CRC Press

Liivak O, Flores A, Lewis R, Jelinski LW (1997) Conformation of the polyalanine repeats in minor ampullate gland silk of the spider Nephila clavipes. Macromolecules 30:7127–7130. doi: 10.1021/ma970834a

Lim WK, Rösgen J, Englander SW (2009) Urea, but not guanidinium, destabilizes proteins by forming hydrogen bonds to the peptide group. Proc Natl Acad Sci USA 106:2595–2600. doi: 10.1073/pnas.0812588106

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

Linge JP, Habeck M, Rieping W, Nilges M (2004) Correction of spin diffusion during iterative automated NOE assignment. J Mag Res 167:334–342. doi: 10.1016/j.jmr.2004.01.010

Linge JP, ODonoghue SI, Nilges M (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. Meth Enzymol 339:71–90.

Linge JP, Williams MA, Spronk CAEM, et al. (2003) Refinement of protein structures in explicit solvent. Proteins 50:496–506. doi: 10.1002/prot.10299

Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104:4546–4559. doi: 10.1021/ja011883c

Lipsitz RS, Sharma Y, Brooks BR, Tjandra N (2002) Hydrogen bonding in high-resolution protein structures: a new method to assess NMR protein geometry. J Am Chem Soc 124:10621–10626.

Liu D, Xu R, Cowburn D (2009) Segmental isotopic labeling of proteins for nuclear magnetic resonance. Meth Enzymol 462:151–175. doi: 10.1016/S0076-6879(09)62008-5

Liu Y, Shao Z, Vollrath F (2005) Relationships between supercontraction and mechanical properties of spider silk. Nat Mater 4:901–905. doi: 10.1038/nmat1534

Liu Y, Sponner A, Porter D, Vollrath F (2008) Proline and processing of spider silks. Biomacromolecules 9:116–121. doi: 10.1021/bm700877g

Lobanov MY, Bogatyreva NS, Galzitskaya OV (2008) Radius of gyration as an indicator of protein structure compactness. Molecular Biology 42:701–706. doi: 10.1134/S0026893308040195

Lorieau J, Yao L, Bax A (2008) Liquid crystalline phase of G-tetrad DNA for NMR study of detergent-solubilized proteins. J Am Chem Soc 130:7536–7537. doi: 10.1021/ja801729f

Losonczi JA, Prestegard JH (1998) Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. J Biomol NMR 12:447–451. doi: 10.1023/A:1008302110884

Lovell SC, Davis IW, Arendall WB, et al. (2003) Structure validation by Cα geometry: phi,psi and Cβ deviation. Proteins 50:437–450. doi: 10.1002/prot.10286

Lu Q, Wang X, Lu S, et al. (2011) Nanofibrous architecture of silk fibroin scaffolds prepared with a mild self-assembly process. Biomaterials 32:1059–1067. doi: 10.1016/j.biomaterials.2010.09.072

Lu R-C, Cao A-N, Lai L-H, et al. (2005) Interaction between bovine serum albumin and equimolarly mixed cationic-anionic surfactants decyltriethylammonium bromide-sodium decyl sulfonate. Colloids Surf B 41:139–143. doi: 10.1016/j.colsurfb.2004.11.011

Lukin JA, Gove AP, Talukdar SN, Ho C (1997) Automated probabilistic method for assigning backbone resonances of (^{13}C,^{15}N)-labeled proteins. J Biomol NMR 9:151–166.

Luy B, Kobzar K, Kessler H (2004) An easy and scalable method for the partial alignment of organic molecules for measuring residual dipolar couplings. Angew Chem Int Ed Engl 43:1092–1094. doi: 10.1002/anie.200352860

Ma J, Goldberg GI, Tjandra N (2008) Weak alignment of biomacromolecules in collagen gels: an alternative way to yield residual dipolar couplings for NMR measurements. J Am Chem Soc 130:16148–16149. doi: 10.1021/ja807064k

MacIntosh AC, Kearns VR, Crawford A, Hatton PV (2008) Skeletal tissue engineering using silk biomaterials. J Tissue Eng Regen Med 2:71–80. doi: 10.1002/term.68

Makhatadze GI (1999) Thermodynamics of protein interactions with urea and guanidinium hydrochloride. J Phys Chem B 103:4781–4785. doi: 10.1021/jp990413q

Marion D, Driscoll PC, Kay LE, et al. (2002) Overcoming the overlap problem in the assignment of proton NMR spectra of larger proteins by use of three-dimensional heteronuclear proton-nitrogen-15 Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. Biochemistry 28:6150–6156. doi: 10.1021/bi00441a004

Marion D, Kay LE, Sparks SW (1989) Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. J Am Chem Soc 111:1515–1517. doi: 10.1021/ja00186a066

Marsh JA, Singh VK, Jia Z, Forman-Kay JD (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. Protein Sci 15:2795–2804. doi: 10.1110/ps.062465306

