DISULPHIDE LOCKING: CONTRASTING EFFECTS ON DISPARATE PROTEINS

by

Anamika Sulekha

Submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

at

Dalhousie University

Halifax, Nova Scotia

November 2021

DEDICATION

In solidarity with the anonymous faces battling the unspoken and rocky road towards protecting their educational rights and pursuing their dreams, I would like to dedicate my thesis to them while fighting against the barriers of age, ethnicity, gender, region and religion for freedom of choice in life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

It is great pleasure to express my gratitude towards my supervisor Prof. Jan K. Rainey for allowing me to join his lab as a graduate student and entrusting me with such wonderful research projects. His constant guidance, support and patience have not only helped me develop as an independent researcher, but also grow as an individual.

I am also grateful to all my former and current colleagues in the Rainey lab for providing a friendly and proactive work environment. I am thankful to Drs. Kyungsoo Shin, Lingling Xu and Andy Song, and lab manager Bruce Stewart who familiarized me with biochemistry research techniques. I would also like to thank Nathan Weatherbee-Martin, Chloe Ma, Jeff Simmons, Tam Pham, Anupama Ghimire, Stefan Warkentin, Lizzy Baker, Dr. Farideh Badichi, Sara Evans and Jayatee Ray for being great lab mates. I would like to acknowledge my supervisory committee members Prof. Laurent Kreplak, Dr. Vanya Ewart and Dr. David Langelaan for availing helpful discussions, constructive criticism, and willingness to troubleshoot throughout my graduate program. I am thankful to Drs. Paul Liu, Stephen Bearne, David Langelaan, Melanie Dobson and their research groups, Paul Briggs and Heidi MacKinnon for offering access to instrumental facilities. Graduate training sessions coordinated by Prof. Allison Thompson and Nazanin Omidvar have helped me broaden my career goals. I am thankful to Dr. Ashok P. Masilamani for providing me with an opportunity to intern with Stratuscent Inc., Montreal. I would also like to express my gratitude to all the NMR people for their expertise and assistance throughout this thesis work: Prof. Lewis Kay and Drs. Ranjith Muhandiram (University of Toronto, CA), Tara Sprules (Quebec/Eastern Canada High Field NMR Facility, Montreal, CA), Mike Lumsden (Nuclear Magnetic Resonance Research Resource (NMR3), Dalhousie University, CA) and Mr. Ian Burton (National Research Council of Canada, Halifax, CA) for helping with solution-state NMR data

acquisition and Drs. Zhehong Gan and Ivan Hung (National High Magnetic Field Laboratory, Tallahassee, USA) for assisting with solid-state NMR data acquisition. I would like to thank the mass spectrometry unit and scientific imaging suite (Dalhousie University, Halifax, CA) for availing the MS and SEM facilities. I would also like to thank Roisin McDevitt for all the clerical support and personal warmth throughout my degree program. Finally, I would like to thank NSERC, CREATE-Bioactives, the department of Biochemistry and Molecular Biology and Dalhousie University for funding this piece of work.

I am grateful to my previous mentors Drs. Vinesh Vijayan, Ravi Maruthachalam and Shyamala Thirunavukkuarasu at IISER, Thiruvananthapuram. My heartful thanks to Prakaash Thirunavukkuarasu, Pooja Srinivasan, Dhivya Jayaraman, Mukund Mohan, Meenakshi Mukund, Malavika Mukund, Sumi Girijan, Aswini Prasannan, Bijish Lakshmanan, Amitha Wilson and Shiju Koshy for being family in Halifax. Heartfelt thanks to Meera G. and Prof. O. Thomas for inculcating the passion for science and being my constants ever since I met them. Words are not just enough to express my gratitude towards Anjali Anilkumar, Arya Aravind, Dr. Kavya Vinayan, Dr. Alphy John, Dr.Vishnu Sreepal, Akhilraj K.V., and Vipin Vikraman for their immense support, care, encouragement, understanding and friendship over the past one decade. More importantly, I thank Ammamma and Appuchan (grandparents), my better half, Dr. Arun Kesavan and his family for being really kind and very supportive through rough times.

Thank you!

LIST OF SYMBOLS AND ABBREVIATIONS

α	Regularization parameter (DLS)
A	Absorbance
A_0	Original cross-sectional area
Aβ	Amyloid-β
AcSp1	Aciniform spidroin 1
can	Acetonitrile
Amp	Ampicillin
ADR	Ambiguous distance restraint
ARIA	Ambiguous restraint for iterative assignment
AS	As spun
$\mathbf{B_0}$	External magnetic field
BME	β-mercaptoethanol
CD	Circular dichroism
CS	Chemical shift
CSα/α	Cys-stabilized helix-loop-helix
CSI	Chemical shift index
CSD	Chemical shift displacement
Cτ	Rotational correlation function
CT-CPMG	Constant time-Carr–Purcell–Meiboom–Gill
CTD	C-terminal non-repetitive domain
δ	Chemical shift
δ	Gradient length (in DOSY)
Δ	Diffusion time
$\Delta\delta$	Secondary chemical shift
$\bar{d}$	Inter-nuclear distances
d	Delay time
D_C	Translational diffusion coefficient
D_2O	Deuterium oxide
DANGLE	Dihedral Angles from Global Likelihood Estimates
d_c	Dipolar constant
D_C	Translational diffusion coefficient
$\overline{d_{NH}}$	Normalized chemical shift displacements
ΔL	Change in length
dH_2O	Distilled water
DLS	Dynamic light scattering
DOSY	Diffusion ordered spectroscopy
DPC	Dodecyl phosphocholine

DSS	2,2-dimethyl-2-silapentane-5-sulfonate
DTT	Dithiothreitol
Ɛ	Molar extinction coefficient
ε	Strain
η	NOE enhancement factor
η	NOE enhancement factor
EtOH	Ethanol
Γ	Gyromagnetic ratio
G	Relative centrifugal force
GSH	Glutathione (reduced)
GSSG	Glutathione (oxidized)
$G_1(\tau)$	Normalized electric field correlation function in DLS
$G_2(\tau)$	Correlation function measuring intensity fluctuations in DLS
GUI	Graphical user interface
H5	Helix 5
H_6 SUMO	Hexa-histidine tagged small ubiquitin-like modifier
het-NOE	Heteronuclear nuclear Overhauser effect
HFA	Hexafluoroacetone
HFIP	1,1,1,3,3,3-hexafluoro-2-propanol
HPLC	High-performance liquid chromatography
HSQC	Heteronuclear single quantum correlation spectroscopy
ICK	Inhibitor cystine knot
IMAC	Immobilized metal affinity chromatographic
IPA	Isopropyl alcohol
IPTG	Isopropyl β-d-1-thiogalactopyranoside
$J(\omega)$	Spectral density function at frequency w
$j(\omega)$	Reduced spectral density function at frequency w
l	Pathlength
L	Regularizer operator (DLS)
LED	Longitudinal eddy current delay
LB	Lysogeny broth
MeOH	Methanol
MRE	Mean residue ellipticity
M	Molecular weight
MWCO	Molecular weight cut off
NaAc	Sodium acetate
$^{15}NH_4Cl$	^{15}N-labelled ammonium chloride
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
NNLS	Non-negative least squares

NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
NRC	National Research Council
NTD	N-terminal non-repetitive domain
OD_{600}	Optical density at 600 nm
O/N	Overnight
ω_0	Larmor frequency
$\parallel$	Parallel
$\perp$	Perpendicular
PAGE	Polyacrylamide gel electrophoresis
PDI	Polydispersity index
PEEK	Polyetheretherketone
PFG	Pulsed field gradient
PB	Phosphate buffer
PS	Post spun
q	Bragg's vector
QANUC	Quebec/eastern Canada high field NMR facility
r	Hydrodynamic radius of the diffusing molecule
R_1	Longitudinal relaxation rate constant
R_2	Transverse relaxation rate constant
$R_{2,eff}$	Conformational sampling rate
RC	Random coil
RF	Reverse purification
R_g	Radius of gyration
RMSD	Root mean square deviation
RN	Reversible needle
R-T	Richardson-Thornton rules
RT	Room temperature
σ	Stress
σ	Variance
σ_{IS}	Cross-relaxation
σ_{NOE}	Error estimated for NOE
S	Generalized order parameter
SD	Standard deviations
SAM	S-adenosylmethionine
SDS	Sodium dodecyl sulphate
SEM	Scanning electron microscopy
SOR	Soricidin
τ	Delay time in DLS measurements

τ_c	Rotational correlation time
$[\theta]$	Mean residue ellipticity
T_1	Longitudinal relaxation time constant
T_2	Transverse relaxation time constant
TCEP	Tris(2-carboxyethyl)phosphine
TCI	Triple resonance inverse
TFA	Trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
T_m	Melting temperature
TOCSY	Total correlation spectroscopy
TRP	Transient receptor potential
[U-^{13}C,^{15}N]	Uniformly ^{13}C and ^{15}N isotope enriched
UTS	Ultimate tensile strength
UV/vis	Ultraviolet/visible
VT-NMR	Variable-temperature NMR
W	Repeat unit
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
WT	Wild type
X	Intensity-weighted arithmetic and harmonic diameters

ABSTRACT

This thesis details a series of structural and biophysical studies on two different proteins containing disulphide bonds: (1) soricidin and (2) engineered mutants of aciniform spider silk. All studies relied upon recombinant protein expression, purification, and refolding, with several biophysical and materials characterization techniques applied, including solution-state nuclear magnetic resonance (NMR) spectroscopy as the major tool. Soricidin is a venomous protein isolated from shrew saliva whose short derivatives without any disulphide linkages are reported as effective anticancer drugs. Here, I report drastic improvements in refolding to produce the paralytic conformer and present for the first time the atomic structure of the full-length bioactive protein. With one disulphide linking the N-terminal end of the protein to an α-helical segment in the otherwise disordered C-terminal tail to form an enclosed loop and two α-helices cross-linked by the other two disulphides, soricidin demonstrates a I–VI/II–IV/III–V disulphide connectivity. Although cysteine motifs and exposed Lys/Arg dyad are consistent with other venoms, soricidin adopts a rarely reported cysteine-stabilized helix-loop-helix fold.

With the objective of identifying the trigger for fibrillogenesis in aciniform spider silk, conformation and dynamics were probed using engineered mutant forms of aciniform spider silk in the reduced (i.e., sulfhydryl-containing cysteine side chains) vs. disulphide-linked state. The reduced state of the aciniform silk mutant is capable of silk-like fibre formation and can be wet-spun into silk-like fibres with improved mechanical properties relative to the wild-type protein. The disulphide-locked state, conversely, is unable to form silk-like fibres. These differences in functionality are correlated to solution-state conformation, dynamics and pre-fibre self-assembly behaviour. Specifically, the disulphide-locked aciniform silk protein loses some α-helical character around the disulphide but forms a more compact unit as a whole, correlating with more

heterogeneous and larger pre-fibre particle self-assembly. The reduced state, conversely, has unchanged structuring relative to the wildtype but exhibits slightly elevated ps-ns-timescale dynamics around the mutation site that correlates with pre-fibre self-assembly into smaller than the wildtype nanoparticles with similar homogeneity. These studies on two different classes of protein demonstrate that disulphide bridges have the potential to amend both native structure and dynamics with clear functional consequences.

Chapter 1

INTRODUCTION

1.1. OVERVIEW OF DISULPHIDE BRIDGE

Among the 20 naturally occurring amino acids, cysteine (Cys) deserves special mention owing to the distinctive reactivity of the thiol (–SH) group. Even though methionine (***Fig. 1.1***.) also has a sulphur-containing side chain, only Cys residue is capable of thiol-based reactivity. While sequential oxidation of the –SH group may take place to produce sulfenic (–SOH), sulfinic (–SO_2H), or sulfonic (–SO_3H) acid derivatives of Cys, it is also susceptible to reaction with reactive oxygen, nitrogen and carbon species (ROS/RNS/RCS) formed during various biochemical processes. With different pKa values for its carboxyl (-COOH), amine (-NH_2) and thiol groups, Cys exhibits variable ionization states in response to changes in physiological pH. (***Fig. 1.2.***)

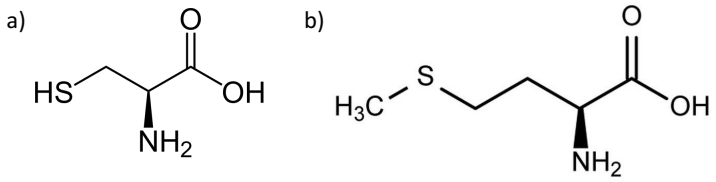


Fig. 1.1. *Structure of sulphur-containing amino acids a) cysteine and b) methionine.*

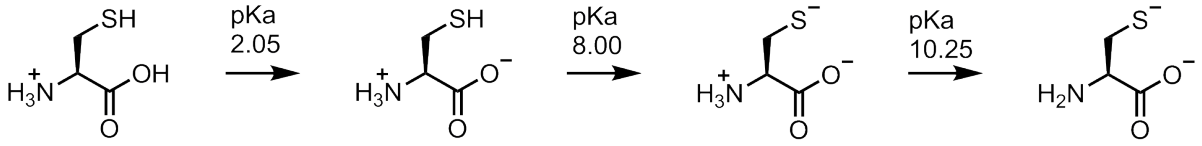


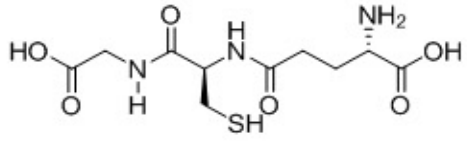
Fig. 1.2. *Ionization states of Cys residue in response to pH changes.*

Here, I am focusing on the oxidation reaction between two Cys thiol groups, resulting in the formation of a covalent bond known as disulphide bond, given by

$$R–SH + R'–SH \rightarrow R–S–S–R' + 2H^+ + 2e^- \quad (\mathbf{1.1}.)$$

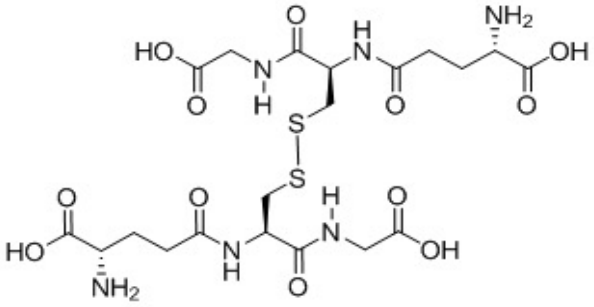
While reactions with low-molecular-weight thiols, like the tripeptide glutathione (GSH; ***Fig. 1.3.a)***) generates a mixed disulphide, covalent bonding with other Cys residue connects different regions of the same or different protein chains. The disulphide bond is also referred to as a disulphide lock, disulphide bridge or crosslink, while the individual Cys residues within the bridge are often referred to as cystine.

a)



GSH: Glutathione (reduced)

b)



GSSG: Glutathione (oxidized)

Fig. 1.3. *Chemical structure of glutathione in a) reduced and b) oxidized states.*

1.1.1. Properties of disulphide bridges

With a bond length of 2.05 Å, bond dissociation energy of 60 kcal·mol^{-1} and preferred dihedral angle (Cβ−Sγ−Sγ−Cβ), χ_{SS} of ± 90°, the disulphide bridge is considered a strong covalent bond in proteins or other biomolecules like polysaccharides, fatty acids etc.[11-15] However; the presence of different functional groups besides this bond renders the R-S- bond polarizable, making it susceptible to polar reagents with the consequential formation of the thiolate, R–S$^-$ ion and thus rendering a redox nature to this covalent bond and providing functionality as an internal pH sensor within the proteins.

With Cys residues reported at higher frequencies in complex organisms, disulphide-rich proteins have been identified and isolated from both plants and animals.[16-18] Venomous (detailed in *section*

1.2.) and fibrous structural proteins from animals, and antimicrobial peptides from plants (e.g., thionins) often rely upon disulphide bridges for their thermal stability and resistance to proteolysis.[19-22]

1.1.2. Formation of disulphide bridges

A cytosolic pH of 6.7-7.3 and the presence of the enzyme glutathione reductase ensures that a reduced form of GSH is present in high concentrations (rather than the oxidized form (GSSG; ***Fig. 1.3.b***)), and this is responsible for the reduced state of Cys in the cytosol in eukaryotes.[23-25] However, oxidative conditions in the rough endoplasmic reticulum of eukaryotic cells and the periplasmic space of prokaryotic cells favour disulphide formation during protein folding, thus a higher frequency of disulphide bridge(s) is reported in secretory, lysosomal and exoplasmic domains of membrane proteins.[26, 27] Notably, the thioredoxin enzyme localized in the endoplasmic reticulum is responsible for the reduction of disulphides in misfolded proteins.[28, 29]

Native disulphide bridging in a multiple Cys-containing reduced, unfolded protein depends on the spatial proximity, reactivity and accessibility of Cys groups and disulphide bonds.[30, 31] Proximity largely relies on chain entropy and non-covalent interactions favouring appropriate positioning of the relevant thiol groups (i.e., Cys residues) with respect to the protein backbone and thermodynamic stabilization of the unfolded state, which in turn is governed by the spatial proximity defined in terms of the effective intramolecular concentration of two thiol groups. The local electrostatic environment determines the potential for thiol-disulphide exchange reaction(s) and, hence, the reactivity of the -SH groups while the position of Cys residues is crucial in determining their accessibility to other Cys residues.[32] Following preliminary disulphide formation, fast reshuffling of non-native disulphides is mediated through the thiol-disulphide

exchange,[33] enabling native protein folding towards the stable tertiary structure under typical conditions.[31, 34] Hence, the native tertiary state of the protein is developed through a complex process termed as oxidative folding which involves native disulphide bridging through oxidation (i.e., disulphide formation), reduction (disulphide breaking) and isomerization/reshuffling (disulphide rearrangement), and conformational folding.[35, 36] (***Fig. 1.4.***)

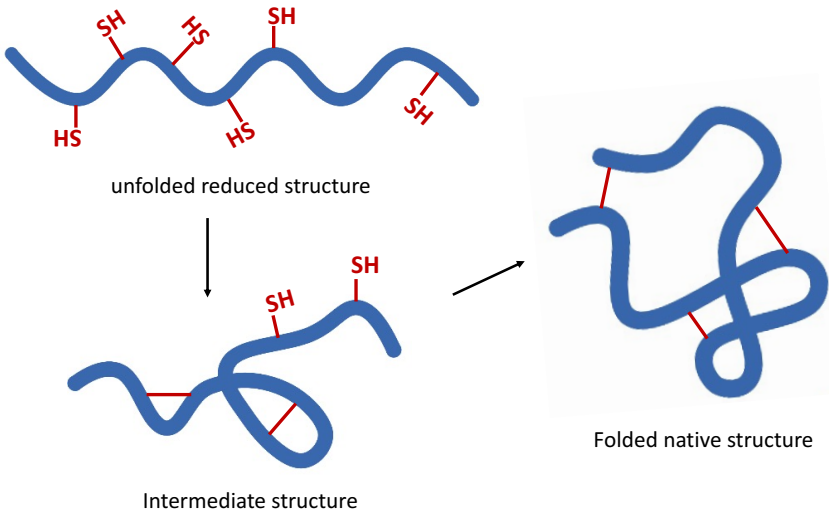


Fig. 1.4. *Schematic representation for oxidative folding of protein. Cys residues and disulphide bridges are highlighted in maroon.*

In the laboratory, an excess concentration of a thiol-containing reagent such as β-mercaptoethanol (BME) or dithiothreitol (DTT) is employed in a buffer to maintain the reduced state of Cys residues in proteins. While BME functions similarly to GSH, the two thiol groups in DTT create an intramolecular disulphide bridge, ensuring the reduced state of Cys residues within the protein. Alternatively, oxidative folding may be attained by incubating the protein with redox-active molecules like cysteine-containing peptides (GSH), dithiols (aromatic thiols), selenol etc. in suitable buffers.[37-41]

1.1.3. Role of disulphide bridges in protein architecture

Early theories of protein folding assumed that the native state corresponds to a well-defined structure while the unfolded protein exists as a random coil, suggesting a thermodynamic two-state protein folding mechanism.[42-44] Accordingly, disulphide bonds were considered to be key for structural stability, reducing the conformational freedom (i.e., entropy) of the unfolded state.[45] Based on this, rules for disulphide bridging were postulated such that the formation of these covalent bonds does not introduce straining effects in the protein structure.[46, 47] The Richardson-Thornton (R-T) rules[46, 47] state that disulphide bridging is forbidden between Cys pairs:

a) adjacent in the sequence in primary structure;
b) on adjacent β-strands;
c) in a single helix or strand; and,
d) on non-adjacent strands of the same β-sheet.

However, disulphide-bridged Cys residues in disagreement with R-T rules were later identified in protein structures.[48, 49] Disulphide bridges within β-sheets and β-hairpins distort these secondary structures, conveying additional constraints on the polypeptide chain, while Cys mutants of some proteins exhibit a decrease in transition temperatures of unfolding, i.e., reduced stability relative to wild type.[50] Additionally, disulphide bonds were observed to reinforce hydrophobic interactions causing local instabilities and favouring local conformational changes related to biological function, in accordance with the fuzzy oil drop model which advocates that hydrophobic core formation in proteins is driven more by their water environment than by pair-wise interactions operating in their local environment.[51-53] Observations of these classes of disulphide revealed that some disulphides may not provide a significant contribution to protein stability while some others could even destabilize the structure or disrupt the native function in the host protein.