Meersman F, Dobson CM, Heremans K (2006) Protein unfolding, amyloid fibril formation and configurational energy landscapes under high pressure conditions. Chem Soc Rev 35:908–917. doi: 10.1039/B517761H

Meiboom S, Gill D (1958) Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. Rev Sci Instrum 29:688–691. doi: 10.1063/1.1716296

Meier S, Häussinger D, Grzesiek S (2002) Charged acrylamide copolymer gels as media for weak alignment. J Biomol NMR 24:351–356. doi:10.1023/A:1021609207024

Merutka G, Dyson HJ, Wright PE (1995) "Random coil" 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. J Biomol NMR 5:14–24. doi: 10.1007/BF00227466

Mielke SP, Krishnan VV (2004) An evaluation of chemical shift index-based secondary structure determination in proteins: influence of random coil chemical shifts. J Biomol NMR 30:143–153. doi: 10.1023/B:JNMR.0000048940.51331.49

Moelbert S, Emberly E, Tang C (2004) Correlation between sequence hydrophobicity and surface-exposure pattern of database proteins. Protein Sci 13:752–762. doi: 10.1110/ps.03431704

Mohan PMK, Hosur RV (2008) pH dependent unfolding characteristics of DLC8 dimer: Residue level details from NMR. Biochim Biophys Acta 1784:1795–1803. doi: 10.1016/j.bbapap.2008.07.007

Mootz HD, Blum ES, Tyszkiewicz AB, Muir TW (2003) Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo. J Am Chem Soc 125:10561–10569. doi: 10.1021/ja0362813

Morris KF, Johnson CS (1992) Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy. J Am Chem Soc 114:3139–3141. doi: 10.1021/ja00034a071

Muir TW (2008) Studying protein structure and function using semisynthesis. Biopolymers 90:743–750. doi: 10.1002/bip.21102

Mumenthaler C, Braun W (1995) Automated assignment of simulated and experimental NOESY spectra of proteins by feedback filtering and self-correcting distance geometry. J Mol Biol 254:465–480. doi: 10.1006/jmbi.1995.0631

Muona M, Aranko AS, Iwaï H (2008) Segmental isotopic labelling of a multidomain protein by protein ligation by protein *trans*-splicing. ChemBioChem 9:2958–2961.

Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. Nat Methods 3:429–438. doi: 10.1038/nmeth886

Myers JK, Pace CN (1996) Hydrogen bonding stabilizes globular proteins. Biophys J 71:2033–2039. doi: 10.1016/S0006-3495(96)79401-8

Nabeshima Y, Mizuguchi M, Kajiyama A, Okazawa H (2014) Segmental isotope-labeling of the intrinsically disordered protein PQBP1. FEBS Lett 588:4583–4589. doi: 10.1016/j.febslet.2014.10.028

Nabuurs SB, Spronk CAEM, Vriend G, Vuister GW (2004) Concepts and tools for NMR restraint analysis and validation. Concepts Magnetic Res 22A:90–105. doi: 10.1002/cmr.a.20016

Nentwig W (2013) Spider Ecophysiology. Springer Science & Business Media, Bern

Nick Pace C, Huyghues-Despointes BMP, Fu H, et al. (2010) Urea denatured state ensembles contain extensive secondary structure that is increased in hydrophobic proteins. Protein Sci 19:929–943. doi: 10.1002/pro.370

Nilges M, Macias MJ, ODonoghue SI, Oschkinat H (1997) Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from beta-spectrin. J Mol Biol 269:408–422. doi: 10.1006/jmbi.1997.1044

Nolis P, Parella T (2007) Simultaneous alpha/beta spin-state selection for ^{13}C and ^{15}N from a time-shared HSQC-IPAP experiment. J Biomol NMR 37:65–77. doi: 10.1007/s10858-006-9104-z

O'Brien EP, Dima RI, Brooks B, Thirumalai D (2007) Interactions between hydrophobic and ionic solutes in aqueous guanidinium chloride and urea solutions: lessons for protein denaturation mechanism. J Am Chem Soc 129:7346–7353. doi: 10.1021/ja069232

Ohgo K, Kawase T, Ashida J, Asakura T (2006) Solid-state NMR analysis of a peptide (Gly-Pro-Gly-Gly-Ala)6-Gly derived from a flagelliform silk sequence of Nephila clavipes. Biomacromolecules 7:1210–1214. doi: 10.1021/bm0600522

Ohgushi M, Wada A (1983) "Molten-globule state": a compact form of globular proteins with mobile side-chains. FEBS Letters 164:21–24. doi: 10.1016/0014-5793(83)80010-6

Olejniczak ET, Xu RX, Fesik SW (1992) A 4D HCCH-TOCSY experiment for assigning the side chain ^{1}H and ^{13}C resonances of proteins. J Biomol NMR. doi: 10.1007/BF02192854

Omenetto FG, Kaplan DL (2010) New opportunities for an ancient material. Science 329:528–531. doi: 10.1126/science.1188936

Ortega A, Amorós D, García de la Torre J (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. Biophys J 101:892–898. doi: 10.1016/j.bpj.2011.06.046

Osapay K, Case DA (1991) A new analysis of proton chemical shifts in proteins. J Am Chem Soc 113:9436–9444.