Expanding the classical idea of protein folding beyond thermodynamics, currently, it is understood that proteins navigate through a funnel-like (i.e., multi-dimensional) energy landscape, such that loss of entropy is recompensed by enthalpy gained through hydrogen bonds, ionic and hydrophobic interactions etc. of the protein backbone and sidechains.[54-56] Disulphides within higher-ordered structures forbidden by R-T rules were discovered to be metastable, performing a redox role that facilitates functional processes like enzymatic activity, binding, catalysis etc., thus compensating for the violations of R-T rules.[57-59] Conformational changes associated with such disulphides are obvious in multifunctional metamorphic proteins, consistent with the structure-function paradigm which states that the function of a protein is determined by its structure, which is dependent on the sequence of the protein.[60] Correspondingly, 14 canonical subtypes of disulphide stereochemistry violating the concepts of R-T rules have been identified in protein structures up to now.[48, 49, 58] Disulphide bridges facilitate different levels of protein folding, higher-order structuring, interdomain/helical mobility and various conformational changes. From the perspective of structural biology, disulphide bridges serve as molecular 'safety pins' within the host protein rendering their functional role.[61 62]

1.1.4. Engineered disulphide bridges in protein

Given their versatile role when naturally occurring in proteins, disulphide bridges are frequently incorporated to aid as a biotechnological tool to introduce, improve and modify the functional character and protein stability for various biomedical and industrial applications.[63] Site-specific mutagenesis for disulphide incorporation may alter the thermodynamics of the native protein structures causing protein unfolding, aggregation, degradation, and/or loss of function.[59, 64] However, disturbances initiated by such disulphides can modify the rates of folding and/or

unfolding which can prove beneficial, with improvements in protein stability in response to biochemical factors like pH, temperature, etc. The addition of disulphides in regions involved in the preliminary stages of unfolding has been reported to substantially improve protein stability.[65] As an example, the scope of these engineered covalent bonds in designing novel biomaterials was demonstrated by Wu *et. al*. using a monomeric streptavidin-ligand conjugate designed with better immobilization capability and reversible interaction, applicable in biosensors and protein chips.[66] This is promising for improvements in affinity optimization, catalyst design, protein construct design, shorter functional proteins, protein-based therapeutics and industrial enzymes.

Apart from this, disulphide bridges at strategic positions may be applied to restrict intramolecular movement and domain motions, generating a locked or stapled conformation of the protein. The resulting alterations in local flexibility then facilitate functional alteration studies and/or characterization of protein structure, dynamics and conformational and functional changes accompanying protein folding, substrate binding, phase separation, etc.[63, 67-71] A relevant example is in the study of amyloid-β (Aβ) aggregation, a pathological indicator in Alzheimer's disease. It is hypothesized that conformational rearrangement of the soluble Aβ protein to an insoluble neurotoxic plague through the distinct pathway(s) involves the formation of higher-order structures like oligomers, globulomers, protofibrils, fibrils, etc. Disulphide mutants of the Aβ protein were employed to test the possible pathways of this progressive oligomerization to understand the structure-function relationship.[72-75]Apart from these classes of study, bioengineered disulphide(s) have been reported to impact enzyme kinetics, resulting in improvements in thermostability, decreases in pH-sensitivity and shifts in bioluminescence colour for a peroxisomal enzyme, luciferase.[76-78]

1.2. VENOM PROTEINS

One of the forthcoming chapters is a detailed discussion about the work I carried out on a mammalian venom-derived protein, soricidin. Before moving on to the experimental details of the project, it is important to understand the class of venom proteins.

Venoms isolated from noxious and predatory species identified in various phyla of the animal kingdom are often complex molecular cocktails.[79] Biochemical examinations have revealed that these venoms are predominantly composed of biologically active disulphide-rich proteins and polypeptides with broad functional roles. These frequently target ion (Na^+, K^+, Ca^{2+}, Cl^- etc.), mechanosensitive, and transient receptor potential (TRP) channels.[80-83]

Small proteins with a size of 2-6 kDa, constituting 20-60 amino acids, stabilized by multiple (≥3) disulphide bridges are typical for this class of proteins. However, larger proteins exhibiting diverse bioactivity have also been recovered from venoms in the snake family.[79, 84-87] Such proteins carry an extra loop, relatively long C-terminal tail and an additional disulphide bridge linking the N-terminus to the C-terminal tail to form an enclosed loop, thus generating a pseudo-cyclic structure. Studies conducted with the neuronal acetylcholine homo-pentameric receptor, α7, confirmed the indispensable role of this additional cyclic loop in providing high binding affinity to long-chain toxins, with short-chain and long-chain toxins lacking this structural feature exhibiting low affinity towards the neuronal receptor.[86, 88] Though most of these proteins are monomeric in nature, dimerization through covalent and/or non-covalent interactions with identical (homodimer) or non-identical (heterodimers) protomers has been reported for some snake-derived proteins. This enhances the exclusion of non-specific binding and provides superior pharmacological activity.[89-92]

Changes in the disulphide bridging pattern of venom proteins have been shown to alter folding patterns, leading to significant variations in the overall conformation and biological functions, in agreement with the structure-function paradigm.[60, 93] Alternate disulphide patterns and protein folds are considered metastable - trapped in a relatively stable but non-functional state and hence, physiologically irrelevant.[94] Comprehensive analyses have indicated that venoms with a similar fold can exert different functions and those with unrelated folds can exert similar functions too.[82, 95, 96]

Even though venom proteins are categorized into different families based on their disulphide bridging patterns and biological properties, a common functional dyad (also referred to as a pharmacophore), constituting a pair of basic and aromatic residues (Lys/ Arg, Tyr /Trp) at a distance of ~6-7 Å, is conserved across venom proteins targeting ion-channels.[82, 95, 97-105] Such convergence of a functional dyad besides their structural and functional divergence indicates a shared molecular basis in these phylogenetically unrelated animals, suggesting a common ancestral genetic origin.[82, 95, 96]

Even though general variations in chain length, the number of Cys residues, their arrangement within the sequence, and the pattern of disulphide bridging have been identified in venom proteins, their exceptional structural stability ensures functional integrity in changing chemical environments.[95] Together with this, the knowledge of structural and divergent evolution identified among the venom proteins from distant animal taxa being insightful towards exploring the potential of grafting novel functional sites through protein engineering expand their commercial potential. Hence identification, research and development of such proteins targeting a wide range of receptors with high potency, selectivity and *in vivo* stability are highly promising towards the

development of novel enzyme catalysts, bioinsecticides, therapeutic agents, and diagnostic and pharmacological tools.[71, 106, 107]

1.3. SPIDER SILK PROTEINS

Spider silk proteins (termed spidroins) are structural proteins constituting the strong, extensible, and/or tough silk strands that exhibit outstanding characteristics and are recognized to be promising biomaterials.[108] Unlike other arthropods, spiders are capable of producing a variety of silks through distinct specialized spinnerets, whose distinctive structure, physical and mechanical characteristics coupled with differences in their constituent amino acid content, domain structuring, and higher-order structuring enable different biological roles.[1, 109-112] (***Fig. 1.5.***) Most of these are employed in spider web architecture, with aciniform silk being engaged in prey wrapping and as the inner lining of the egg case while tubuliform silk is used to in egg casing.

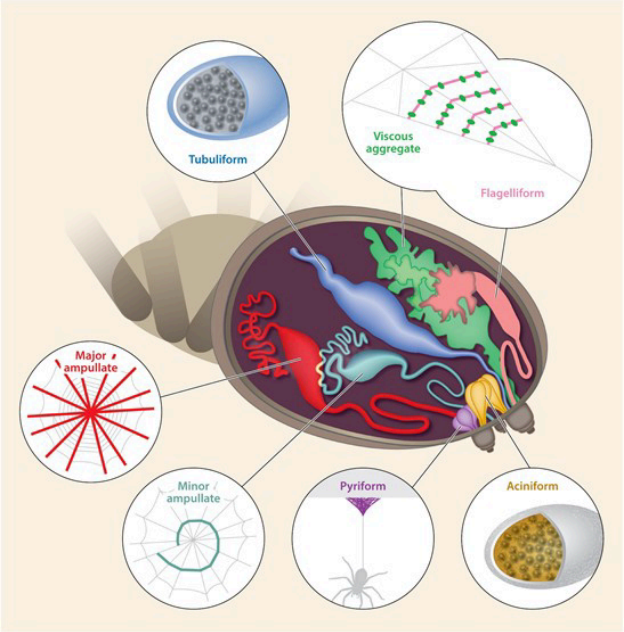


Fig. 1.5. *Anatomy of spider silk glands producing seven different types of silk proteins. Each spinning gland is depicted in the same colour as the corresponding silk type. Spinnerets are shown as brown protrusions from the abdomen. (Reproduced from Blamires et. al.*[1]*).*

In contrast with venom proteins, spidroins are large proteins (> 100 kDa) with repetitive sequence motifs linked end-to-end, adopting a multimeric state of elongated narrow strands of single filaments that form β-sheet secondary structures and coalesce into fibres.[108, 113, 114] These repetitive sequence motifs, distinct for each type of silk, are flanked by relatively shorter, highly conserved non-repetitive N- and C-terminal domains (CTD and NTD) with 50-200 residues. (***Fig. 1.6.***)

Fig. 1.6. *Schematic diagram depicting the general tripartite structure of the repetitive sequence motif in spider silk. Value of n usually decrease with increase in the length of the repeat unit.*

Like fibre forming proteins used in various animals (keratin, collagen, elastin, etc.) and like silkworm silk, spidroins also contain a high proportion of glycine, alanine and serine with small, non-reactive side chains facilitating H-bonding between the antiparallel polypeptide chains with an alternating C → N orientation to form β-sheets and ensuring close packing with minimal steric hindrance.[115-117] The amino acid sequence of the repetitive domain ultimately decides the secondary structural motifs (α-helix, β-sheet motifs and disulphide bridges) that can form, and these are primarily considered to be responsible for the mechanical characteristics of the fibre or other biomaterials derived from it.

Some silks are also associated with the phenomenon of supercontraction. This is defined as the irreversible conformational change and lengthwise-contraction of a fibre in response to the incorporation of water molecules into the structure (i.e., wetting).[118] Mobility of the protein (backbone and sidechains), supercontraction and extensibility of fibres when exposed to water is attributed to the proline content in major ampullate silk.[119]

While the secondary structure dictates local conformation, which is important for binding to other proteins as in venom proteins, the relative positioning of such motifs determines the tertiary architecture and aggregation of spidroins, following the structure-function paradigm.[120, 121] Extensive H-bonding within well-aligned β-sheets generates crystallinity that imparts exceptional strength, rigidity and water insolubility to spider silk fibres.[108, 121-123] Flexible disordered structures in the monomeric state and α-helices retained in the fibrous form generate amorphous regions permitting two-dimensional stretching of spider silk fibres, responsible for the extensibility and toughness, as is also the case in elastin fibres.[121, 124, 125]

The α-keratin proteins are mainly constituted by a coiled-coil heterodimer of two polypeptides, which are connected through regular intra-chain H-bonding whereas β-keratins in reptiles and birds are characterized by β-sheets twisted together, then stabilized and modified to a hardened state by disulphide bridges. Even though the secondary structure of monomeric spidroins is characterized as being α-helix-rich, as in α-keratins (hair, nails, horns, etc.), variations are evident at higher levels of ordering with spidroins are recognized to adopt β-sheets in the fibre state while α-keratins adopt a coiled-coil configuration. It should be noted, however, that β-sheets have been reported in α-keratins as well.[126] Owing to their repetitive amino acid sequence, flexible and elastic nature similar to other fibroin protein block copolymers like avian scales, feathers etc., spidroins are often classified as β-keratins.

A higher frequency of disulphide bridging confers higher thermal stability, water insolubility and structural rigidity to hard keratins (horns, nails, hooves and claws of mammals etc.), as in vulcanized rubber. In contrast, relatively soft, flexible and elastic structural proteins like hair (14% Cys in human hair), wool etc. are reported to contain a lower frequency of interchain disulphides.[108, 127-129] Although disulphide bridge-driven dimerization is reported for the C-

terminal domains of aciniform and ADF-3 silks, Cys residues are usually rare in spider silk proteins.[130, 131]

Spider silks are superior to almost all natural and synthetic materials, with the highest fibre strength of the ampullate and tubiliform silks being imparted by a higher density of well-aligned β-sheet motifs. The highest toughness among spider silks is reported for aciniform silk – the focus of my work, a silk which has a significant amount of α-helicity retained in the fibre structure. Finally, the highest extensibility is observed in flagelliform silks and being conferred by its predominately amorphous structuring.[132-134] Spider silks also exhibit exceptional thermal conductivity (416 $W \cdot m^{-1} \cdot K^{-1}$; silkworm silk: 0.5-6.5 $W \cdot m^{-1} \cdot K^{-1}$; copper: 401 $W \cdot m^{-1} \cdot K^{-1}$), a property that is largely attributed to the nature of the self-assembly process, which results in well-organized structures with minimal defects.[113, 135-138]

With the ability to form versatile materials like films, nanospheres, sponges with variable pore size, and hydrogels that could potentially serve as scaffoldings for cell attachment and growth and implanted tissue supports, the biocompatible nature of silks makes them amenable to biomedical and pharmaceutical applications.[139-144] The development of textiles, microfluidic surfaces and films, nanomaterials, optical fibres, chemical sensors, air filters, catalytic reaction chambers etc. with desirable mechanical and biocompatible properties also imply disparate potential commercial applications for this class of proteins.[144-148] Long-term degradation studies establishing the *in vitro* biodegradability of these protein-based materials is yet another aspect of spider silk that indicates promise for tissue engineering applications like medical sutures, tendon and ligament regeneration, etc.[149, 150] Amyloidogenic nature of spider silk proteins makes it a potential candidate as a model system to understand the aggregation patterns in clinically important fibrillogenic proteins like Aβ, Tau etc. involved in neurodegenerative diseases. Finally, the thermal conductivity properties

of silks are being explored, with one example being an aim to develop light-weight winter jacket fabrics which keep warm air close to the skin.[113] Given the combination of such remarkable mechanical properties, biocompatibility and biodegradability, it is important to identify the driving force(s) and processes involved in the hierarchical arrangement of spidroins atomic-level to allow the introduction of newly engineered biomaterials for desired purposes.

1.4. RECOMBINANT PROTEIN PRODUCTION

The cannibalistic and territorial nature of spiders poses a hindrance in rearing these organisms like silkworms for commercial applications allowing the collection of silks.[3, 151, 152] Also, spiders are reported to produce lower amounts of silk in captivity, thus making large-scale silk collection even more challenging.[153, 154] Owing to the poorly understood fibre formation process and non-feasibility of large-scale harvesting of native spider silk fibres, recombinant spidroin production using various host cells has been explored by numerous labs.[131, 155-158]The well-established bacterial expression and protein purification protocols of the Rainey lab for recombinant protein production using *Escherichia coli* as the host system were employed throughout the course of the work detailed in this thesis.[3, 9] The potential for higher yields and cost-effective protein production makes this a highly effective expression host for laboratory-scale protein production. An N-terminal hexahistidine tagged small ubiquitin-like modifier (H_6-SUMO) protein tag was used for the recombinant expression of all proteins discussed herein. This renowned fusion tag not only increases protein expression by improving the stability and solubility of the fusion protein[159] but also facilitates upfront immobilized metal affinity chromatographic (IMAC) purification, after which the protein-of-interest can be obtained by site-specific proteolytic cleavage using SUMO proteases.[3, 159] SUMO is entirely cleaved, with the added advantage compared to the proteolytic

removal of some other fusion-tags where there are no residual amino acids C-terminal to the cleavage site following proteolysis. (***Fig. 1.7.***).

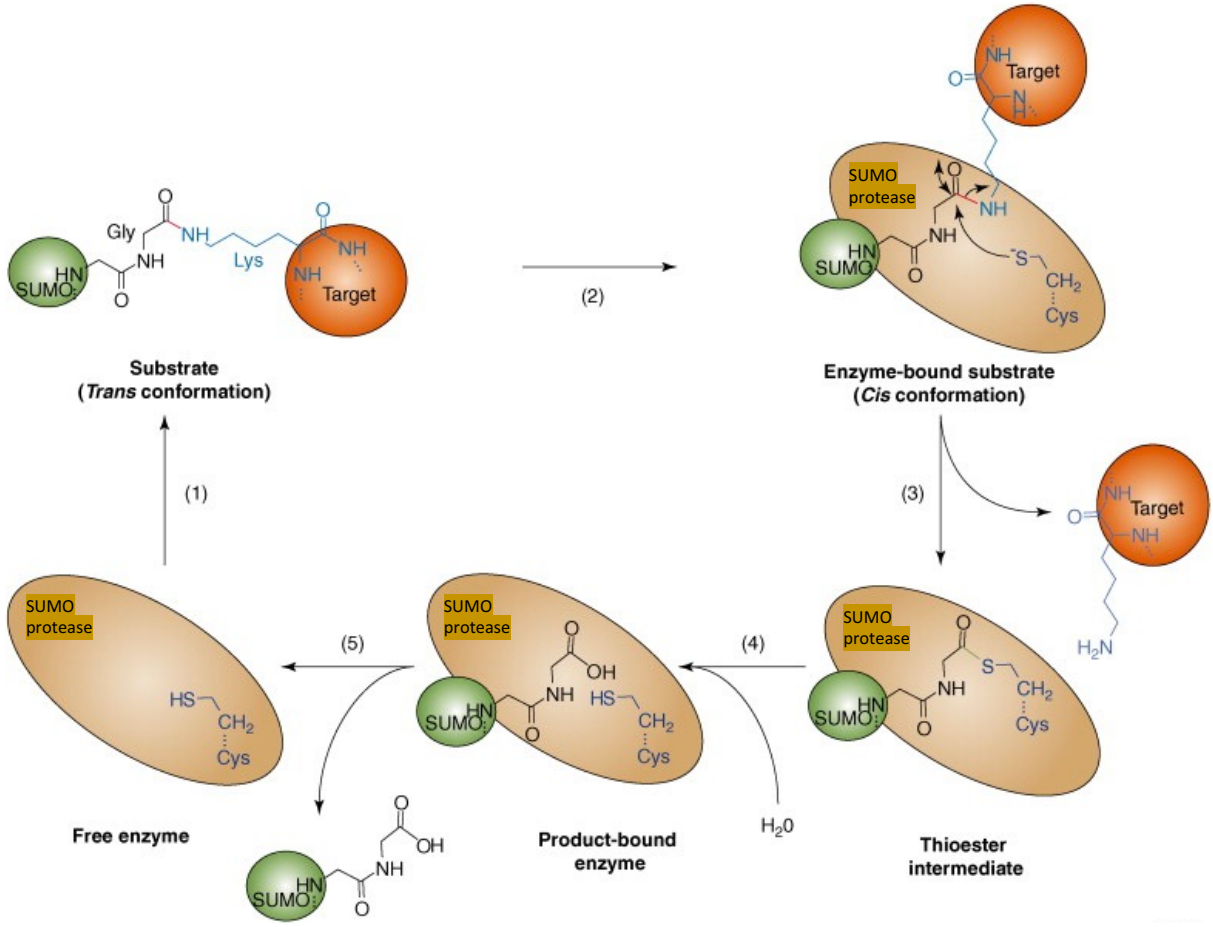


Fig. 1.7. *A schematic representation of SUMO deconjugation by SUMO protease from target protein. In the case of fusion protein, the Target is connected to SUMO moiety through a peptide bond instead of an isopeptide bond on a lysine side chain, as shown in the figure. Image adopted and modified from Mukhopadhyay, D., et. al.* [2]

1.5. PROJECTS

1.5.1. Project 1: Determine the structure of bioactive soricidin

Soricidin (SOR) is a protein secreted from the submaxillary glands of the northern short-tailed shrew (*Blarina brevicauda*), a mammal native to northeastern regions of North America.[160-162]

This voracious insectivore uses its saliva, containing SOR, as a prey-immobilizing agent for predatory purposes, making them one of the rare venomous mammalian species reported so far.[163-168]

This small protein contains 54 amino acids with six Cys residues, all of which are disulphide bridged (i.e. cystines) within the bioactive conformation (***Fig. 1.8.***) Studies conducted by *Soricimed Biopharma Inc.* identified a bifunctional nature for this protein. The binding of a relatively short C-terminal portionof soricidin, SOR-13, was identified to disrupt the function of TRPV6, a highly-selective calcium channel over-expressed in epithelial cancer cells, thus, inducing apoptosis in breast, ovarian, and prostate cancer.[162, 169-171] Interestingly, the N-terminal region is reported to inhibit Na^+ channels involved in neuronal signal transmission pathway(s).[172] Thus, soricidin is considered a prospective non-opioid candidate for neuropathic pain treatment, similar to analgesics derived from the disulphide bridged compounds isolated from venomous cone snails (conotoxins) etc.[173-179]

```
1           11          21
DCSQD CAACS ILARP AELNT ETCIL ECEGK
31          41          51
LSSND TEGGL CKEFL HPSKV DLPR
```

Fig. 1.7. *Amino acid sequence of soricidin with Cys residues highlighted in yellow. Prior to this work, the topology of disulphide-bonding remained unknown but all Cys were known to be in the disulphide-bridged state.*

Successful recombinant expression of soricidin protein in a bacterial system for the first time is reported in ***Chapter 2***. Development of an efficient refolding strategy to accelerate native disulphide bridging as in the bioactive form and to allow resolution of the tertiary structure of the full-length bifunctional protein is also reported. Here, I report the high-resolution structure of this venom protein for the first time, which also has allowed me to delineate the disulphide-bond pattern in the bioactive form delineated.