Otomo T, Teruya K, Uegaki K, et al. (1999) Improved segmental isotope labeling of proteins and application to a larger protein. J Biomol NMR 14:105–114. doi: 10.1023/A:1008308128050

Otzen DE (2002) Protein Unfolding in Detergents: Effect of micelle structure, ionic strength, pH, and temperature. Biophys J 83:2219–2230. doi: 10.1016/S0006-3495(02)73982-9

Ozawa T, Kaihara A, Sato M, et al. (2001) Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. Anal Chem 73:2516–2521. doi: 10.1021/ac0013296

Palladino P, Rossi F, Ragone R (2010) Effective critical micellar concentration of a zwitterionic detergent: a fluorimetric study on n-dodecyl phosphocholine. J Fluoresc 20:191–196. doi: 10.1007/s10895-009-0537-0

Park SH, Son WS, Mukhopadhyay R, et al. (2009) Phage-induced alignment of membrane proteins enables the measurement and structural analysis of residual dipolar couplings with dipolar waves and lambda-maps. J Am Chem Soc 131:14140–14141. doi: 10.1021/ja905640d

Parkhe AD, Seeley SK, Gardner K, et al. (1997) Structural studies of spider silk proteins in the fiber. J Mol Recognit 10:1–6. doi: 10.1002/(SICI)1099-1352(199701/02)10:1<1::AID-JMR338>3.0.CO;2-7

Pauli J, Baldus M, van Rossum B, de Groot H (2001) Backbone and side-chain ^{13}C and ^{15}N signal assignments of the α-spectrin SH3 domain by magic angle spinning solid-state NMR at 17.6 tesla. ChemBioChem 2:272–281. doi: 10.1002/1439-7633(20010401)2:4<272::AID-CBIC272>3.0.CO;2-2

Perry DJ, Bittencourt D, Siltberg-Liberles J, et al. (2010) Piriform spider silk sequences reveal unique repetitive elements. Biomacromolecules 11:3000–3006. doi: 10.1021/bm1007585

Pettersen EF, Goddard TD, Huang CC, et al. (2004) UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. doi: 10.1002/jcc.20084

Plaxco KW, Morton CJ, Grimshaw SB, et al. (1997) The effects of guanidine hydrochloride on the "random coil" conformations and NMR chemical shifts of the peptide series GGXGG. J Biomol NMR 10:221–230. doi: 10.1023/A:1018340217891

Plaza GR, Corsini P, Marsano E, et al. (2009) Old silks endowed with new properties. Macromolecules 42:8977–8982. doi: 10.1021/ma9017235

Prabhu NV, Sharp KA (2005) Heat capacity in proteins. Annu Rev Phys Chem 56:521–548. doi: 10.1146/annurev.physchem.56.092503.141202

Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. Chem Rev 104:3519–3540. doi: 10.1021/cr030419i

Prevost M, Wodak SJ, Tidor B, Karplus M (1991) Contribution of the hydrophobic Effect to protein stability: analysis based on simulations of the Ile-96 → Ala mutation in barnase. Proc Natl Acad Sci USA 88:10880–10884. doi: 10.2307/2358356

Privalov PL (1990) Cold denaturation of protein. Crit Rev Biochem Mol Biol 25:281–306.

Pronk S, Páll S, Schulz R, et al. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854. doi: 10.1093/bioinformatics/btt055

Quirt AR, Lyerla JR, Peat IR, et al. (1974) Carbon-13 nuclear magnetic resonance titration shifts in amino acids. J Am Chem Soc 96:570–574.

Rainey JK, Fliegel L, Sykes BD (2006) Strategies for dealing with conformational sampling in structural calculations of flexible or kinked transmembrane peptides. Biochem Cell Biol 84:918–929. doi: 10.1139/o06-178

Rath A, Davidson AR, Deber CM (2005) The structure of "unstructured" regions in peptides and proteins: Role of the polyproline II helix in protein folding and recognition. Peptide Science 80:179–185. doi: 10.1002/bip.20227

Reddy JG, Hosur RV (2013) Parallel acquisition of 3D-HA(CA)NH and 3D-HACACO spectra. J Biomol NMR 56:77–84. doi: 10.1007/s10858-013-9735-9

Reddy T, Rainey JK (2010) Interpretation of biomolecular NMR spin relaxation parameters. Biochem Cell Biol 88:131–142. doi: 10.1139/o09-152

Reif MM, Hünenberger PH, Oostenbrink C (2012) New interaction parameters for charged amino acid side chains in the GROMOS force field. J Chem Theory Comput 8:3705–3723. doi: 10.1021/ct300156h

Rey AD, Herrera-Valencia EE (2011) Liquid crystal models of biological materials and silk spinning. Biopolymers 97:374–396. doi: 10.1002/bip.21723

Reynolds JA, Tanford C (1970) The gross conformation of protein-sodium dodecyl sulfate complexes. J Biol Chem 245:5161–5165.

Riekel C, Bränden C, Craig C, et al. (1999) Aspects of X-ray diffraction on single spider fibers. Int J Biol Macromol 24:179–186. doi: 10.1016/S0141-8130(01)00166-0

Riekel C, Madsen B, Knight D, Vollrath F (2000) X-ray diffraction on spider silk during controlled extrusion under a synchrotron radiation X-ray beam. Biomacromolecules 1:622–626. doi: 10.1021/bm000047c

Riekel C, Vollrath F (2001) Spider silk fibre extrusion: combined wide- and small-angle X-ray microdiffraction experiments. Int J Biol Macromol 29:203–210. doi: 10.1016/S0141-8130(01)00166-0

Rieping W, Habeck M, Bardiaux B, et al. (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. Bioinformatics 23:381–382. doi: 10.1093/bioinformatics/btl589

Rising A, Widhe M, Johansson J, Hedhammar M (2010) Spider silk proteins: recent advances in recombinant production, structure–function relationships and biomedical applications. Cell Mol Life Sci 68:169–184. doi: 10.1007/s00018-010-0462-z

Ropars V, Bouguet-Bonnet S, Auguin D, et al. (2007) Unraveling protein dynamics through fast spectral density mapping. J Biomol NMR 37:159–177. doi: 10.1007/s10858-006-9091-0

Rostkowski M, Olsson MHM, Søndergaard CR, Jensen JH (2011) Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. BMC Struct Biol 11:1-6. doi: 10.1186/1472-6807-11-6

Roth CM, Neal BL, Lenhoff AM (1996) Van der Waals interactions involving proteins. Biophys J 70:977–987. doi: 10.1016/S0006-3495(96)79641-8

Ruan K, Tolman JR (2005) Composite alignment media for the measurement of independent sets of NMR residual dipolar couplings. J Am Chem Soc 127:15032–15033. doi: 10.1021/ja055520e

Rule GS, Hitchens TK (2006) Fundamentals of Protein NMR Spectroscopy. Springer Science & Business Media

Rückert M, Otting G (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. J Am Chem Soc 122:7793–7797. doi: 10.1021/ja001068h

Sass H-J, Musco G, Stahl SJ, et al. (2000) Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. J Biomol NMR 18:303–309. doi: 10.1023/A:1026703605147

Sass J, Cordier F, Hoffmann A, et al. (1999) Purple Membrane Induced Alignment of Biological Macromolecules in the Magnetic Field. J Am Chem Soc 121:2047–2055. doi: 10.1021/ja983887w

Schmidt JM, Blümel M, Löhr F, Rüterjans H (1999) Self-consistent ^{3}J coupling analysis for the joint calibration of Karplus coefficients and evaluation of torsion angles. J Biomol NMR 14:1–12. doi: 10.1023/A:1008345303942

Schubeis T, Lührs T, Ritter C (2015) Unambiguous assignment of short- and long-range structural restraints by solid-state NMR spectroscopy with segmental isotope labeling. ChemBioChem 16:51–54. doi: 10.1002/cbic.201402446

Schumann FH, Riepl H, Maurer T, et al. (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. J Biomol NMR 39:275–289. doi: 10.1007/s10858-007-9197-z

Schwarzinger S, Kroon GJ, Foss TR, et al. (2000) Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. J Biomol NMR 18:43–48. doi: 10.1023/A:1008386816521

Schwarzinger S, Kroon GJA, Foss TR, et al. (2001) Sequence-dependent correction of random coil NMR chemical shifts. J Am Chem Soc 123:2970–2978. doi: 10.1021/ja003760i

Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor–NIH for NMR molecular structure determination. Prog Nucl Mag Res Sp 48:47–62. doi: 10.1110/ps.034561.108/full

Scott CP, Abel-Santos E, Wall M, et al. (1999) Production of cyclic peptides and proteins in vivo. Proc Natl Acad Sci USA 96:13638–13643. doi: 10.1073/pnas.96.24.13638

Shen Y, Delaglio F, Cornilescu G, Bax A (2009a) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223. doi: 10.1007/s10858-009-9333-z

Shen Y, Lange O, Delaglio F, et al. (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci USA 105:4685–4690. doi: 10.1073/pnas.0800256105

Shen Y, Vernon R, Baker D, Bax A (2009b) De novo protein structure generation from incomplete chemical shift assignments. J Biomol NMR 43:63–78. doi: 10.1007/s10858-008-9288-5