1.5.2. Project 2: Characterize the initiation of the aciniform spidroin fibrillogenesis pathway

Aciniform (or wrapping) silk, one of the seven types of natural silk spun by araneid spiders, has a tripartite structure comprising a long repetitive domain flanked by shorter non-repetitive NTD and CTD. Native wrapping silk, employed in prey wrapping and inner egg casing, is the toughest of all of the natural silk types and dramatically differs from any other silk in terms of its domain architecture and mechanical behaviour. A combination of high strength and elasticity makes aciniform the toughest of all spider silk varieties.[180] Unlike the short repetitive amino acid motifs (e.g., A_n, GX, GGX, $(GA)_n$) found in other types of silk, aciniform has a relatively large repetitive domain (typically ~200 amino acids) that is concatenated end-to-end.

Structure-mechanics studies conducted on other types of silk fibres, primarily major ampullate silk, have related strength[123] and extensibility[125, 181] exclusively to β-sheets and β-turns, respectively. Unlike other silk fibres, which mostly lack α-helical content, aciniform fibres have ~30% α-helical content alongside ~30% β-sheet and β-turn/disordered structure.[182] The functional role of the significant degree of α-helical content that is retained in the aciniform silk fibre is thus enigmatic.

Native aciniform silk has been minimally studied. The aciniform silk of *Argiope trifasciata* that we are studying is predominantly made up of aciniform spidroin (AcSp1) protein (~300-430 kDa) which is composed of ≥14 highly conserved 200-aa long iterated repeat units (termed W units, from wrapping silk) flanked by NTD and CTD regions.[110, 180, 183] The solution-state 3D structure of the W unit was previously reported by our group in 2015.[5]

Inspired by the use of site-specific Cys mutation and disulphide chemistry to probe protein structure and dynamics,[184] a bioengineered disulphide mutant (stapled form) of the W unit was

created (discussed in ***section 3.1***.; ***Fig. 1.9.)*** Comparative analyses of the stapled to the reduced mutant and wildtype (WT) W units were employed to test the effects of disulphide-locking upon 1) functionality concerning fibre forming ability and changes in fibre characteristics (***Chapter 3***); 2) the precursor state to fibre assembly in terms of the nature of protein pre-assembly in the dope solution state before wet-spinning (***Chapter 3***); and, 3) secondary structural variation, conformational changes and backbone dynamics of the soluble protein (***Chapters 4 and 5***).

```
1           11          21          31          41
AGPQGGFGAT GGASAGLISR VANALANTCT LRTVLRTGVS QQIASSVVQR
51           61           71          81          91
AAQSLASTLG VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL
101         111         121         131         141
NASNIDTLGS RVLSALLNGV SSAAQGLGIN VDSGSVQSDI SSCSSFLSTS
151         161         171         181         191
SSSASYSQAS ASSTSGAGYT GPSGPSTGPS GYPGPLGGGA PFGQSGFGGS
```

Fig. 1.9. *Amino acid sequence of the mutant W unit with Ser→ Cys mutation sites highlighted in yellow. The last Ser (highlighted in red) is removed from the last W unit for cloning purposes, hence W_2M has 399 residues while W_1M has 199 residues.*

The presence of thiol side chains, retained in the reduced form with DTT, was observed to affect protein packing, making the structure more compact, and hence, an improvement in mechanical properties was observed for the reduced mutant, while the fibre-forming ability was compromised in the stapled form of the mutant. Preliminary 2D NMR experiments with single W unit (W_1) and double unit (W_2) mutant proteins established the modular nature of the mutated unit as previously reported for the WT W unit.[5] With the modular nature of the Cys-containing W unit established, structural and dynamics characterization through solution-state NMR spectroscopy revealed variations between the native and mutant protein and, very pronouncedly, between the reduced and stabled forms of the mutant protein.

Chapter 2

DISULPHIDE TOPOLOGY AND STRUCTURAL DETERMINATION OF THE MAMMALIAN VENOM-DERIVED PROTEIN SORICIDIN

2.1. INTRODUCTION

During the development of SOR-13, containing the last 13 residues at the C-terminal tail of soricidin as an anti-cancer drug, the structural features of smaller derivatives of the protein were explored. However, neither the proper disulphide bridge pattern in the bioactive form nor the tertiary structure of the full-length protein had been resolved. The abundance of Cys residues in the N-terminal region of soricidin (***Fig. 1.5.***) leads to challenges to the synthetic production of the bioactive conformer of this mammalian protein with three disulphide bridges at levels of high purity. Namely, given the 6! possibilities in the number of distinct disulphide-containing versions of this protein (ranging from 1-3 disulphide bridges) and the slow kinetics of disulphide formation and reshuffling,[32] it was important first to develop an efficient protein refolding protocol allowing high-scale production of the bioactive soricidin protein.

Even though venomous compounds have been identified in various animals (as detailed in section ***1.2.***), venomous mammals are very rare.[164, 166-168] Soricidin, with its multiple disulphide bridges, thus presents an interesting situation. Hence, it is fundamental to understand the structure of this bifunctional protein to establish structure-function relationships for venomous disulphide-rich proteins produced in mammals and to understand how these are similar or distinct from those produced by other venomous animals.

2.1.1. Objectives

The work detailed in this chapter was carried out with aims to:

- recombinantly produce full-length soricidin using a bacterial expression system,
- develop a folding technique to accelerate native folding and maximize the yield of bioactive conformer; and,
- solve the structure of soricidin at the atomic-level to determine both its functional conformation and delineate the disulphide bridge pattern required for bioactivity.

Here, heteronuclear correlation-based solution-state nuclear magnetic resonance (NMR) spectroscopy was employed as the major tool to elucidate the topology and high-resolution structure of the disulphide-linked bioactive conformer of soricidin.

2.2. MATERIALS AND METHODS

2.2.1. Materials

A recombinant plasmid DNA encoding H_6-SUMO-soricidin was obtained from *BioBasic Inc.* (Markham, ON) and stored at -20 °C. The C41(DE3) *E. coli* strain was purchased from *Lucigen* (*Middleton, WI*) and stored at -80 °C. Isopropyl-β-*D*-thio-galactoside (IPTG), ampicillin, sodium dodecyl sulphate (SDS), acetonitrile (ACN) (high-performance liquid chromatography (HPLC) grade), and reagents for lysogeny broth (LB) medium were purchased from *Fisher Scientific* (*Ottawa, ON*). 15% Next gel solution was purchased from *Amresco* (*Solon, OH*). Dialysis tubing (3.5 kDa and 14 kDa cut-offs) was purchased from *BioDesign Inc.* (*Carmel Hamlet, NY*). Uniformly ^{15}N-labelled ammonium chloride ($^{15}NH_4Cl$), uniformly ^{13}C-labelled *D*-glucose and deuterium oxide (D_2O; all isotopes at 99% enrichment) were purchased from *Cambridge Isotope Laboratories* (*Tewksbury, MA*). Ni-NTA resins were purchased from *Takara Bio USA Inc.*

(*Mountain View, CA*). 0.45 μm syringe filters were obtained from *Millipore (Billerica, Massachusetts*) and 15 mL *Amicon Ultra* Centrifugal Filter units were purchased from *Fisher Scientific* (*Ottawa, ON*). SUMO protease was produced *in-house* using an expression plasmid obtained as a gift from Dr. Paul Liu (*Department of Biochemistry & Molecular Biology, Dalhousie University*). All other chemicals were obtained from *Sigma-Aldrich* (*Oakville, ON*) unless otherwise specified. The compositions of buffers employed at various purification stages are tabulated in the concerned sections.

2.2.2. Recombinant Protein Production and Purification

2.2.2.1. Bacterial expression

Recombinant plasmid DNA (0.5 μL) encoding an N-terminally-fused hexahistidine-tag and small ubiquitin-like modifier protein (H_6-SUMO) soricidin (***Fig. 2.1.***) was transformed into C41(DE3) *E. coli* cells (200 μL) through electroporation. After the addition of 500 μL LB media, *E. coli* cells were incubated at 37 °C for 30 min. 200 μL of this culture sample was plated on an agar plate and allowed to grow at 37 °C for 16 h. An isolated single colony of *E. coli* was selected from the agar plate, transferred to 3 mL LB media (***Table 2.1.)*** supplemented with ampicillin (Amp; 50 $\mu g \cdot mL^{1}$) at 37 °C and allowed to grow until an optical density at 600 nm (OD_{600}) of ~0.8 was attained to perform test expression. Cell cultures were then supplemented with IPTG (0.8 mM) to induce expression from the T7 promoter and allowed to grow at 37 °C after induction for 3 h. Once over-expression was achieved/confirmed, large-scale expression was initiated with a 20 mL starter culture of *E. coli* cells carrying the expression plasmid grown in LB media with ampicillin (50 $\mu g \cdot mL^{-1}$) at 37 °C for 12 h. 10 mL of this starter culture was used to inoculate 1 L of LB media with ampicillin (50 $\mu g \cdot mL^{-1}$) and cells were allowed to grow at 37 °C. Once an OD_{600} of ~0.8 was

attained, 0.8 mM IPTG was added to the culture medium to induce expression and the cells were further allowed to grow at 30 °C overnight (O/N).

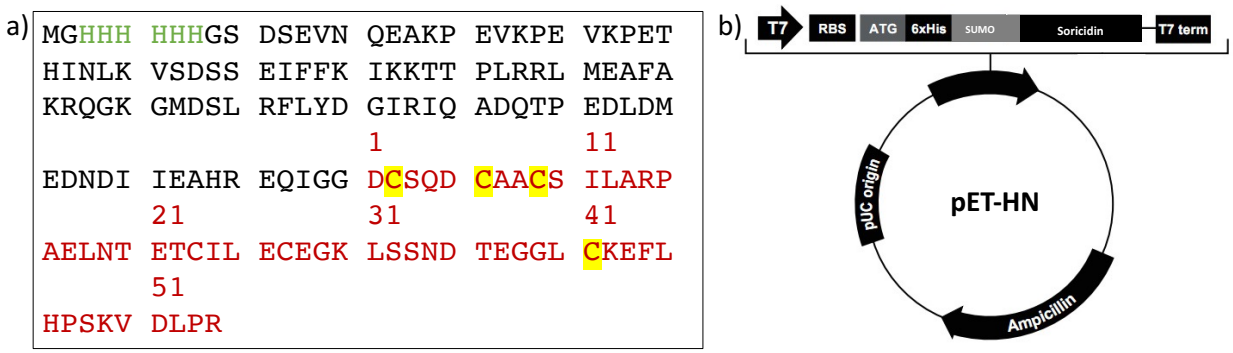


***Fig. 2.1**. a) Amino acid sequence for H_6-SUMO-soricidin protein. The fusion protein sequence consists of an N-terminal hexaHis (green) followed by SUMO (black) tags followed by soricidin (red). Cysteine residues are highlighted in yellow. b) Recombinant plasmid DNA encoding H_6-SUMO-soricidin.*

2.2.2.2. ^{15}N and ^{13}C-isotope labelling for NMR experiments

For isotope-labelled protein production, cells were spun down (20 min/4 °C/6500 rcf) after attaining an OD_{600} ~ 0.8, resuspended and transferred to a ½-equivalent volume of M9 minimal medium (***Table 2.1.***) containing $^{15}NH_4SO_4$ (1 g·L^{-1}) and ^{13}C-*D*-glucose (2 g·L^{-1}) and supplemented with 4 mM magnesium sulphate ($MgSO_4$) and 1.8 µM ferric chloride ($FeCl_3$) to achieve ^{15}N and ^{13}C isotope enrichment.[185, 186] Cells were then grown for 30 min at 37 °C before inducing with 0.8 mM IPTG and grown O/N at 30 °C. From here onwards, the labelled protein will be referred to as [U-^{13}C,^{15}N] soricidin.

2.2.2.3. Protein purification

Following expression, cells were harvested through centrifugation (6500 rcf /20 min/4 °C) and resuspended in ice-cold lysis buffer to facilitate cell lysis (3×) through a French Pressure Cell Press

(*American Instrument Company Silver Spring, MD*). The insoluble fraction obtained after centrifugation of the cell lysate (13000 rcf /40 min/4 °C) was discarded while the fusion protein was purified from the supernatant through affinity chromatography using a column packed with Ni-NTA agarose beads were pre-equilibrated with ice-cold lysis buffer. The column was then washed with native wash buffer. Finally, bound proteins were eluted with elution buffer and pooled together. Cleavage of the H_6-SUMO tag was achieved using 50 µM SUMO protease (produced *in-house*) in the presence of 1 M DTT in 50 mM PB at pH 8 ± 0.05. Imidazole and other salts were removed through O/N dialysis against distilled water at 4 °C. All buffers used in the affinity chromatography purification are detailed in ***Table 2.2***., following which the protein was lyophilized following shell freezing by submersion in liquid N_2 and stored at -20 °C until use.

Table 2.1.: GROWTH MEDIUM FOR BACTERIAL CULTURE

Name	**Compsition**	**Concentration**
Lysogeny broth (LB) medium	LB broth Ampicillin IPTG	20.0 $g \cdot L^{-1}$ 50 $\mu g \cdot mL^{-1}$ 0.8 mM
Minimal (M9) medium	$^{15}NH_4SO_4$ ^{13}C-*D*-glucose $MgSO_4$ $FeCl_3$ Ampicillin IPTG	1.0 $g \cdot L^{-1}$ 2.0 $g \cdot L^{-1}$ 4.0 mM 1.8 µM 50 $\mu g \cdot mL^{-1}$ 0.8 mM

pH adjusted to 7.4 using NaOH.

Table 2.2.: BUFFERS USED IN Ni-NTA CHROMATOGRAPHY

#	**Buffer Nomenclature**	**Concentration (mM)**		
		NaH_2PO_4	**NaCl**	**Imidazole**
1.	Lysis	50	300	10
2.	Wash	50	300	20
3.	Elution	50	300	250

pH adjusted to 8.0 using NaOH.

Lyophilized samples were redissolved in deionized water and purified at a semi-preparative scale using a C-18 column (5 µm bead size, 10 mm × 250 mm; *W.R. Grace and Company*; *Columbia, MD*) on a ProStar HPLC (*Varian Canada Inc.*; *Mississauga, ON*) and binary solvent system

(solvent A: 0.1% trifluoroacetic acid (TFA) in H_2O; solvent B: 0.1% TFA in acetonitrile (ACN); flow rate 5 mL·min^{-1}; mobile phase gradient: 2-20% solvent B from 0-5 min, 20-45% B from 5-30 min and a linear gradient of 2% min^{-1} B from 30-31 min). The eluent peak corresponding to the fusion-tag-free protein was collected, pooled, concentrated, flash-frozen in liquid N_2, lyophilized, and stored at -20 °C until use.

At every instance, the protein concentration in the sample was determined according to the Beer-Lambert law:

$$A = \varepsilon C l \qquad \textbf{(2.1.)}$$

where A is the absorbance, ε is molar extinction coefficient[187] (M^{-1}·cm^{-1}), C is protein concentration (M) and l is the pathlength of the cell (cm). Samples were diluted (10×) with 6 M guanidinium chloride and absorbance values were determined at 210 nm (*Olis HP 8452 Diode Array UV spectrophotometer; Bogart, GA*). The molar extinction coefficient[187] at 214 nm, (ε_{214} = 70, 992 M^{-1} cm^{-1}) was estimated using the Prot pi website server.

(https://www.protpi.ch/Calculator/ProteinTool)

2.2.2.4. Oxidative refolding

An oxidative protein refolding methodology based on the concept of thiol-disulphide exchange,[33,] (detailed in *section **1.1.2.***) was employed to accelerate disulphide bond shuffling in the purified soricidin. Small-scale trials were performed using glutathione (GSH), its oxidized counterpart (GSSG) and their mixtures added in different ratios - 1:1, 2:1, 3:1 and 1:2 – to the lyophilized HPLC-purified protein fraction (0.5 mg·mL^{-1}) redissolved in deionized water and shaken to facilitate the refolding process O/N at RT. Analytical HPLC purification runs were carried out following the same methodology as in section ***2.2.3.*** with the exception that a narrower diameter

C-18 column (5 µm, i. d. 5 mm × 250 mm); *AAPPTec; Louisville, KY*) and lower flow rate (flow rate 3 $mL \cdot min^{-1}$) were employed to test the folded-state status of the protein as a function of time. Two major eluents were targeted based on an RP-HPLC elution profile provided by *Soricimed Biopharma Inc.* for the soricidin protein produced through solid-phase peptide synthesis and refolding using their *in-house* protocol. Namely, one major species and a second minor species were separately collected at two different elution timepoints, lyophilized and analyzed by electrospray ionization mass spectrometry (ESI^+ MS)[188] at the Mass Spectrometry Facility, Department of Chemistry, Dalhousie University, Halifax, NS.

2.2.2.5. Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE)

Protein samples for SDS-PAGE were prepared by mixing an equal volume of the sample and 2× SDS-PAGE loading buffer either using non-reducing (i.e., without β-mercaptoethanol) for protein samples after SUMO tag removal or reducing buffers for all the other samples. Samples were then incubated in a hot water bath (~95 °C) for 5 min and centrifuged (14,000 rcf/RT/5 min) to pellet out insoluble debris. SDS-PAGE gels were prepared *in-house* using 12.5% Next gel solutions with the supernatant following centrifugation loaded. Samples were typically run at 190 V for 50 min in a *Mini-Protean II cell* (*Bio-Rad Laboratories Inc.; Mississauga, ON*) using a *Power Station 200 power supply* (*Labnet International Inc.; Edison, NJ*). Gel soaked in staining solution was microwaved for 30 s and allowed to stain for 5 min. Further, the gel was washed and soaked in the de-staining solution and framed for visualization.

2.2.2.6. Bioactivity testing on mealworms

The functionality of both the protein fractions collected through RP-HPLC purification following oxidative refolding was tested on adult *Tenebrio molitor* mealworms (purchased from a local pet

shop) following a protocol provided by *Soricimed Biopharma Inc.* Briefly, each protein fraction was solubilized in *Insect Ringers'* saline solution (***Table 2.3***.). Individual worms were weighed and injection dosage was determined according to their body weight (1 µg· mg^{-1} of body weight). The injection was dorsally administered at the fourth segment from the head just under the integument. Two groups, each constituting eight mealworms, were injected with blank saline solution as a control. The protein solutions, prepared as per the given protocol, were injected into similar batches of worms (2 groups of 4 each) at different dosages (2×, 3×) to test the functional efficacy and effect of the higher dosage.

Table 2.3.: INSECT RINGERS' SALINE SOLUTION		
#	**Composition**	**[Stock]**
1.	Krebs-ringer solution (1 L)	172 mM
	a) Sodium chloride (NaCl) b) Potassium chloride (KCl) c) Sodium phosphate (NaH_2PO_4) d) Sodium bicarbonate ($NaHCO_3$)	
2. 3. 4.	Calcium chloride ($CaCl_2$) Magnesium sulphate ($MgSO_4$) Dextrose in Deionized water	2 mM 1 mM 10 mM

2.2.2.7. Far-UV- CD spectroscopy

20 µM protein samples were prepared by dissolving refolded soricidin in 50 mM sodium phosphate buffer (pH 7.00 ± 0.05) with concentration determined using absorbance at 214 nm (molar extinction coefficient,[187] ε_{214} = 70,992 M^{-1} cm^{-1}; ***Eq. 2.1.***). 10 mM TCEP was added to one of the protein samples and incubated at 65 °C for 30 min in a water bath to reduce disulphide bonds.