Shenderovich IG, Burtsev AP, Denisov GS, et al. (2001) Influence of the temperature-dependent dielectric constant on the H/D isotope effects on the NMR chemical shifts and the hydrogen bond geometry of the collidine--HF complex in CDF3/CDClF2 solution. Magn Reson Chem 39:S91–S99. doi: 10.1002/mrc.938

Shi X, Yarger JL, Holland GP (2014) Elucidating proline dynamics in spider dragline silk fibre using ^{2}H-^{13}C HETCOR MAS NMR. Chem Commun 50:4856–4859. doi: 10.1039/c4cc00971a

Silvers R, Buhr F, Schwalbe H (2010) The molecular mechanism of spider-silk formation. Angew Chem Int Ed 49:5410–5412. doi: 10.1002/anie.201003033

Simmons A, Ray E, Jelinski LW (1994) Solid-State ^{13}C NMR of *Nephila clavipes* dragline silk establishes structure and identity of crystalline regions. Macromolecules 27:5235–5237. doi: 10.1021/ma00096a060

Simmons AH, Michal CA, Jelinski LW (1996) Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. Science 271:84–87. doi: 10.1126/science.271.5245.84

Skrisovska L, Schubert M, Allain FH-T (2010) Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. J Biomol NMR 46:51–65. doi: 10.1007/s10858-009-9362-7

Slepkov ER, Rainey JK, Li X, et al. (2005) Structural and functional characterization of transmembrane segment IV of the NHE1 isoform of the Na+/H+ exchanger. J Biol Chem 280:17863–17872. doi: 10.1074/jbc.M409608200

Slotta U, Hess S, Spiess K, et al. (2007) Spider silk and amyloid fibrils: a structural comparison. Macromol Biosci 7:183–188. doi: 10.1002/mabi.200600201

Southworth MW, Adam E, Panne D, et al. (1998) Control of protein splicing by intein fragment reassembly. EMBO J 17:918–926. doi: 10.1093/emboj/17.4.918

Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and C. alpha. and C. beta. 13C nuclear magnetic resonance chemical shifts. J Am Chem Soc 113:5490–5492. doi: 10.1021/ja00014a071

Spronk CAEM, Nabuurs SB, Krieger E, et al. (2004) Validation of protein structures derived by NMR spectroscopy. Prog Nucl Mag Res Sp 45:315–337. doi: 10.1016/j.pnmrs.2004.08.003

Spyracopoulos L (2006) A suite of Mathematica notebooks for the analysis of protein main chain ^{15}N NMR relaxation data. J Biomol NMR 36:215–224. doi: 10.1007/s10858-006-9083-0

Starr FW, Bellissent-Funel MC, Stanley HE (1999) Structure of supercooled and glassy water under pressure. Phys Rev E 60:1084–1087. doi: 10.1103/PhysRevE.60.1084

Stauffer SL, Coguill SL, Lewis RV (1994) Comparison of physical properties of three silks from Nephila clavipes and Araneus gemmoides. J Arachnology 22:5–11. doi: 10.2307/3705704

Stejskal EO, Tanner JE (1965) Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. J Chem Phys 42:288. doi: 10.1063/1.1695690

Steven E, Park JG, Paravastu A, et al. (2011) Physical characterization of functionalized spider silk: electronic and sensing properties. Sci Technol Adv Mater 12:055002. doi: 10.1088/1468-6996/12/5/055002

Sticke DF, Presta LG, Dill KA, Rose GD (1992) Hydrogen bonding in globular proteins. J Mol Biol 226:1143–1159. doi: 10.1016/0022-2836(92)91058-W

Stumpe MC, Grubmüller H (2007) Interaction of Urea with Amino Acids: Implications for Urea-Induced Protein Denaturation. J Am Chem Soc 129:16126–16131. doi: 10.1021/ja076216j

Suzuki Y, Takahashi R, Shimizu T, et al. (2009) Intra- and intermolecular effects on ^{1}H chemical shifts in a silk model Peptide determined by high-field solid state ^{1}H NMR and empirical calculations. J Phys Chem B 113:9756–9761. doi: 10.1021/jp903020p

Szilagyi A, Kardos J, Osvath S, et al. (2007) Protein Folding. In: Lajtha A, Banik N (eds) Handbook of Neurochemistry and Molecular Neurobiology. Springer US, Boston, MA, pp 303–343

Szilágyi L, Jardetzky O (1989) α-Proton chemical shifts and secondary structure in proteins. J Mag Res 83:441–449. doi: 10.1016/0022-2364(89)90341-7

Søndergaard CR, McIntosh LP, Pollastri G, Nielsen JE (2008) Determination of Electrostatic Interaction Energies and Protonation State Populations in Enzyme Active Sites. J Mol Biol 376:269–287. doi: 10.1016/j.jmb.2007.09.070

Takegoshi K, Nakamura S, Terao T (2001) ^{13}C–^{1}H dipolar-assisted rotational resonance in magic-angle spinning NMR. Chem Phys Lett 344:631–637.

Teng Q (2005) Structural Biology: Practical NMR Applications. Springer Science & Business Media, New York

Thanabal V, Omecinsky DO, Reily MD, Cody WL (1994) The ^{13}C chemical shifts of amino acids in aqueous solution containing organic solvents: application to the secondary structure characterization of peptides in aqueous trifluoroethanol solution. J Biomol NMR 4:47–59.

Thiel BL, Guess KB, Viney C (1997) Non-periodic lattice crystals in the hierarchical microstructure of spider (major ampullate) silk. Peptide Science 41:703–719. doi: 10.1002/(SICI)1097-0282(199706)41:7<703::AID-BIP1>3.0.CO;2-T

Tian M, Lewis RV (2005) Molecular characterization and evolutionary study of spider tubuliform (eggcase) silk protein. Biochemistry 44:8006–8012. doi: 10.1021/bi050366u

Tokareva O, Jacobsen M, Buehler M, et al. (2014) Structure-function-property-design interplay in biopolymers: spider silk. Acta Biomater 10:1612–1626. doi: 10.1016/j.actbio.2013.08.020

Tonan K, Ikawa S-I (2003) Effect of solvent on an NMR chemical shift difference between glycyl geminal a-protons as a probe of b-turn formation of short peptides. Spectrochim Acta, Part A 59:111–120. doi: 10.1016/s1386-1425(02)00115-4

Tremblay M-L, Banks AW, Rainey JK (2010) The Predictive Accuracy Of Secondary Chemical Shifts Is More Affected By Protein Secondary Structure Than Solvent Environment. J Biomol NMR 46:257–270. Doi: 10.1007/S10858-010-9400-5

Trempe J-F, Morin FG, Xia Z, et al. (2002) Characterization of polyacrylamide-stabilized Pfl phage liquid crystals for protein NMR spectroscopy. J Biomol NMR 22:83–87.

Tsai C-J, Maizel JV, Nussinov R (2002) The hydrophobic effect: a new insight from cold denaturation and a two-state water structure. Crit Rev Biochem Mol Biol 37:55–69. doi: 10.1080/10409230290771456

Turro NJ, Lei XG, Ananthapadmanabhan KP, Aronson M (1995) Spectroscopic probe analysis of protein-surfactant interactions: the BSA/SDS system. Langmuir 11:2525–2533. doi: 10.1021/la00007a035

Tusnády GE, Dosztányi Z, Simon I (2004) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. Nucleic Acids Res 33:D275–D278. doi: 10.1093/nar/gki002

Twomey EC, Cordasco DF, Wei Y (2012) Profound conformational changes of PED/PEA-15 in ERK2 complex revealed by NMR backbone dynamics. Biochim Biophys Acta 1824:1382–1393. doi: 10.1016/j.bbapap.2012.07.001

Tyn MT, Gusek TW (1990) Prediction of diffusion coefficients of proteins. Biotechnol Bioeng 35:327–338. doi: 10.1002/bit.260350402

Ulrich EL, Akutsu H, Doreleijers JF, et al. (2008) BioMagResBank. Nucleic Acids Res 36:D402–D408. doi: 10.1093/nar/gkm957

Valafar H, Prestegard JH (2004) REDCAT: a residual dipolar coupling analysis tool. J Mag Res 167:228–241. doi: 10.1016/j.jmr.2003.12.012

Valluzzi R, Szela S, Avtges P, Kirschner D (1999) Methionine redox controlled crystallization of biosynthetic silk spidroin. J Exp Biol 103:11382–11392. doi: 10.1021/jp991363s

van Beek JD, Hess S, Vollrath F, Meier BH (2002) The molecular structure of spider dragline silk: Folding and orientation of the protein backbone. Proc Natl Acad Sci USA 99:10266–10271. doi: 10.1073/pnas.152162299

van Beek JD, Meier BH, Schäfer H (2003) Inverse methods in two-dimensional NMR spectral analysis. J Mag Res 162:141–157. doi: 10.1016/S1090-7807(02)00193-3

Van Der Spoel D, Lindahl E, Hess B, et al. (2005) GROMACS: Fast, flexible, and free. J Comput Chem 26:1701–1718. doi: 10.1002/jcc.20291

Vasanthavada K, Hu X, Falick AM, et al. (2007) Aciniform spidroin, a constituent of egg case sacs and wrapping silk fibers from the black widow spider Latrodectus hesperus. J Biol Chem 282:35088–35097. doi: 10.1074/jbc.M705791200

Vila-Perelló M, Muir TW (2010) Biological Applications of Protein Splicing. Cell 143:191–200. doi: 10.1016/j.cell.2010.09.031

Viles JH, Donne D, Kroon G, et al. (2001) Local Structural Plasticity of the Prion Protein. Analysis of NMR Relaxation Dynamics. Biochemistry 40:2743–2753. doi: 10.1021/bi002898a