Far UV-CD spectra were recorded at 30 °C in 0.01 cm quartz cuvettes (*Hellma*; *Vaughan, ON*) using a J-810 spectropolarimeter (*Jasco*; *Easton, MD*) over the range of 260-185 nm at 100 nm·min^{-1} with a data pitch of 0.1 nm and three accumulations. Measurements were repeated with three separately prepared batches. The collected data were averaged over all accumulations, blank-

subtracted, and converted to mean residue ellipticity (MRE; °·cm^2·dmol^{-1}·residue^{-1}) according to the following equation,

$$[\theta] = \frac{A * M * 3298}{l * C} \quad \textbf{(2.2.)}$$

where A is the absorbance at 214 nm, M is the molecular weight, *l* is the pathlength of the cell used for measurements and *C* is the protein concentration. The secondary structural features of each state – refolded and selectively reduced for disulphides- were inferred through comparison to canonical protein CD spectra.[189, 190]

2.2.2.8. Solution-state Nuclear magnetic resonance (NMR) spectroscopy

2.2.2.8.1. NMR sample preparation

Lyophilized [U-^{13}C,^{15}N] soricidin was dissolved in 20 mM phosphate buffer (95/5 (v/v) H_2O/D_2O, 1 mM 2,2-dimethyl-2-sila-pentane-5-sulfonic acid (DSS), 1 mM sodium azide (NaN_3); pH 6.00 ± 0.05 (without correction for effects of deuterium)). The protein suspension was vortexed (2 min) and centrifuged (13000 rcf/5 min/10 °C) to remove aggregates if any. The supernatant containing soluble protein was decanted, filtered through a 0.45 μm syringe filter and concentrated using 15 mL *Amicon* centrifugal filter unit (*3,000 NMWL; Fisher Scientific; Ottawa, ON*). The final soricidin concentration was determined to be 0.2 mM using absorption spectroscopy at 210 nm likewise as for CD spectroscopy. The sample was transferred into a 5 mm Shigemi susceptibility-matched NMR tube shaped for cryoprobe usage (*Bruker Biospin; Fällanden, Switzerland*).

2.2.2.8.2. Data acquisition

NMR experiments were acquired either using an Avance III 16.4 T spectrometer equipped with a 5 mm triple resonance inverse (TCI) cryoprobe (*Bruker Biospin; Billerica, MA*) at the *National*

Research Council, Halifax, Canada or an Ascend 18.8 T spectrometer equipped with a TCI cryoprobe (*Bruker BioSpin*) at the *QANUC facility, Montréal, QC.*

2.2.2.8.3. Variable-temperature NMR

Variable temperature NMR (VT-NMR) experiments are often applied for better dispersion and resolution of overlapped resonances.[191, 192] Apart from improving spectral quality, VT-NMR studies are often employed in protein studies to infer crucial hints about intramolecular hydrogen bonding in various secondary structural motifs, free energy lanscapes, protein folding/unfolding, aggregation kinetics and to identify phenomena like conformational exchange.[193-195] This is often executed through a series of ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) experiments collected at different temperatures. Combined chemical shift displacements (CCSD) of the amide proton, H^N, and amide nitrogen, N, may then be determined as a function of temperature and the dependence of amide chemical shifts on temperature is estimated by extrapolating the slope of chemical shift vs. temperature based upon a linear fit to calculate a linear temperature coefficient. The linear temperature coefficient for a given H^N is an indication of the extent of intramolecular hydrogen bonding of that individual amide.[194, 196] This can, in turn, suggest temperature-dependent secondary/teritiary structural changes and, hence, is useful to identify extend of H-bonding of residues within the ordered vs. disordered regions of the protein and further thermal expansion upon decompaction.[195, 197, 198] Interestingly, any deviation from linearity is often interpreted as thermally-driven conformational exchange being operational – i.e. fast exchange between disticnt conformers of the protein. [198]

2D ^{1}H-^{15}N HSQC spectra were acquired at a range of temperatures (5 °C, 15 °C, 25 °C and 37 °C) to probe for any temperature-dependent conformational changes, if any. Following this, all

subsequent NMR experiments for structural and dynamics characterization were acquired at 37 °C as this mimics the human physiological conditions where soricidin may be applied in future as a bioactive species.

2.2.2.8.4. 3D experiments for backbone and sidechain assignment

The following 3D experiments were performed for backbone assignment alongside ^{1}H-^{15}N HSQC spectra: HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCANH,[199, 200, 201] HNHA[202], HNHB[203] and ^{15}N-edited NOESY-HSQC.[204, 205] In addition to the latter four experiments, side-chain chemical shifts were assigned using 3D HCCH-TOCSY and HCCH-COSY,[199, 206] ^{13}C-edited TOCSY-HSQC and ^{1}H-^{13}C HSQC-NOESY[207] experiments. Arg and Gln side-chain assignments were determined using both 2D ^{1}H-^{15}N HSQC and the ^{15}N-edited NOESY-HSQC experiments.

2.2.2.8.5. NOESY Experiments

Magnetization transfer mediated through dipolar interactions between two spatially proximal nuclei (internuclear distance ≤ 6 Å) gives rise to the nuclear Overhauser enhancement (NOE), with the intensity of this enhancement reflecting the rate of cross-relaxation (σ_{IS}). This is, in turn, proportional to the inverse sixth power of the distance between the pair of interacting nuclei (namely, I and S). In a 2D or 3D NOESY experiment, the resulting cross-peaks, thus, provide distance restraints to establish spatial contacts and further determine the 3D contacts used to define the tertiary fold of the protein.[208] 2D ^{1}H-^{1}H nuclear Overhausser enhancement spectroscopy (NOESY; ^{13}C- and ^{15}N-decoupled) and 3D ^{13}C-edited and ^{15}N-edited NOESY-HSQC experiments (mixing time: 200 ms) were acquired at 18.8 T for NOE peak assignment.

2.2.2.8.6. Steady-state heteronuclear-NOE

In a system containing two types of nuclei, *I* and *S,* when spin *S* is irradiated, thermal equilibrium can be re-established by transferring the magnetization to a nearby spin *I* through dipole-dipole interactions.[209] This cross-relaxation process between two dipolar-coupled spins follows the NOE. The change in peak intensity observed for the *I* spin defines the extent of dipole-dipole interactions operating within the system and is referred to as the NOE enhancement factor, η.

The [^{1}H]-^{15}N steady-state heteronuclear-NOE NMR experiment involves the acquisition of two different spectra: one recorded with and one without proton presaturation. Unlike the situation above, here an underlying assumption is made that all ^{1}H-^{15}N spin pairs have the same internuclear distance. In this instance, molecular motion is reflected in any observed differences in NOE. A sufficiently long RF pulse (weak enough not to perturb ^{15}N spin) is used to saturate proton nuclei i.e., achieving a state with a population difference of 0. Dipole-dipole interactions of ^{1}H and ^{15}N would initiate cross-relaxation that eventually leads to the evolution of ^{15}N magnetization to a steady-state. Experimentally, the NOE effect is expressed as the ratio of peak intensities (with and without ^{1}H saturation) given by,[210, 211]

$$NOE = \frac{\mathrm{I}}{\mathrm{I}_0} \tag{2.3.}$$

while enhancement factor, η is given by

$$\eta = \frac{\mathrm{I} - \mathrm{I}_0}{\mathrm{I}_0} \tag{2.4.}$$

where I and I_0 are the peak intensities observed in the spectrum with and without ^{1}H saturation, respectively.

A [^{1}H]-^{15}N steady-state heteronuclear-NOE experiment was measured at 16.4 T with saturation during the recycle delay (5 s) carried out in an interleaved manner for the recording of NOE enhanced (with saturation applied during final 4 s of the recycle delay period) vs. unenhanced (without saturation) signal intensity.

2.2.2.9. Spectral processing

All spectra were processed using NMRpipe[212] and analyzed using CcpNmr Analysis 2.2.1.[213] ^{1}H frequencies were referenced to DSS at 0 ppm and ^{13}C and ^{15}N were indirectly referenced to the ^{1}H zero-point DSS frequency.[214] Backbone and side-chain resonances were sequentially assigned using 2D and 3D heteronuclear correlation NMR experiments (detailed in ***Table 2.4.)*** through the standard backbone walk methodology.[215]

Table 2.4.: EXPERIMENTAL PARAMETERS FOF 2D/3D- NMR EXPERIMENTS

Experiment *pulse sequence*	**Recycle delay** (s)	**# scans**	**% NUS; NUS points**	**Dimension**	**# complex points**	**Sweep width** (ppm)
^{1}H-^{15}N HSQC *fhsqcf3gpph*	1	4	50; 50	^{1}H ^{15}N	2048 200	16.02 24.00
HNCO *hncogpwg3d*	1	16	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.02 24.00 11.00
HNCACO *hncacogpwg3d*	1	64	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.02 24.00 11.00
HNCA *hncagpwg3d*	1	16	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.02 24.00 30.00
HNCOCA *hncocagpwg3d*	1	16	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.02 24.00 30.00
CBCACONH *cbcaconhgpwg3d*	1	16	12.5; 613	^{1}H ^{15}N ^{13}C	2048 90 218	16.02 24.00 80.00
HNCACB *hncacbgpwg3d*	1	16	12.5; 720	^{1}H ^{15}N	2048 90	16.02 24.00

				^{13}C	256	80.00
het-NOE *hsqcnoef3gpsi*	1	32	0	^{1}H ^{15}N	2048 90	16.02 24.00
CC[CO]NH *hccconhgp3d3*	1	4	25; 192	^{1}H ^{15}N ^{13}C	2048 90 128	16.08 24.50 75.00
HCCH_COSY *hcchcogp3d*	1	4	25; 384	^{1}H ^{15}N ^{13}C	2048 48 128	14.00 35.00 06.00
HCCH *hcchdigp3d*	1	4	25; 384	^{1}H ^{15}N ^{13}C	2048 48 128	14.00 35.00 06.00
DIPSI_TOCSY *dipsihsqcf3gpsi3d*	1	4	25; 256	^{1}H ^{15}N ^{13}C	2048 32 128	106.08 24.50 10.00
C_NOESY *noesyhsqcetgp3d*	1	16	25; 640	^{1}H ^{15}N ^{13}C	2048 30 160	16.02 30.00 09.50
3D_NOESY *noesyhsqcf3gpwg3d*	1	16	25; 500	^{1}H ^{15}N ^{13}C	2048 50 160	16.02 25.00 10.00
2D-NOESY *noesyesfbgpph*	1	96	50; 190	^{1}H ^{15}N	2048 760	16.02 9.50
H[CCCCO]NH *hccconhgp3d2*	1	4	25; 256	^{1}H ^{15}N ^{13}C	2048 32 128	16.08 24.50 06.00

2.2.2.10. Data analysis and validation

2.2.2.10.1. Heteronuclear-NOE

NOE for each residue was calculated as in ***Eq. 2.3.*** using CcpNmr Analysis 2.4.2.[213] and reported with error (standard deviation) estimated[210] as described by *Farrow et. al.*[216]:

$$\frac{\sigma_{NOE}}{NOE} = \sqrt{\left(\frac{\sigma_I}{I}\right)^2 + \left(\frac{\sigma_{I_0}}{I_0}\right)^2} \quad (\mathbf{2.5.})$$

where σ_I and σ_{I_0} are the respective values of root-mean-square background noise for each spectrum. A positive enhancement of typically >0.65 was interpreted as signifying order (rigidity)

while smaller or negative enhancements were interpreted to be indicative of regions with elevated backbone disorder (dynamic nature).[215]

Combined CSD values ($\overline{d_{NH}}$) between two temperature conditions were determined, following the approach of Williamson, (2013):

$$\overline{d_{NH}} = \frac{\sqrt{(\delta_H)^2 + 0.2(\delta_N)^2}}{2} \quad \textbf{(2.6.)}$$

with, $(\delta_H) = \delta H_{278} - \delta H_T$

and $(\delta_N) = \delta N_{278} - \delta N_T$

where δH_T and δN_T are the chemical shift values of amide protons and nitrogens (in ppm), respectively at temperature T (in Kelvin) while the corresponding values at 278 K are represented with a subscript. The quality of backbone chemical shift assignments was validated using the PANAV (Probabilistic Approach to NMR Assignment and Validation) webserver (http://panav.wishartlab.com/) that employs the algorithm developed by Wang *et. al.* (2010).[217]

2.2.2.10.2. Secondary structure

Assigned chemical shift values are often used to locate and identify the type of local secondary structure within a protein. [218,219] This approach evaluates the observed chemical shift (CS) of an atom relative to the random coil value, in terms of a secondary chemical shift,

$$\Delta\delta = \delta_{assigned} - \delta_{RC} \quad \textbf{(2.7.)}$$

where $\delta_{assigned}$ and δ_{RC} are the assigned and the corresponding random coil CS values, respectively for atom X (= backbone ^{15}N, ^{1}H (H^N and Hα) and ^{13}C (Cα, Cβ and C'). Depending on the magnitude and sign of the secondary CS value, each residue is assigned a chemical shift index (CSI), defined as,

$$CSI = \begin{cases} \pm 1, & \Delta\delta \geq 0.1\ ppm \\ 0, & \Delta\delta < 0.1\ ppm \end{cases} \quad \textbf{(2.8.)}$$

i.e., residues with upfield shift >0.1 ppm are assigned +1 while a corresponding downfield shift is assigned -1 and all other values in between this regime are assigned 0. A bar graph representation of this 3-state index plotted over residue number is used to identify the type of secondary structures: consecutive residues with -1 values are identified as α-helices, +1 values as β-strands and 0 values as random coil segments. A consensus assignment obtained by combining assignments of multiple chemical shifts (^{1}Hα, ^{13}Cα, ^{13}Cβ and ^{13}C′) improve the accuracy to >90% for this prediction.[220]

DANGLE (Dihedral ANgles from Global Likelihood Estimates) is an algorithm employing Bayesian inference to predict the backbone torsional angles, phi (φ) and psi (ϕ) of all the residues within the protein. This algorithm works based on the chemical shift assignments and the conformational preferences of each amino acid. Hence, this is used for predicting the secondary structural features of a protein.[221]

2.2.2.11. Experimental restraint refinement and structural calculation

Dihedral angles for structural calculations were predicted *in silico* by the DANGLE[221] algorithm, directly integrated into the CcpNMR Analysis assignment program.[213] Bayesian inference based on the assigned backbone chemical shifts and amino acid conformational preference is used to estimate the backbone φ and ψ dihedral angles for each residue. Disulphide bonds and these restraints are directly employed as J-coupling restraints structure calculation protocols in ARIA (***Fig. 2.2.)***.

^{1}H-^{1}H NOE-derived distance restraints were determined using the ^{1}H-^{1}H NOSEY, ^{13}C-edited NOESY-HSQC, and ^{15}N-edited NOESY-HSQC spectra assignments within CcpNmr Analysis. [213,

[222] Since the S/N ratio was generally weak for the protein, the assignment was initiated for the most intense and less overlapped peaks and peak intensity was exported in terms of heights, (and not volume) to avoid arising from peak overlap.

The restraints were exported from Analysis into CNS format and interactively refined using *ARIA version 2.3*.[4] Even though one NOE peak ideally corresponds to one distance pair, an observed peak could be an outcome of more than one overlapping peak in a real setting. To account for such instances, an ambiguous distance restraint (ADR) list was generated.

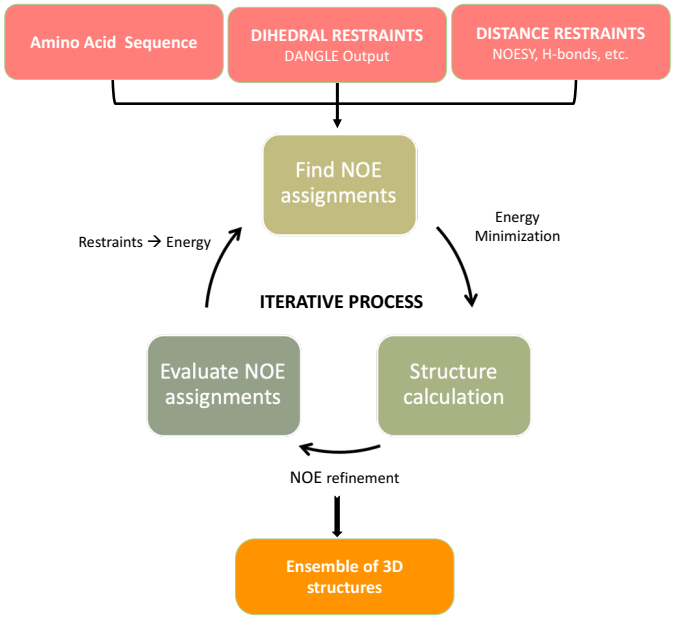


Fig. 2.2. *Schematic representation of steps involved in tertiary structure calculation.*

The list is produced with peak intensity defined as a cumulative sum of all possible inter-nuclear distances, given by,

$$\bar{d} = \left[\sum_{k=1}^{N_k} \frac{1}{d_k^6}\right]^{-6} \qquad \textbf{(2.9.)}$$

such that $\bar{d}$ is the effective distance, N_k the number of possible assignments, and d_k is the interatomic distance between two atoms corresponding to the kth contributions.[223] Often, some or all of the ambiguity can be addressed in this manner during iterative structure calculations as some ambiguous assignment pairs will be removed as violations due to frequent inconsistency with the rest of the experimental data.

All restraint lists were concatenated into a single restraint file. The ambiguous assignment was mostly preferred to avoid bias in any case. More assignments were added at later stages in accordance to agreement with the structure. The general protocol for simulated annealing involved ten iterations 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. Lower and upper bound corrections with values of 1.8 and 6.0, respectively were used for peak filtering. Default force constants in ARIA2, except for the dihedral and H-bond terms, were set up for the cool1 and cool2 steps 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. Restraint violations were evaluated using Dr. Rainey's *in-house* tcl/tk scripts. 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. [224] 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 refinement.[4]

2.3. RESULTS

2.3.1. Producing bioactive soricidin

The H_6-SUMO-soricidin fusion protein was best expressed in C41 (DE3) *E. coli* strain and purified using Ni^{2+}-affinity chromatography. This is evident from the SDS-PAGE gel image with lane I_3 indicating overexpression achieved through IPTG induction relative to lane U_3. (***Fig. 2.3.***). Chelation of the H_6-SUMO fusion tag onto the immobilized Ni^{2+} ions ensures retention of soricidin protein to the column while all other components of the cell lysate were washed out using wash buffer (lane W). The fusion protein is recovered by increasing imidazole concentration, as in elution buffer and collected in the elution fraction (lane E). Following SUMO protease-mediated proteolysis and dialysis, RP-HPLC purification yielded tag-free soricidin.

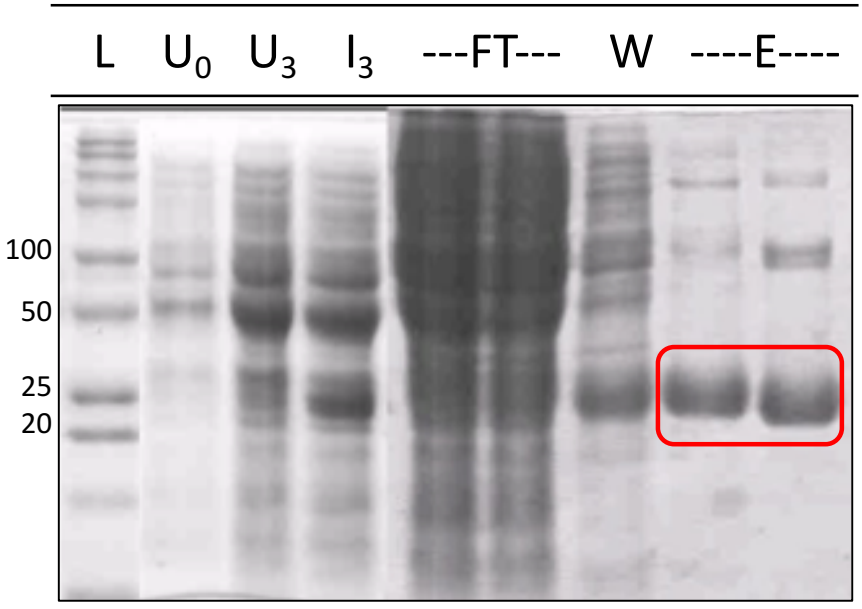


Fig. 2.3. *SDS-PAGE gel image depicting overexpression of H_6-SUMO-soricidin in E. coli cells and representative purification of cell lysate through a Ni-NTA column, allowing the H_6-SUMO-tagged protein to be retained in the column while all other products were collected in the flow-through. (L: ladder, U_0: uninduced cell culture at 0 hr, U_3: uninduced cell culture after 3 hrs, I_3: induced cell culture after 3 hrs, FT: flow-through containing the soluble fraction of cell lysate, W: wash, E: elution).*

Oxidative refolding was achieved O/N through the thiol-exchange technique[33] by adding GSH: GSSG mixture in 1: 1 ratio (final concentration of 30 mM) to the tag-free protein. This was evidenced by the replacement of the relatively broad peak observed before oxidative refolding (***Fig. 2.4.a.***) eluting between 20-24.3 min (wavelength, λ = 214 nm) with two new and better-resolved eluent peaks at shorter retention times in the RP-HPLC chromatogram (***Fig. 2.4.b.***). These peaks were consistent with the chromatographic profile for refolded synthetic soricidin that was provided by *Soricimed Inc.*

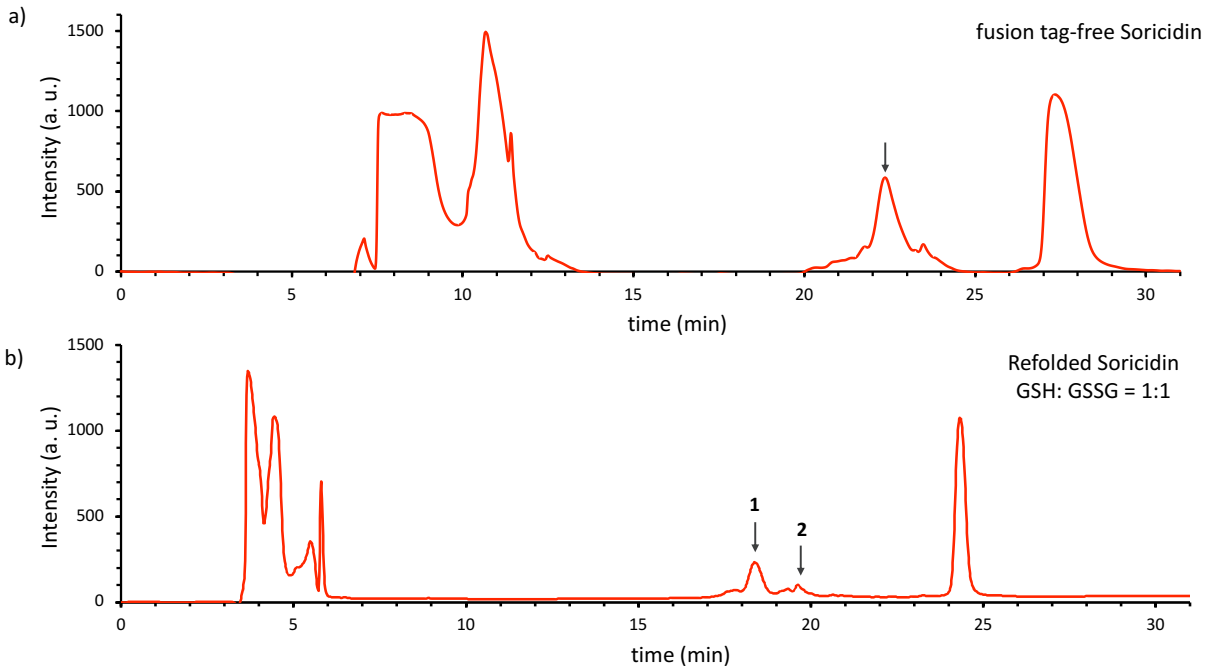


Fig. 2.4. *RP-HPLC chromatograms for fusion tag-free and refolded soricidin. a) Tag-free soricidin obtained after SUMO protease treatment eluted as a broad peak, around 20-24.5 minutes which was replaced by b) two narrower peaks following oxidative refolding and were collected at 18.8-19 mins (fraction-1) and 19.4-20.0 (fraction-2) minutes respectively.*

Both of the new eluent peaks observed after oxidative folding were collected and analyzed by electronspray ionization- mass spectrometry (ESI-MS; ***Fig. 2.5.***). The reconstructed molecular weight observed for both fractions was consistent with full-length soricidin having all Cys residues

fully oxidized (5806 Da; ***Fig. 2.5***.). Hence, both fractions collected after oxidative refolding are consistent by ESI-MS with fully oxidized conformers of soricidin. As each has a distinct elution time under the RP-HPLC conditions, this implies a difference in the disulphide bonding pattern.