Vinogradov SN, Linnell RH (1971) Hydrogen Bonding. New York

Voet D, Voet JG (2010) Biochemistry, 4th Edition. Wiley Global Education

Volkmann G, Iwaï H (2010) Protein trans-splicing and its use in structural biology: opportunities and limitations. Mol BioSyst 6:2110. doi: 10.1039/c0mb00034e

Volkmann G, Murphy PW, Rowland EE, et al. (2010) Intein-mediated cyclization of bacterial acyl carrier protein stabilizes its folded conformation but does not abolish function. J Biol Chem 285:8605–8614. doi: 10.1074/jbc.M109.060863

Vollrath F, Knight DP (2001) Liquid crystalline spinning of spider silk. Nature 410:541–548. doi: 10.1038/35069000

Vollrath F, Knight DP (1999) Structure and function of the silk production pathway in the Spider Nephila edulis. Int J Biol Macromol 24:243–249. doi: 10.1016/S0141-8130(98)00095-6

Vranken WF, Boucher W, Stevens TJ, et al. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696. doi: 10.1002/prot.20449

Vriend G (1990) WHAT IF: A molecular modeling and drug design program. J Mol Graph 8:52–56. doi: 10.1016/0263-7855(90)80070-V

Vuister G, Tessari M, Karimi-Nejad Y, Whitehead B (2002) Pulse sequences for measuring coupling constants. In: Modern Techniques in Protein NMR. Springer, pp 195–257

Vuister GW, Bax A (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNHa) coupling constants in ^{15}N-enriched proteins. J Am Chem Soc 115:7772–7777. doi: 10.1021/ja00070a024

Wang B, He X, Merz KM (2013) Quantum mechanical study of vicinal J spin–spin coupling constants for the protein backbone. J Chem Theory Comput 9:4653–4659. doi: 10.1021/ct400631b

Wang S, Huang W, Yang D (2012) NMR structure note: repetitive domain of aciniform spidroin 1 from Nephila antipodiana. J Biomol NMR 54:415–420. doi: 10.1007/s10858-012-9679-5

Wang S, Huang W, Yang D (2014) Structure and function of C-terminal domain of aciniform spidroin. Biomacromolecules 15:468–477. doi: 10.1021/bm401709v

Wang Y, Jardetzky O (2002a) Probability-based protein secondary structure identification using combined NMR chemical-shift data. Protein Sci 11:852–861. doi: 10.1110/ps.3180102

Wang Y, Jardetzky O (2002b) Investigation of the neighboring residue effects on protein chemical shifts. J Am Chem Soc 124:14075–14084. doi: 10.1021/ja026811f

Wen Y, Li J, Yao W, et al. (2010) Unique structural characteristics of the rabbit prion protein. J Biol Chem 285:31682–31693. doi: 10.1074/jbc.M110.118844

Wendt H, Hillmer A, Reimers K, et al. (2011) Artificial skin-culturing of different skin cell lines for generating an artificial skin substitute on cross-weaved spider silk fibres. PLoS ONE 6:e21833. doi: 10.1371/journal.pone.0021833

Williamson MP (1990) Secondary-structure dependent chemical shifts in proteins. Peptide Science 29:1423–1431. doi: 10.1002/bip.360291009

Williamson MP, Asakura T, Nakamura E, Demura M (1992) A method for the calculation of protein alpha-CH chemical shifts. J Biomol NMR 2:83–98.

Williamson MP, Craven CJ (2009) Automated protein structure calculation from NMR data. J Biomol NMR 43:131–143. doi: 10.1007/s10858-008-9295-6

Wishart DS, Arndt D, Berjanskii M, et al. (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Res 36:W496–502. doi: 10.1093/nar/gkn305

Wishart DS, Bigam CG, Holm A, et al. (1995a) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J Biomol NMR 5:67–81. doi: 10.1007/BF00227471

Wishart DS, Bigam CG, Yao J, et al. (1995b) 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 6:135–140. doi: 10.1007/BF00211777

Wishart DS, Sykes BD (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 4:171–180. doi: 10.1007/BF00175245

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 31:1647–1651. doi: 10.1021/bi00121a010

Wolff JO, Grawe I, Wirth M, et al. (2015) Spider's super-glue: thread anchors are composite adhesives with synergistic hierarchical organization. Soft Matter. doi: 10.1039/c4sm02130d

Wong Po Foo C, Kaplan DL (2002) Genetic engineering of fibrous proteins: spider dragline silk and collagen. Adv Drug Deliv Rev 54:1131–1143. doi: 10.1016/S0169-409X(02)00061-3

Wu DH, Chen AD, Johnson CS (1995) An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. J Mag Res Ser A 115:260–264. doi: 10.1006/jmra.1995.1176

Wu H, Hu Z, Liu XQ (1998) Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. Proc Natl Acad Sci USA 95:9226–9231.