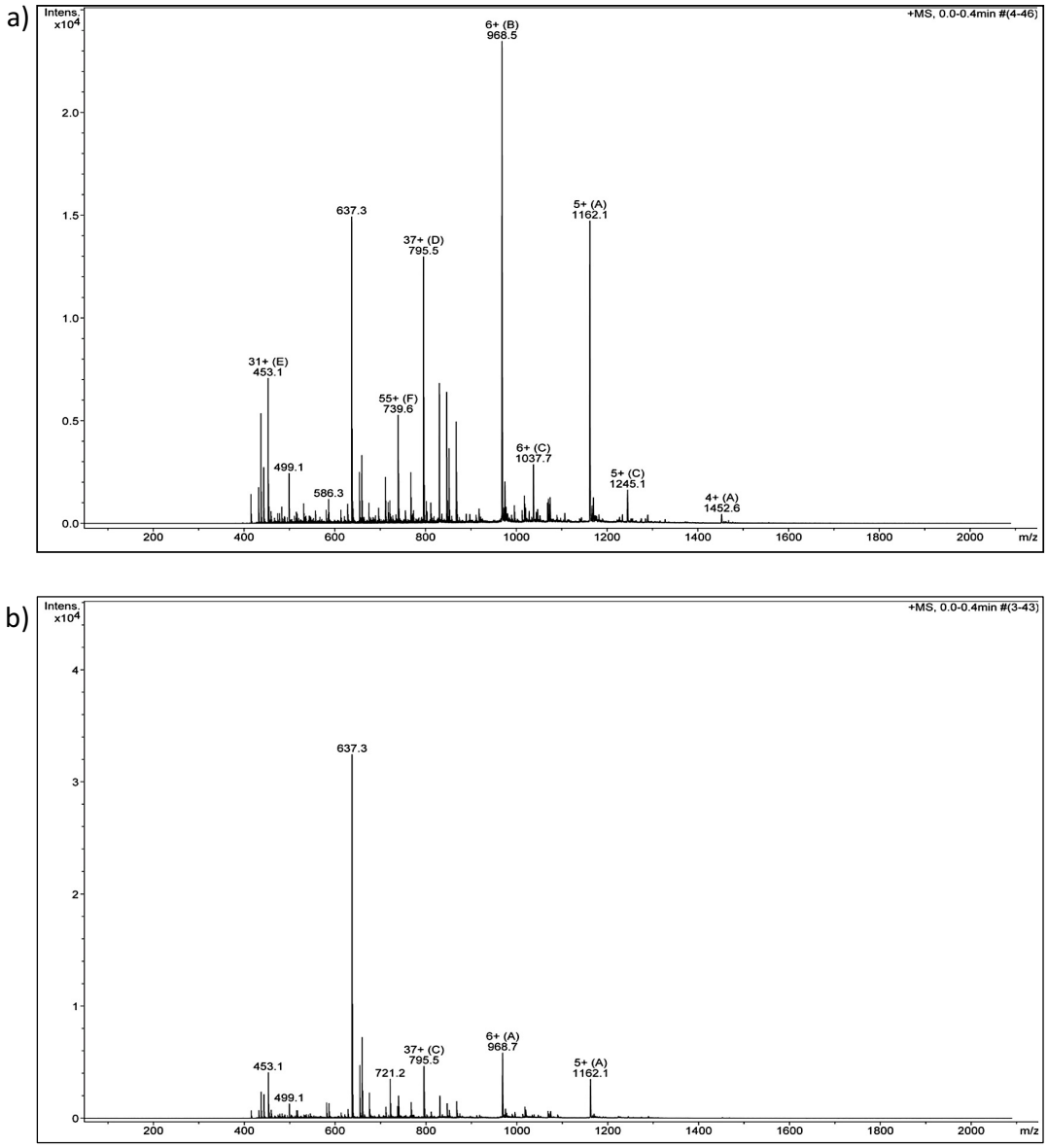


Fig. 2.5. *Mass spectra obtained for the major and minor peaks. Mass spectrum obtained for the lyophilized protein collected as a. Fraction-1 and b. Fraction-2 after RP-HPLC purification. Peaks observed at 1162.1(M^{5+}) and 968.7 Da (M^{6+}) confirm the presence of the soricidin (molecular weight: 5806 Da) with all Cys residues in an oxidized state.*

2.3.2. Bioactivity testing

To determine which fully oxidized soricidin species corresponded to the bioactive form, each fraction was employed in bioactivity assays with mealworms. In the case of injection with the major soricidin product (**fraction-1, *Fig. 2.5.a.*),** mealworms exhibited partial to full paralysis at lower doses and lethality at higher doses (***Table 2.5.***). All these worms were monitored over a period of 14 days to see the long-term effect of the protein. Conversely, despite the MS-based confirmation of its identity as an oxidized soricidin species, the minor conformer (**fraction-2, *Fig. 2.6.5.b)*** did not elicit the same response. This observation is consistent with the previous studies using chemically synthesized and refolded protein performed by *Soricimed Inc.* wherein two fully-oxidized forms of the protein were recovered, but only one exhibited bioactivity in a mealworm-based assay. Hence, the major soricidin conformer obtained following oxidative refolding was confirmed to be the bioactive form responsible for paralytic activity in invertebrates and was further investigated in depth.

Table 2.5.: FUNCTIONALITY TESTING ON MEALWORMS

Injectant	Concentration	Observation time following injection	
		10 min	14 days
Saline blank	-	-	-
Fraction 1	1× 2× 3×	backward crawling; rolling up involuntary movement no response to touch	recovered; slow movement partial paralysis dead
Fraction 2	1× 2× 3×	- - -	- - -

2.3.3. Evaluating the role of disulphides in soricidin structuring

The bioactive form of soricidin was first investigated using far-UV CD spectroscopy to evaluate secondary structuring. The triply-disulphide-bonded bioactive soricidin exhibited a clear

difference from the TCEP-mediated disulphide reduced form. Subtraction of the spectrum for reduced soricidin from that for oxidized soricidin shows a difference spectrum with a positive band at ~193 nm, negative band at ~208 nm and maintained negative ellipticity in the ~220-230 nm regime. This is consistent with the presence of α-helical structures[190] in the oxidized bioactive form of soricidin which are lost upon the reduction of the disulphides, rendering the protein more disordered (***Fig. 2.6.)***.

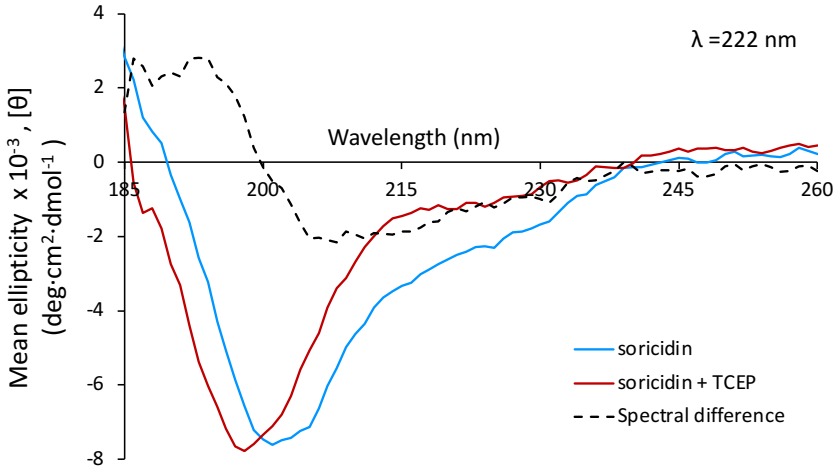


Fig. 2.6. *Far UV-CD spectra of reduced and refolded soricidin indicate the significance of disulphide bonds in maintaining secondary structuring in the functional conformer of soricidin, with TCEP-mediated reduction of disulphides leading to clear changes in protein secondary structure as is clear from the spectral difference obtained by subtraction of the reduced from the oxidized soricidin spectrum. Data were recorded in triplicate for 20 μM protein samples prepared in 50 mM sodium phosphate buffer with or without 10 mM TCEP (pH 7.00 ± 0.05, 30 °C).*

The disulphide bonds in soricidin, thus, appear to play a critical role in its higher-order structuring. This is certainly consistent with the requirement for disulphide-stabilization in venomous ion-channel blocking neurotoxic peptides found in reptiles or arthropods;[61, 225] however, whether or not soricidin has similar topology and structural features to these phylogenetically distinct toxins[226-229] is an important question to answer. Hence, solution-state NMR spectroscopy was next

employed to determine both the pattern of disulphide-bonding in the bioactive state of the protein and the high-resolution structure at the atomic level.

2.3.4. Anomaly in peak numbers and confirmation of sample purity

[U-^{13}C,^{15}N] soricidin was stable in solution with well-dispersed backbone amide resonances observable in ^{1}H-^{15}N HSQC experiments, consistent with a folded protein (***Fig. 2.7.***). However, more than the expected number of resonances were observed in the ^{1}H-^{15}N HSQC experiment. Specifically, for a 54 amino acid-long protein with three Pro residues, ~50 backbone amide resonance cross-peaks would be expected, however a larger number of peaks were clearly observable. This is easiest to disentangle in the distinct Gly cross-peak region of the spectrum where, for example, ~7-8 Gly cross-peaks were observable for this protein which has three Gly residues in total (***Fig. 2.7., inset)***. This anomoly in the observed number of peaks could be attributed to the presence of a distinct protein impurity, the presence of another soricidin conformer, or to the dynamic nature of the pure protein itself due to chemical exchange on the slow timescale with respect to the NMR experiment.

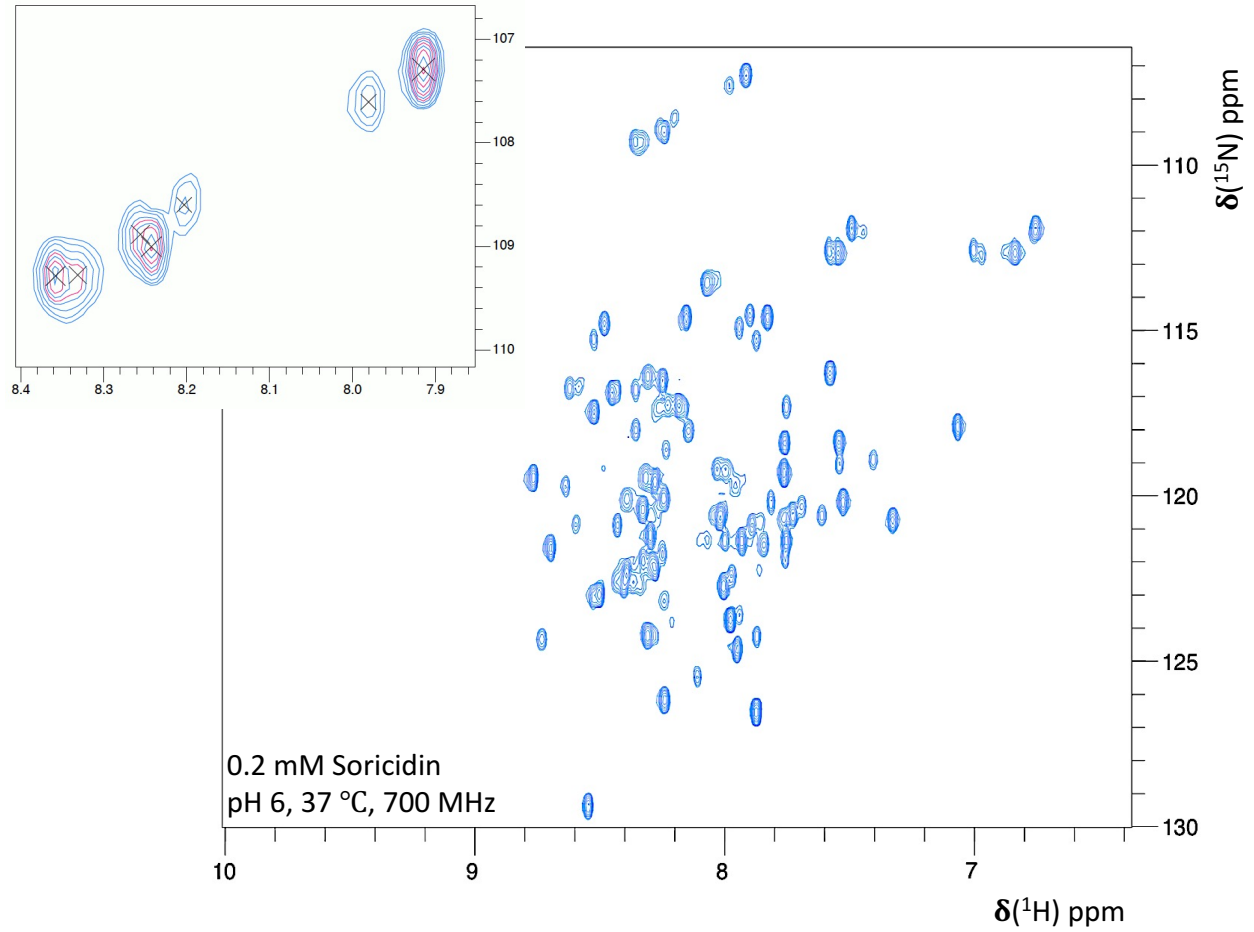


Fig. 2.7*. 2D ^{1}H-^{15}N HSQC spectrum of bioactive soricidin depoicting anomalous number of observed resonances for bioactive soricidin is depicted using an inset of the Gly region of the ^{1}H-^{15}N HSQC spectrum. Namely, uniformly ^{13}C/^{15}N-labelled soricidin showed more than the expected number of peaks for the 54-residue soricidin. For example, 7-8 Gly peaks (blue; expansion of the corresponding region for spectrum in Fig. 2.8), instead of 3, were observed. However, at higher plot levels (maroon), peak counts were found to be consistent with the expected number.*

^{1}H-^{15}N HSQC spectra collected at temperatures ranging from 278-310 K did not show any change in peak pattern (***Fig. 2.8.);*** i.e., neither significant differences in peak intensities nor a deviation from linearity was observed, thus ruling out the potential of these additional peaks arising from a conformational sampling equilibrium within the investigated temperature range which would be expected to lead to temperature-dependent variation(s) in peak number, position or intensity.[230]

Further analysis (as detailed in ***section 2.2.2.8.3.***) was not carried out in terms of calculation of temperature-dependent chemical shift changes.

Based upon the lack of apparent conformational sampling, analytical RP-HPLC was performed on the NMR sample under various conditions –temperature, solvent gradient and elution rate to test for the presence of additional folded-state conformer(s) or impurities. The chromatographic elution profile remained the same under all of the tested chromatographic conditions with the peak purity consistently observed to be at >95% (***Fig. 2.9.***). Given this set of findings, the additional peaks observable in HSQC experiments seem highly unlikely to be due to an impurity or distinctly folded soricidin conformer unless it perfectly co-elutes under many RP-HPLC conditions. Furthermore, at a level of 95% purity for the primary protein species, any impurity or distinct conformers present at a level of ~5% or less would not be at a sufficient concentration to give rise to peaks with the observed intensities given that ^{1}H-^{15}N HSQC cross-peaks – assuming relatively ideal experimental conditions – have an intensity proportional to concentration.

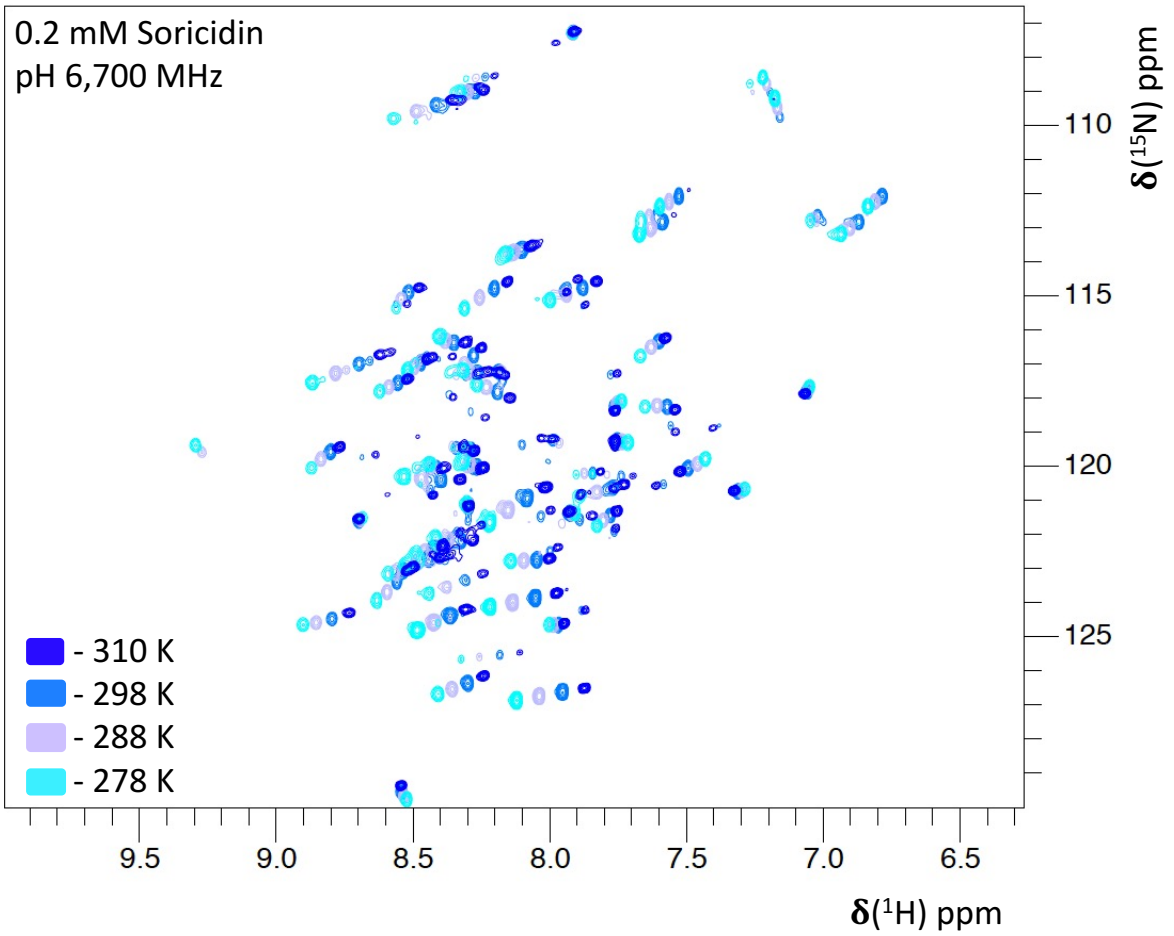


Fig. 2.8. *2D ^{1}H-^{15}N HSQC spectra acquired at various temperatures showed no pronounced differences in peak intensities, ruling out the possibility of an equilibrium based on temperature-dependent conformational sampling being responsible for the additionally observed peaks.*

Scaling of the ^{1}H-^{15}N HSQC spectrum to lower intensity leads to a situation where the number of remaining (i.e., most intense; ***blue***) cross-peaks is consistent with that expected for soricidin (***Fig. 2.7. inset, magenta***). Hence, it was concluded that the minor peaks observable by NMR spectroscopy could be arising from: (1) a differently folded and co-eluting conformer or impurity and/or (2) dynamic behaviour of pure, bioactive soricidin itself. This behaviour remains enigmatic as the latter hypothesis is hard to reconcile with the lack of temperature-dependent changes in the

peak pattern, while the former is difficult to rationalize since perfect co-elution under RP-HPLC would not be expected under the variety of HPLC conditions that were employed.

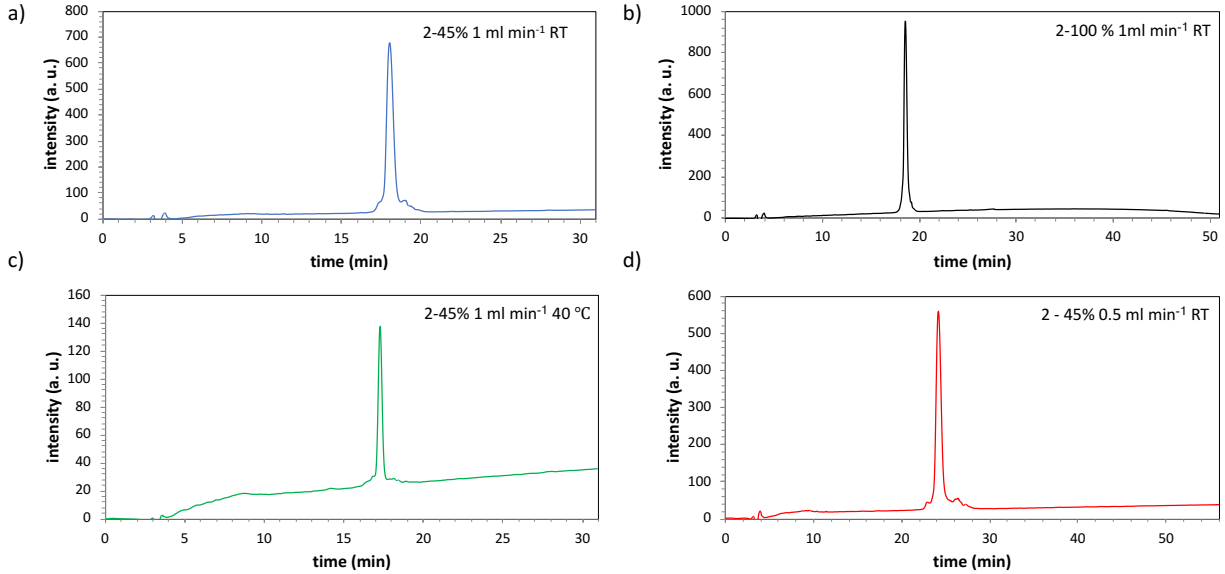


***Fig. 2.9**. HPLC chromatograms showing purity of NMR sample under various schemes. The purity of refolded protein was reconfirmed through RP-HPLC under various conditions. (Buffer A: 0.1% TFA in water, Buffer B: 0.1% TFA in ACN): a. Standard condition: 2-45% buffer B against buffer A, flow rate: 1ml min $^{-1}$; RT; b. Extended buffer gradient scale: 2-100% buffer B against buffer A, flow rate: 1ml min $^{-1}$; RT; c. Elevated temperature: 2-45% buffer B against buffer A, flow rate: 1ml min $^{-1}$, 40 °C; d. Lower flow rate: 2-45% buffer B against buffer A, flow rate: 0.5 mL min $^{-1}$, RT.*

Nevertheless, further analysis of the structure and backbone dynamics was carried out. This specifically focused on the predominant set of NMR resonances under the assumption that this is representative of the primary state of soricidin present in solution, with this state being referred to hereafter as the bioactive form of the protein. 37 °C was used for all subsequent NMR experiments as this mimics human physiological conditions where soricidin may be applied as a bioactive species. The similarity of ^{1}H-^{15}N HSQC spectral cross-peak patterns as a function of temperature (***Fig. 2.8.***) implies that the structure of soricidin is consistent over a wide range of temperatures,

over 5-37 °C. Good dispersion in the amide region indicates a structured protein, with a good likelihood of allowing for near-complete resonance assignment.

2.3.5. Analyzing solution-state triple resonance NMR data

2.3.5.1. Backbone and side-chain chemical shift assignment

The ^{1}H, ^{13}C, and ^{15}N chemical shifts were assigned for the bioactive state of soricidin (***Appendix A***) with assignment % detailed in ***Table 2.6***. (e.g., assigned ^{1}H-^{15}N HSQC in ***Fig. 2.10.)*** A series of triple-resonance experiments, listed in ***Table 2.4.*****,** were analyzed to perform backbone walk-based assignment.[231] Chemical shifts for the C', Cα, Cβ, N and H(N) nuclei were thus assigned. Side chain assignment was accomplished in a similar manner.

2.3.5.2. Assignment quality verification

Backbone chemical shift assignment quality was verified using the PANAV server, which specifically checks for chemical shift referencing offsets, mis-assignments, and provides a global assignment quality score, CONA.[217] The chemical shift reference offset(s) reported in this program allow for accurate compensation for any potential referencing errors, with cutoff re-referencing values of 1.0 ppm for Cα, Cβ and C' nuclei and 1.5 ppm for N nuclei. Any assignments that have standard deviations (SD) > ±4 relative to the expected random coil values are flagged as misassignments.[232] The summary of the PANAV assessment report is presented in ***Appendix B***, based on which chemical shift assignments are of high quality.

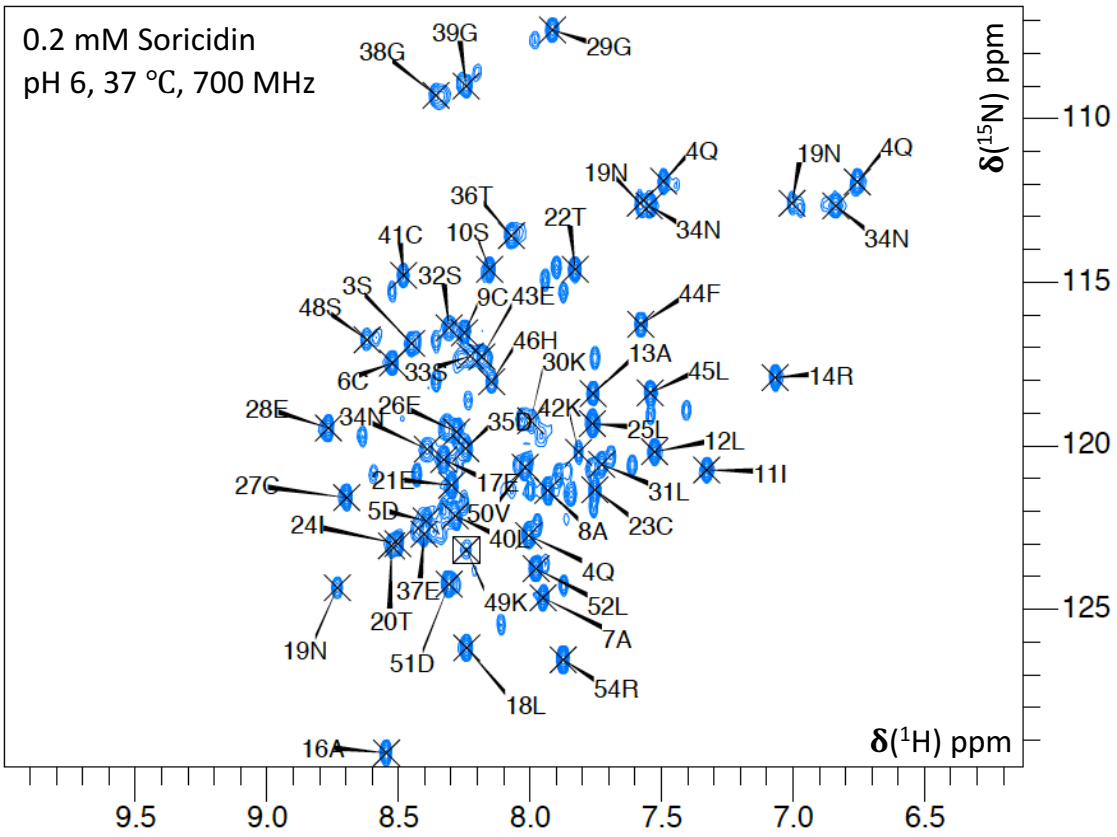


Fig. 2.10. *Assigned 2D ^{1}H-^{15}N HSQC spectrum of [U-^{13}C,^{15}N] (main chain) soricidin (0.2 mM) in H_2O/D_2O (90/10 v/v). (Data collected at* 37 °C, *pH 6, at 16.4 T using NRC-Halifax instrument).*

Table 2.6.: CHEMICAL SHIFT ASSIGNMENT REPORTS	
Backbone (%)	
backbone N	90.7
backbone H+(H + Hα)	97.2
backbone C (C+Cα)	97.2
Side chain (%)	
Cβ	98.1
H	89.3
non-H	73.7

2.3.6. Secondary structure

Large deviations in Cβ chemical shifts were observed relative to standard random coil values for all six Cys residues (***Fig. 2.11.****a)*)**.** The observed values are consistent with those reported for Cys in its oxidized state, i.e. cystines,[233] consistent with the mass spectrometry data detailed above. CSI values of -1 observed for two major and a few minor regions suggest, as discussed in section ***2.2.2.9.****a)*)**, the presence of α-helices. Consensus dihedral predictions using the assignments for Hα, Cα, Cβ, and C also indicate two major α-helices. In short, the consensus of secondary structure inferences on the basis of both the CSI and the DANGLE algorithm[221] suggested two extended helical regions in the bioactive soricidin conformer (***Fig. 2.11.***) as was suggested by far-UV CD spectroscopy.

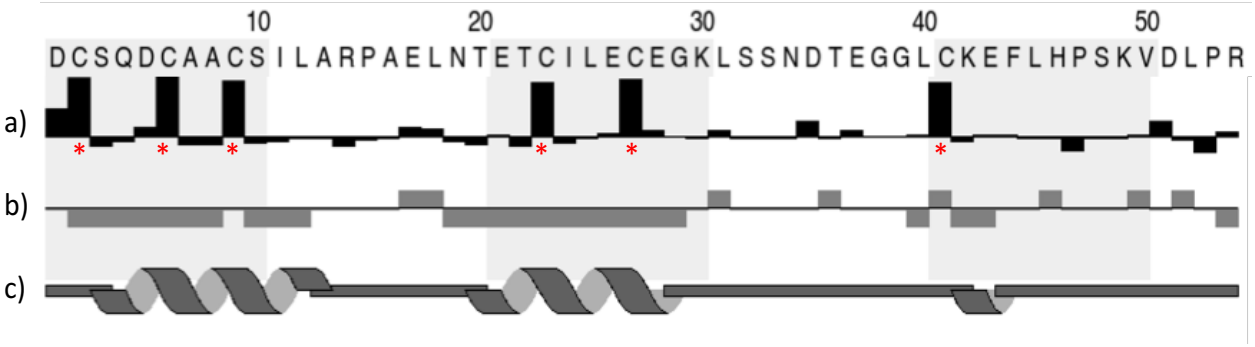


Fig. 2.11. *Residue-level summary of the secondary structure of soricidin**.** a) ΔδCβ values for all the residues are shown, with those marked with red asterisks showing significant deviations relative to random coil values, confirming the oxidized state of all the six Cys residues. b)Chemical shift index (CSI) bar graph, based upon Hα, Cα, Cβ, and C' assignments, consistent with the presence of two extended helical regions (-ve CSI). c) Secondary structure predicted by the DANGLE algorithm on the basis of assigned backbone chemical shifts (Hα, Cα, Cβ, and C'). The bar graphs and DANGLE secondary structure prediction were generated by CCPNMR-Analysis.*

2.3.7. Local dynamics

Enhancement factors derived from heteronuclear-NOE-modulated ^{1}H-^{15}N HSQC spectral analysis show a rigid protein core with a dynamic C-terminal tail region (***Fig. 2.12.***). Notably, the less dynamic segment encompasses both long helical segments that were inferred from chemical shift information (***Fig. 2.11.***) and also covers all six cystines. For comparison, previous NMR spectroscopy-based studies on SOR-13 analyzed a short peptide comprising only residues 41-54. In this prior study, two helical turns were identified followed by a dynamic tail, features which are in general agreement with the data being reported here for the full-length, disulphide-bonded protein.[162]

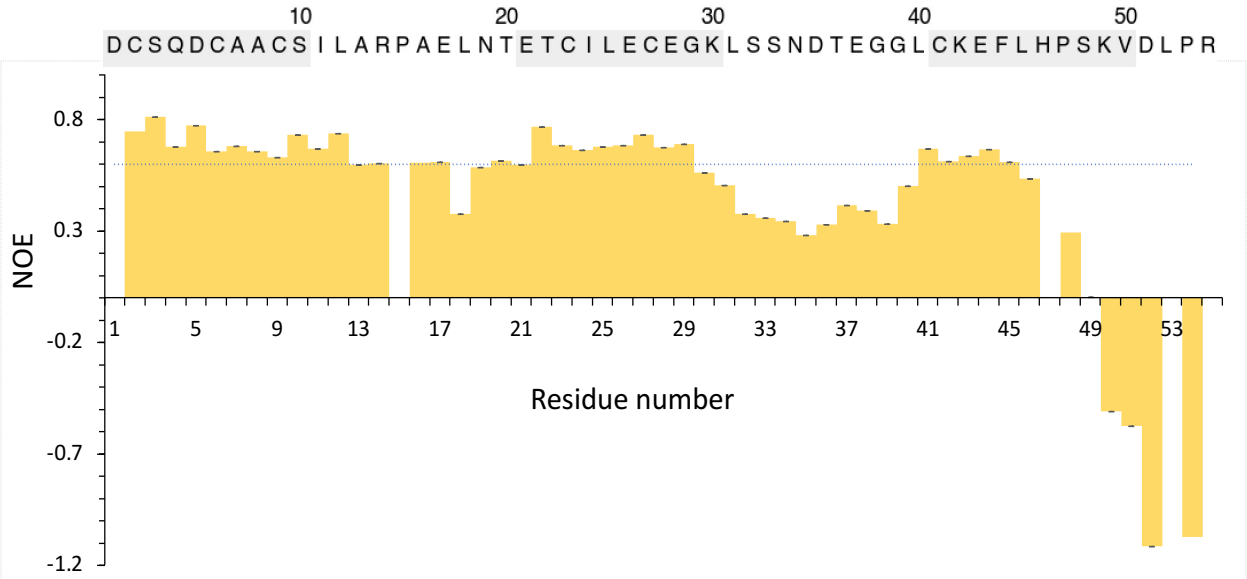


Fig. 2.12. *^{1}H-^{15}N heteronuclear NOE for the backbone amide peaks suggest a highly rigid protein core with a dynamic C-terminal tail in the ps-ns timescale.*

2.4. STRUCTURE OF SORICIDIN

2.4.1. Final structure calculations in ARIA 2.3

The initial NOE list with a large number of ambiguous assignments was filtered using ARIA2.[234] Distance restraints imported from CcpNMR Analysis[213] and DANGLE[221]-predicted dihedral

angles were used as input restraints with ARIA2. A series of ~18 iterative structure calculations and manual restraint refinement rounds were carried out in CNS 1.2.[235] following detailed protocols outlined previously.[236] Disulphide bonds were defined as part of the covalent topology of the protein following explicit verification that these were consistent with the distance and dihedral angle restraints. For each round of structure calculation, 100 NOE- and dihedral angle restraint-based structures were calculated and used to further refine NOE restraints. Following an iterative assignment and refinement process, a set of >1300 NOE distance restraints were determined corresponding to through-space ^{1}H-^{1}H interactions within a distance of 6 Å (***Table 2.7.)*** (***Fig. 2.13.)***.

Table 2.7.: NOE RESTRAINTS USED IN STRUCTURE CALCULATION	
a) Unique Distance Restraints	
Total NOE	1339
Intra-residue	730
Inter-residue	
Sequential (|i-j| = 1)	336
Medium-range (|i-j| <4)	170
Long-range (|i-j| >5)	60
Ambiguous	43
b) Dihedral Angle Restraints	
Phi (ϕ)	45
Psi (φ)	45

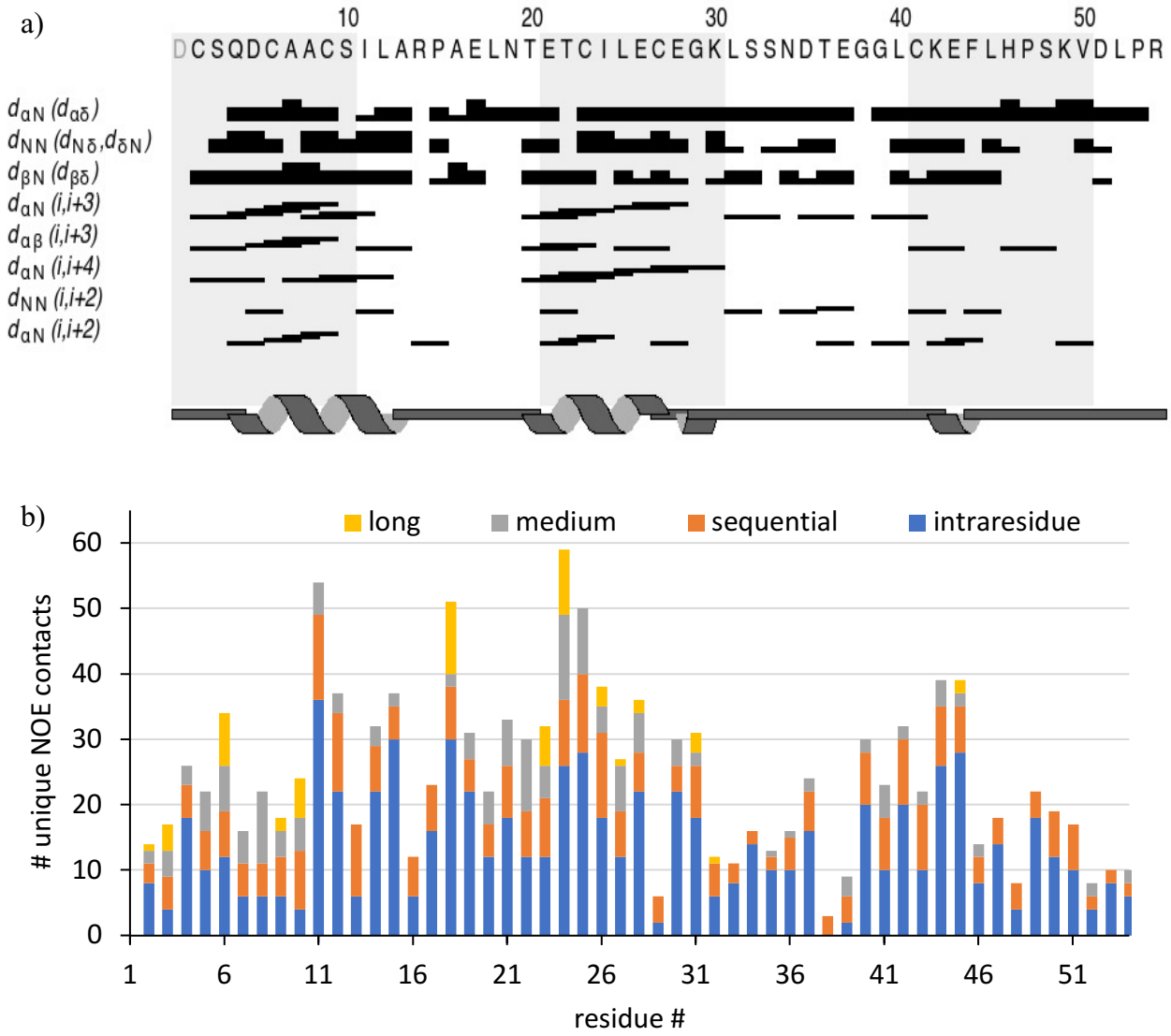


Fig. 2.13. *NOE restraints retained in the final round of structural calculations. Dangle predicted secondary structure is shown alongside a) the distribution and b) a graphical summary of NOE restraints across residue number.*

Simulated annealing structural calculations using these NOE restraints, DANGLE-based dihedral angle restraints based upon the chemical shift assignments, and disulphide links determined through careful evaluation of NOE patterns provided an ensemble of 20 structures (retained from a set of 100 calculated structures) (***Fig. 2.14.)*** in excellent agreement with the experimental restraints (***Table 2.8***). The 20 lowest-energy structures following water refinement (of 100

calculated structures) were retained in the final structural ensemble and visualized using Pymol.[237] The average backbone RMSD values, calculated relative to the lowest energy structure using *in-house* scripts along with dihedral angle parameters were used to identify local divergence/dynamics among the ensemble members.