Wüthrich K (1986) NMR of Proteins and Nucleic Acids. Wiley-Interscience, New York

Wüthrich K, Wagner G (1979) Nuclear magnetic resonance of labile protons in the basic pancreatic trypsin inhibitor. J Mol Biol 130:1–18. doi: 10.1016/0022-2836(79)90548-5

Xu L, Rainey JK, Meng Q, Liu X-Q (2012a) Recombinant Minimalist Spider Wrapping Silk Proteins Capable of Native-Like Fiber Formation. PLoS ONE 7:e50227. doi: 10.1371/journal.pone.0050227

Xu L, Tremblay M-L, Meng Q, et al. (2012b) ^{1}H, ^{13}C and ^{15}N NMR assignments of the aciniform spidroin (AcSp1) repetitive domain of *Argiope trifasciata* wrapping silk. Biomol NMR Assign 6:147–151. doi: 10.1007/s12104-011-9344-z

Xu L, Tremblay M-L, Orrell KE, et al. (2013) Nanoparticle self-assembly by a highly stable recombinant spider wrapping silk protein subunit. FEBS Lett 587:3273–3280. doi: 10.1016/j.febslet.2013.08.024

Xu X-P, Case DA (2001) Automated prediction of 15N, 13Cα, 13Cβ and 13C′ chemical shifts in proteins using a density functional database. J Biomol NMR 21:321–333. doi: 10.1023/A:1013324104681

Xue B, Dunbrack RL, Williams RW, et al. (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 1804:996–1010. doi: 10.1016/j.bbapap.2010.01.011

Yamauchi K, Kuroki S, Fujii K, Ando I (2000) The amide proton NMR chemical shift and hydrogen-bonded structure of peptides and polypeptides in the solid state as studied by high-frequency solid-state ^{1}H NMR. Chem Phys Lett 324:435–439. doi: 10.1038/pj.2012.95

Yamazaki T, Otomo T, Oda N, Kyogoku Y (1998) Segmental isotope labeling for protein NMR using peptide splicing. ChemBioChem 120:5591–5592. doi: 10.1021/ja980776o

Yang Z, Liivak O, Seidel A, et al. (2000) Supercontraction and Backbone Dynamics in Spider Silk: ^{13}C and ^{2}H NMR Studies. J Am Chem Soc 122:9019–9025. doi: 10.1021/ja0017099

Yao S, Hinds MG, Norton RS (1998) Improved estimation of protein rotational correlation times from ^{15}N relaxation measurements. J Mag Res 131:347–350. doi: 10.1006/jmre.1998.1382

Yazawa K, Suzuki F, Nishiyama Y, et al. (2012) Determination of accurate 1H positions of an alanine tripeptide with anti-parallel and parallel β-sheet structures by high resolution 1H solid state NMR and GIPAW chemical shift calculation. Chem Commun 48:11199–11201. doi: 10.1039/c2cc36300c

Yuan Z, Zhang F, Davis MJ, et al. (2006) Predicting the solvent accessibility of transmembrane residues from protein sequence. J Proteome Res 5:1063–1070. doi: 10.1021/pr050397b

Zech SG, Wand AJ, McDermott AE (2005) Protein structure determination by high-resolution solid-state NMR spectroscopy: application to microcrystalline ubiquitin. J Am Chem Soc 127:8618–8626. doi: 10.1021/ja0503128

Zhang X, Baughman CB, Kaplan DL (2008) In vitro evaluation of electrospun silk fibroin scaffolds for vascular cell growth. Biomaterials 29:2217–2227. doi: 10.1016/j.biomaterials.2008.01.022

Zhao A, Zhao T, Sima Y, et al. (2005) Unique molecular architecture of egg case silk protein in a spider, Nephila clavata. J Biochem 138:593–604. doi: 10.1093/jb/mvi155

Zhao A-C, Zhao T-F, Nakagaki K, et al. (2006) Novel molecular and mechanical properties of egg case silk from wasp spider, *Argiope bruennichi*. Biochemistry 45:3348–3356. doi: 10.1021/bi052414g

Zhao C, Asakura T (2001) Structure of silk studied with NMR. Prog Nucl Mag Res Sp 39:301–352. doi: 10.1016/S0079-6565(01)00039-5

Zuiderweg ERP, Nettesheim DG, Mollison KW, Carter GW (1989) Tertiary structure of human complement component C5a in solution from nuclear magnetic resonance data. Biochemistry 28:172–185. doi: 10.1021/bi00427a025

Züger S, Iwaï H (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nat Biotechnol 23:736–740. doi: 10.1038/nbt1097

Zweckstetter M, Bax A (2001) Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. J Biomol NMR 20:365–377. doi: 10.1023/A:1011263920003

Zweckstetter M, Bax A (2002) Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. J Biomol NMR 23:127–137.

Zweckstetter M, Bax A (2000) Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. J Am Chem Soc 122:3791–3792. doi: 10.1021/ja0000908