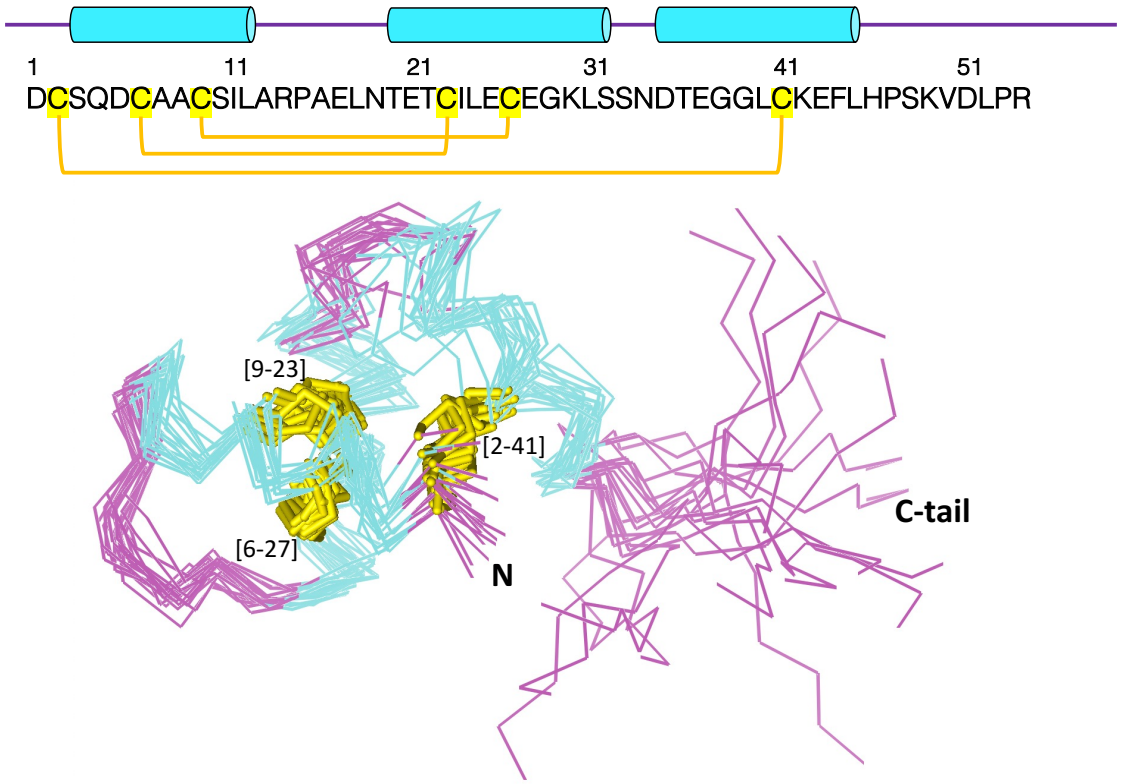


Fig. 2.14*. Overlay of 20 lowest energy structures from the 100 lowest energy structures generated in the final round of structure calculation. All the ensemble members are aligned along Cα of the converged region consisting of residues 3S-45L (blue: α-helix, purple: random coil, yellow: disulphide bridges).*

2.4.2. Tertiary structure

Rigorous structural analyses involving the evaluation of φ/ψ dihedral angle deviations[238] and dihedral angle order parameter,[239] S (***Fig. 2.15***.) were performed to distinguish between the structurally converged and dynamic regions of the protein. Low dihedral angle deviations alongside high S values were observed from 3S-45L suggesting structural convergence in this

region (***Fig. 2.16.)***, echoing the chemical shift-based secondary structuring inferences[240, 241] (***Fig. 2.13.a)***). High RMSD values for the C-terminal region of soricidin are consistent with the lack of structural convergence reported in previous NMR spectroscopy-based studies of this portion of soricidin, SOR-13.[162]

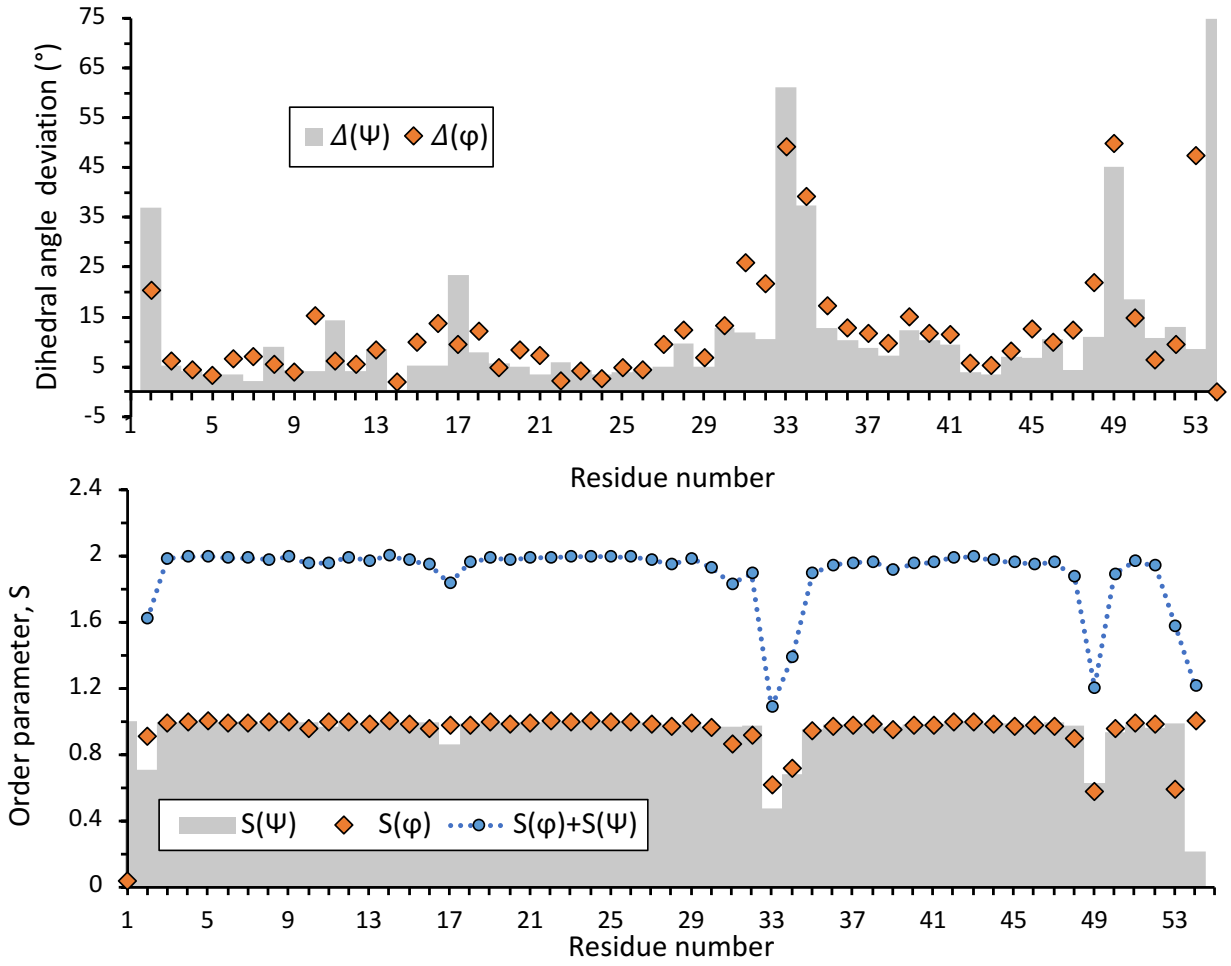


Fig. 2.15. *Structurally convergent segments were identified by analysing a) φ/ψ dihedral angle deviations and b) corresponding order parameters.*

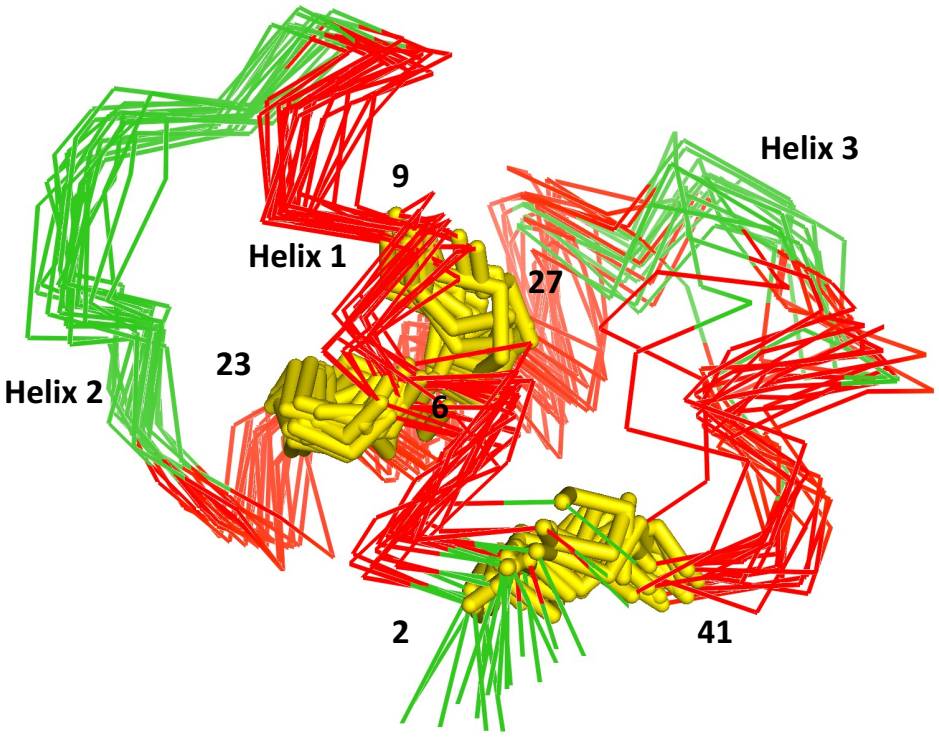


Fig. 2.16. *Overlay of the 20 lowest energy structures over the residues 2-45 shows high structural convergence. A figure-8-like enclosure formed through the three disulphide linkages includes the major α-helical regions.*

2.4.3. Disulphide connectivity

The disulphide bridges were identified as 2-41, 6-27 and 9-23, thus adopting an I–VI, II–IV, III–V arrangement. While two of these (6-27, 9-23) inter-lock the first two α-helices, the third one (2-41) connects the N-terminus to the third helix, proximal to the C-terminal tail, conferring an enclosed loop structure with an "figure-8-like" shape flanked by a flexible C-terminal tail (***Fig. 2.17.***). Thus, the three disulphide bridges play a crucial role in building such a unique 3D architecture to this protein.

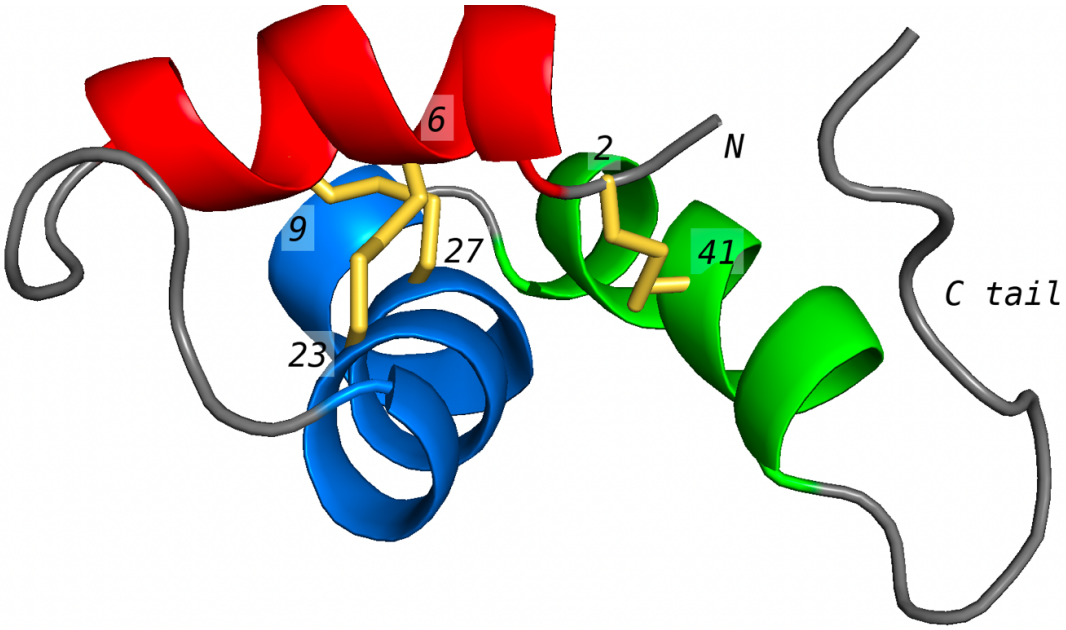


Fig. 2.17. *Representative 3D structure of functional soricidin with N, C-termini and Cys labelled. Bioactive conformer forms a structure with three intramolecular disulphide bonds – while 2 of these occur within the major α-helices, the third one connects the N- terminus to the third helix just ahead of the dynamic C-tail region.*

2.4.4. Surface potential

Specific pharmacological functions of venoms are often attributed to the molecular surface of these proteins such that the type and position of hydrophobic interactions and salt bridges between charged residues guide the inter-helical angle.[242] The surface charge distribution of soricidin indicates a large acidic surface locally disrupted by basic residues at two particular sites. (***Fig. 2.18.***) Upon investigation, one site with positive surface potential was identified within the enclosed loop corresponding to residues Arg14 (loop 1) and Lys30 (helix 2), with their side chains placed at a distance of ≤ 6-7 Å in most of the ensemble members (***Fig. 2.19.***) and the other site was identified as Lys49 on the C-terminal tail. Notably, the basic-aromatic amino acid pair identified here is attributed to the functioanl dyads mediating ion-channel activities in other venom-derived bioactive proteins; hence this could be the potential binding site involved in Na^+

channel inhibition[243] leading to paralytic activity, while the antineoplastic activity of the positively charged SOR-13[162] could be attributed to the electrostatic interaction with the selectivity filter ring (-ve charged) in TRPV6.[244]

Table 2.8.: STRUCTURAL ENSEMBLE STATISTICS	
Residual restraint violations	
RMSD NOE Restraints (Å)	0.016 +/- 0.002
Distance restraint violations	
> 0.5 Å	0
> 0.3 Å	0
Average Violations From 0.1-0.3 Å	6.7 +/- 2.5
RMSD Dihedral angle restraints (°)	0.41 +/- 0.1
Average no. of dihedral angle violations per structure > 5˚	0.1 +/- 0. 3
Residual restraint violations:	
RMSD NOE Restraints (Å)	0.016 +/- 0.002
Distance restraint violations > 0.5 Å	0

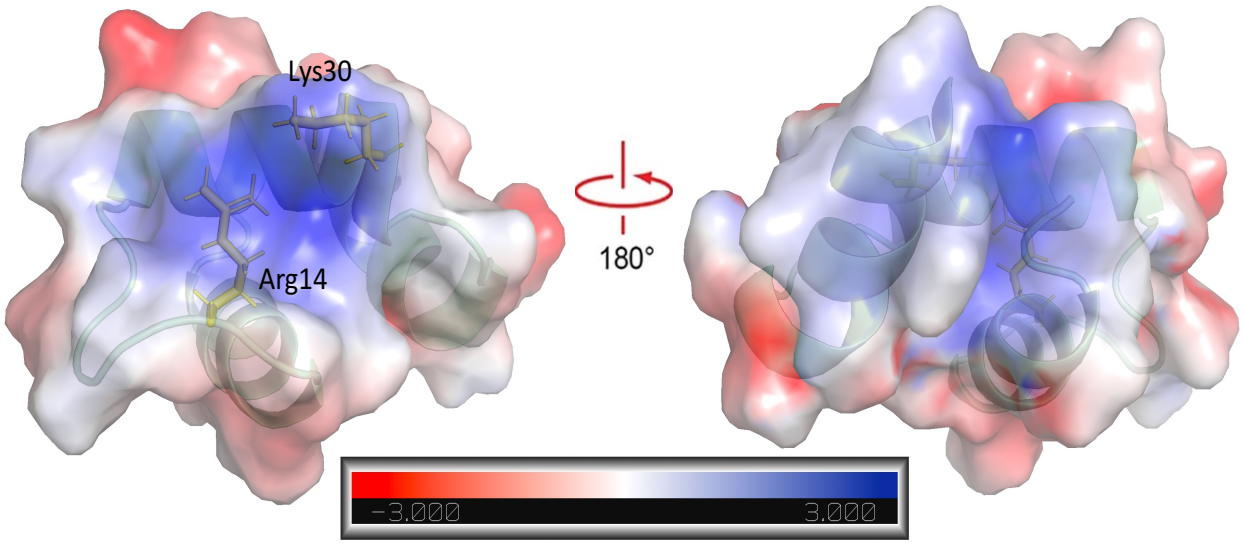


Fig. 2.18. *Electrostatic surface potential for soricidin was calculated using Adaptive Poisson-Boltzmann Solver (ABPS) Electrostatics Plugin included in Pymol Version2.4.2.. Highly positive electrostatic potential correspond to Arg14 (loop 1) and Lys30 (helix 2) and residues on the*

structured region, indicative of a probable binding/active site. Surface is colour coded as red: acidic, white: neutral, and blue: basic residues.

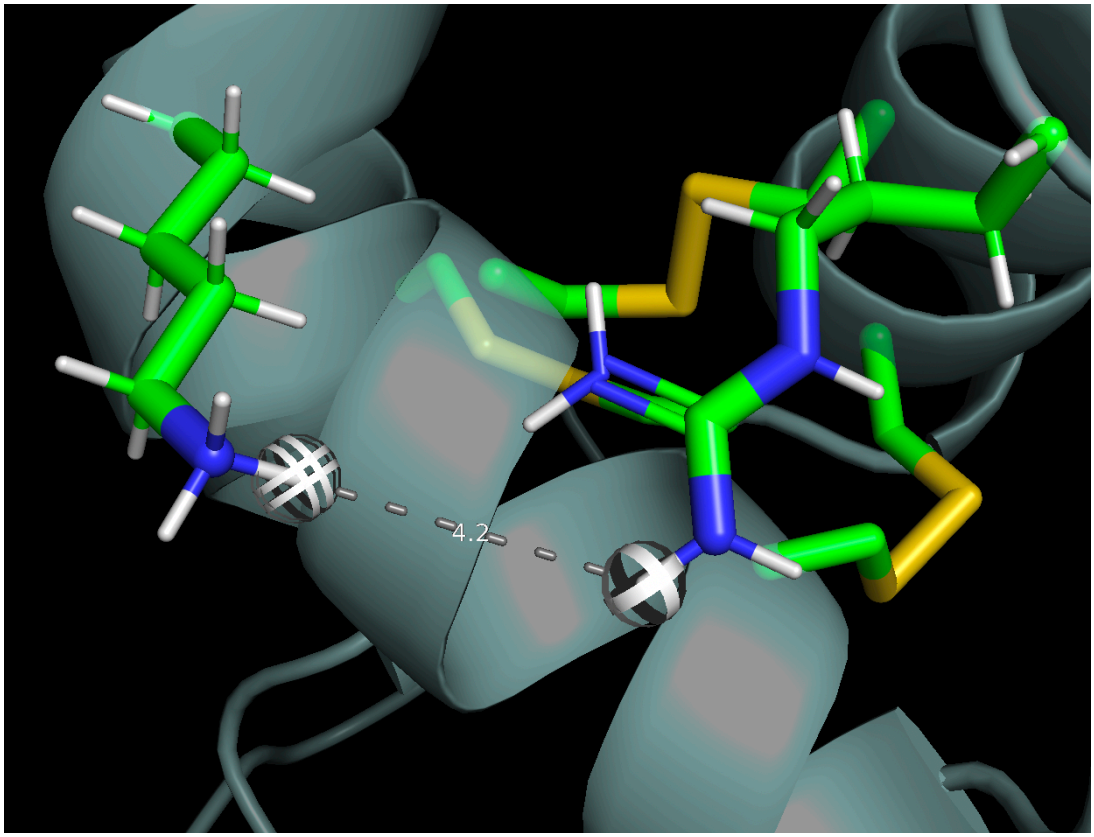


***Fig. 2.19**. Representative diagram showing the putative dyad present in soricidin. Distance between the side chains of Arg14 and Lys30 (4.2 Å) is shown. (grey, protein, blue: basic residues, yellow: disulphide bridge).*

2.5. DISCUSSION

2.5.1. Conserved Cys motifs identified

Looking at the primary structure, soricidin contains both CXXXCXXC and CXXXC (where X is any amino acid residue other than Cys) motifs, identified as conserved motifs in many disulphide-rich venom proteins. The CXXXC motif within the α-helix region is often conserved in K^+ channel blockers isolated from scorpion venom of *Tityus stigmurus* and *Buthus tamulus*[245-247] while CXXXCXXC motif is understood to modulate enzymatic function, ion-channels etc. by providing

a specific metal-binding site, as in viperin,[248] ceramide kinase[249] and the radical S-adenosylmethionine (SAM) superfamily of enzymes etc.[250, 251] Interestingly, the CXXXCXXC motif is located in the N-terminal half of the SAM enzymes, as in soricidin.[250, 251]

2.5.2. Cysteine-stabilized protein folds

Studies on botulinum toxin have recognized the decisive role of interhelical disulphide bridges in conferring the functional conformation to the protein. Reduction of these disulphide bridges could potentially cause a conformational shift of the helices and associated amendment(s) in the catalytic site could adversely affect its pharmacological role – ranging from a decline in neurotoxicity to complete inactivation.[252] Site-specific mutations and disulphide reduction in the fungal toxin restrictocin exemplify the decisive role of these covalent bonds not only in providing resistance towards proteolytic degradation but also in furnishing the requisite functional structure to the protein.[253] Hence, the interhelical disulphide pattern identified in soricidin likely contributes to the temperature independence of the protein structure (as observed in ***Fig. 2.8.***) and its physiological role, with the probability for loss of function with reduction of disulphide bridges. It is interesting to note the abundance of Cys in the enclosed loop in comparison to in the C-terminal tail, which exhibits a higher degree of dynamics over a ps-ns timescale. The unambiguous identification of the disulphide pattern in the bioactive conformer is key for the synthetic production of this protein with the desired disulphide bonds as in the bioactive form.

With the N terminal and central helices interlocked through two disulphide bridges (6-27, 9-23), and the N-terminus connected to the third helix that is found proximal to the C-tail through a third disulphide bridge (2-41), an enclosed and well-structured rigid loop is formed which is flanked by the flexible C-tail region. (***Fig. 2.20.)*** While the first loop between the initial two helices is eight

residues long, the second loop is as small as two residues and the C-terminal tail is constituted by the last 13 residues SOR-13 which is reported to showcase antineoplastic activity.[162] Even though soricidin is a small protein with a molecular weight of ~ 5 kDa, the enclosure formed through terminal disulphide demonstrates a remarkable resemblance to the pseudocylic structures reported for long-chain toxins exhibiting high binding affinity.[86, 88] Notably, the third helix present in the calculated ensemble structure was not initially suggested in the secondary structure predicted by the DANGLE algorithm. The agreement of the chemical shift data with the calculated structural ensemble implies that the helix is indeed supported by the chemical shift data but may not be as tightly restrained to the helical region of the Ramachandran plot on the basis of the chemical shifts as the other helices in the protein. Two turns of an α- helix previously reported in the NMR-based structural calculations performed on SOR-27 fragment in 100 mM SDS micelles[162] is in agreement with the later part of the third helix in the current structure of the full-length protein.

With its ladder-like pattern of disulphides stapling the α-helices in the protein, soricidin adopts a Cys-stabilized helix-loop-helix (CSα/α) hairpin fold.[254] With hydrophic hairpins thought to be critical in the formation of a functional pore,[63] the alternative CSα/β fold is more common, although examples of CSα/α also exist in the literature. For example, similar structures have been reported in the vicilin-buried peptide (VBP) family constituted by bioactive molecules exhibiting a plethora of functions including antimicrobial, ribosome inactivator, ion-channel blocking, protease inhibition, antinociception, paralysis, neurotoxicity etc.[254] isolated from scorpions, spiders, sea anemones,[95, 242, 255-260] and plants.[257, 258, 261] However, it should be noted that the K^+ channel blockers in the κ-KTx subfamily of scorpion venom adopting CSα/α fold are reported as relatively poor blockers. Also, Kunitz-type toxins (e.g., Hg-1, BmKTT-1 and BmKTT-2) with bifunctional properties, inhibiting K^+ channels and proteases share the same disulphide patterns.[82]

Interestingly all protein structures reported with a helix-loop-helix fold either have 0, 2 or 4 interhelical disulphides, but none have 1 or 3.[94, 178, 254, 256, 257, 262-264] Looking at the initial two helices linked through two disulphides, this generalization holds in the case of soricidin as well.

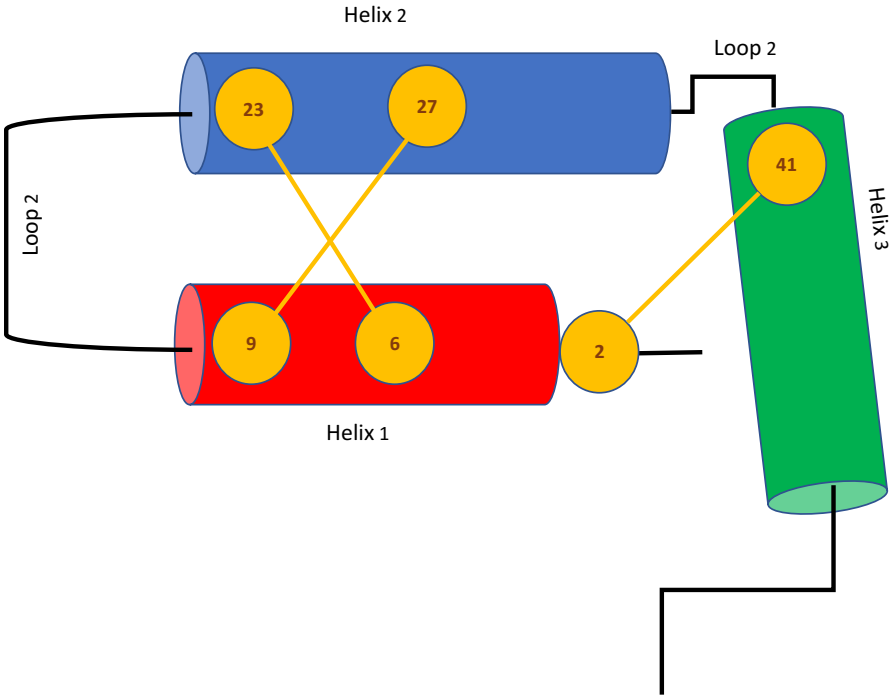


Fig. 2.20. *Schematic diagram depicting the topology of soricidin structure. While two interhelical disulphides confer the CSα/α fold, terminal disulphide generates an enclosed loop flanked by a C-terminal tail.*

2.5.3. Disulphide pattern

Even though there are three disulphides as in the popular inhibitor cystine knot (ICK) (I-IV, II-V, III-VI), and three α-helices connected via two loops of variable length as in centipede toxin SSD609, the I-VI, II-IV, III-V disulphide connectivity reported in soricidin differs from both these proteins.[81] However, a similar disulphide pattern with the CSα/α protein fold has been reported in defensins and the Kunitz inhibitor family constituting calcicludin from snake venom, SHPI-1 from the sea anemone, κ-theraphotoxin-Hh1a from Chinese bird spider.[265-269]

2.5.4. Identification of a putative surface dyad

The Lys/Arg residue pair identified on the surface of all ensemble members in the soricidin structure is in agreement with those corresponding to K^+ channel blockers isolated from scorpions and cone snails, where the functional dyad is thought to interact with the pore of the ion channel.[72, 270-272] Furthermore, the localized positioning of this residue pair on the protein surface could allow for targeted protein engineering to improve the specificity/potency of this bifunctional protein as a molecular probe or drug.[273]

2.6. SUMMARY

The study detailed in this chapter showcases the first demonstration of recombinant expression of a mammalian venom-derived protein using a bacterial expression system. Following cell lysate purification through Ni-affinity chromatography, and fusion-tag cleavage, RP-HPLC was employed for further purification. This strategic change in the purification method was adopted since the presence of multiple cysteines could promote random disulphide bond formation, triggering protein coagulation or aggregation[274] at a higher concentration as would be the case in the purification of the protein through reverse Ni-NTA column chromatography.

The disulphide bridging process was accelerated through the thiol-disulphide exchange method, facilitated by the usage of equivalent amounts of GSH in both reduced and oxidized forms. The predominance of the fraction (~90-95% of total protein based upon chromatography analysis) containing the bioactive form of soricidin, alongside significant minimization of the second fraction O/N, indicates a drastic improvement in the yield of the desired product in comparison to prior methods (***Fig. 2.4.)***. Namely, refolding of reduced soricidin peptides prepared by solid-phase peptide synthesis by *Soricimed Inc.* required a reported ~14 days to recover the bioactive

conformer with a <60% yield. My methodology thus accelerated this ~14 fold while improving the yield from 60 to >90%.

The presence of intermolecular disulphide-linked multimers was not observed by MS, indicating that the protein solution was sufficiently dilute over the course of the refolding process I developed to facilitate the formation of the specific, energetically favourable intramolecular disulphide bond pattern over random, intra- and/or intermolecular bonds that would have been likely to result in a heterogeneous protein mixture of monomers and oligomers.

The recombinantly-prepared bioactive conformer of this mammalian venom-derived protein allowed me to determine the tertiary structure of the full-length protein. This was previously an unknown structure, with even the pattern of disulphides being elusive before my studies.

Chapter 3

INTRODUCING DISULPHIDE LOCK IN AcSp1 ACINIFORM SPIDROIN

3.1. INTRODUCTION

3.1.1. Previous studies on AcSp1 aciniform spider silk

3.1.1.1. W unit adopts a globular structure in solution state

Recombinant aciniform silk proteins ranging in size from a single W unit (W_1; consisting of 199 aa's with the N-terminal Ser being omitted for cloning purposes) through to four concatenated W units (W_4, with 799 amino acids, again omitting the N-terminal Ser) were first developed and tested for fibre formation capability, the results of which demonstrated a linear improvement in mechanical properties with an increase in the number of W units.[3] Isotope-enrichment of W_1 was subsequently performed, enabling atomic-level investigations by NMR spectroscopy using uniformly ^{13}C- and ^{15}N- isotope-enriched proteins.[9] This enabled the determination of the high-resolution structure of the protein in its soluble state.[9] Complete structural characterization using triple resonance solution-state NMR methodologies revealed that the W unit adopts a globular structure consisting of five α-helices interconnected by unstructured regions, with the N- and C-terminal segments being intrinsically disordered in the soluble state.[5] (***Fig. 3.1.***).

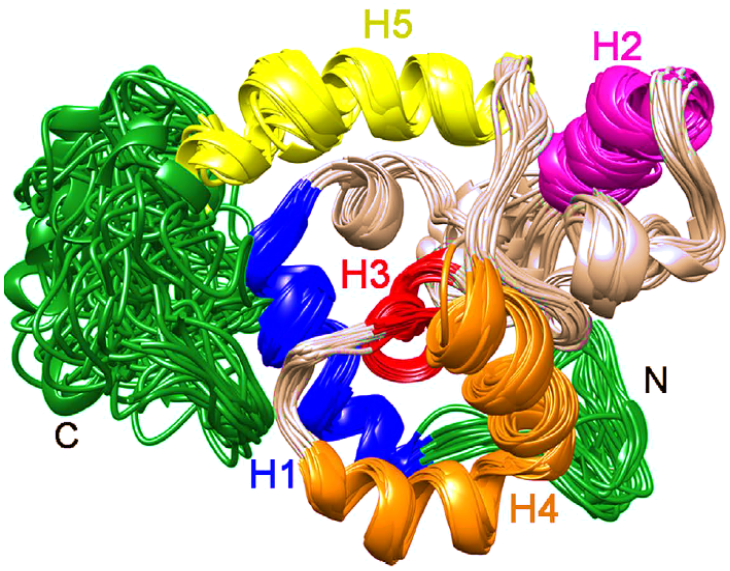


Fig. 3.1. *Structure of AcSp1 W unit reported through solution-state NMR spectroscopy (as adapted from Tremblay et. al.[5])*

3.1.1.2. W units behave as a modular system

Based upon the difference in fibre formation capability that was observed between W_1 and longer concatemers of W units, NMR studies were carried out on segmentally-labelled W_2 proteins – i.e., proteins having one W unit at natural abundance and the other ^{13}C- and ^{15}N-enriched - to investigate the effects of concatenation upon W units. Quantitative comparison of backbone chemical shifts and dynamics over ps-ns timescale demonstrated that each W unit within the W_2 concatemer behaves as a discrete module, without structural perturbation and with minimal restriction in dynamics.[5, 275, 276] Coupling segmental-labelling-based W_2 chemical shifts, structural restraints from W_1, and NMR-based hydrodynamics measurements, structural models were produced that show W_2 to be a compact "beads-on-a-string" concatemeric state. In this model which is consistent with all of the experimental data, the connected C- and N-terminal domains of joined W units form linkers between globular units that are intrinsically disordered, but relatively compact in the soluble state.[5] Compellingly, NMR samples containing these W_2 proteins were

capable of fibre formation upon application of shear force, implying that this is a constructive starting point for fibre formation.

3.1.1.3. Fibre formation involves a significant structural transition

CD and Raman spectral analyses have revealed that the conversion of soluble silk proteins to rigid/solid insoluble fibres involves the introduction of β-sheet content. Unlike many other classes of silk fibre (most notably, the highly studied dragline/major ampullate spider silk and silkworm silk), which are devoid of α-helicity, both recombinant and native aciniform silk show a significant amount of α-helicity being retained in the fibrous state.[122, 275] This, in turn, seems likely to relate to the functionality of this silk, which requires greater extensibility than the non α-helical silks. To date, the direct functional role of α-helicity has not yet been proven.

3.1.1.4. Localized instability at H5 region elevate dynamics

Titration experiments performed with both W_1 and W_2 to induce protein denaturation with the chaotropic reagent urea and structural perturbations induced by the zwitterionic detergent dodecyl phosphocholine (DPC) demonstrated that destabilization of helix 5 (H5; residues 135–149; yellow helix in ***Fig. 3.1.***) was the first conversion to take place. A simultaneous structural rearrangement was reported in the protein core beneath it, consistent with these regions being proximal to each other. Measurement of backbone dynamics through ^{15}N experiments following these chemical shift perturbation-based observations from heteronuclear-NMR and ^{19}F-NMR studies confirmed the elevated dynamics of H5, which is immediately N-terminal to the intrinsically disordered C-terminal region, relative to the remainder of the protein.[276, 277]

3.1.2. Proposed hypothesis for fibrillogenesis mechanism in aciniform spider

Based on our current understanding from previous structural and backbone dynamics studies conducted in the Rainey lab on the soluble structure of AcSp1 proteins,[9, 277, 278] we know that the W unit before fibre formation has a relatively plastic and destabilized portion – helix 5 (H5; residues 135-149) and the protein core beneath it which undergoes significant perturbation upon H5 unfolding – and we speculate these specific regions to be involved in the conversion to fibres. A hypothesis for the transformational change from the highly concentrated soluble state to start the journey on to eventual, insoluble fibre formation of *AcSp1* was proposed. Namely, the localized unfolding of the fifth helix (***Fig. 3.2.)*** coupled with dissociation of this segment of the protein from the globular core of the W unit, would lengthen the linker region and decrease the compactness of the structure. Upon encounter with W unit(s) of another molecule(s), the decompacted state with elongated linkers would serve to promote intermolecular interactions, hypothetically leading to the formation of β-sheets in the H5 region and/or linker region and, ultimately, result in β-sheet enriched and α-helically depleted fibre formation.

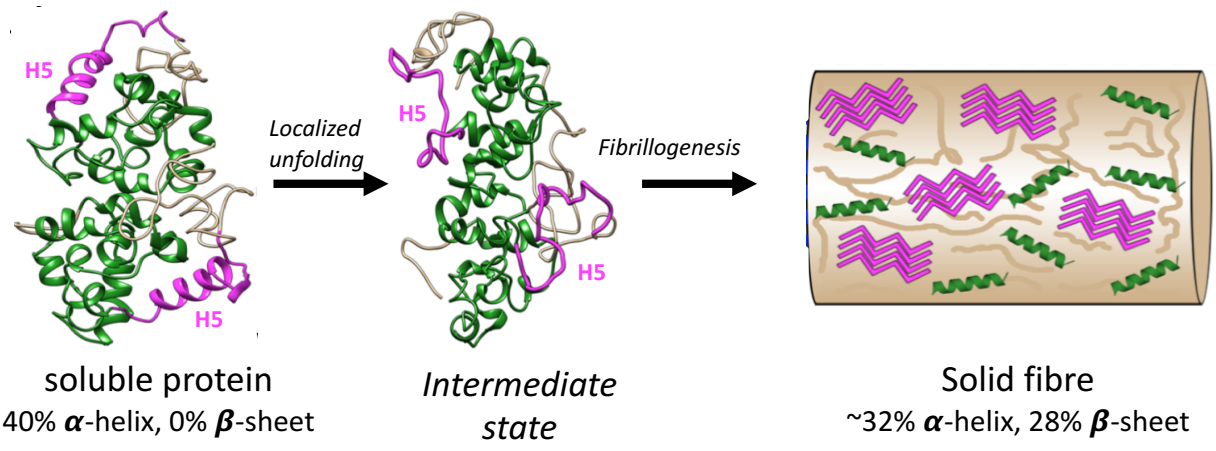


Fig. 3.2. *Schematic diagram illustrating the proposed fibrillogenesis mechanism for AcSp1 protein. Intermediate state structure was produced by Prof. Jan K. Rainey using XPLOR-NIH[4] in the same manner as for W_2 protein structute[5], except that restraints in/to helix 5 were excluded.*

3.1.3. Engineered mutant silk protein

Based upon the spatial proximity of side chains (hydroxyls at a distance of ~3.4 Å and dihedral angle between them being ~90 ° (Fig. 3.3.)), as determined using the solution-state NMR-derived W1 structural ensemble, and the potential of a disulphide bridge to form a structural "lock," 29Ser, located in the core domain beneath H5, and 143Ser, within the Ser5 motif of H5 were chosen for site-directed Ser →Cys mutations to investigate the hypothesis proposed above (Fig. 3.2.). Site-specific replacement of the Ser gamma-hydroxyl (-OH) with a Cys gamma-thiol (-SH) would increase steric bulk for these two residues (covalent radii of single-bonded sulphur = 1.05 ± 0.03 Å and single-bonded oxygen = 0.66 ± 0.02 Å)[279] in the case of reduced mutant while oxidation would introduce a disulphide bridge in the stapled form. This uniqueness in terms of Cys oxidation states and absence of Cys residues in the wild-type (WT) W unit[280] was advantageous while engineering this mutation since undesired side reactions did not need to be considered. As an added benefit, this particular mutation also provides the opportunity to introduce heterogeneity in the otherwise exclusively Ser chemical shift patterns within the Ser5 motif (S141-S145) present in the H5 region. As this region is speculated to seed the events of structural transition, the mutation would be likely to enhance the likelihood of successful NMR-based characterization of this key region.

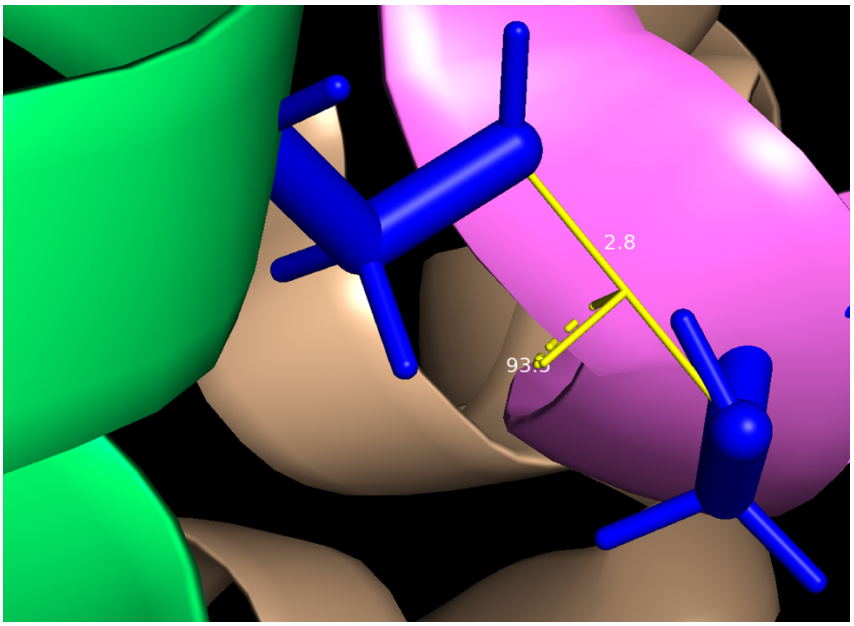


Fig. 3.3. *Representative diagram showing the interatomic distance between the oxygen molecules on the hydroxyl side chains of 29Ser and 143Ser and the dihedral angle between them. (green: H5, purple: core beneath H5, blue: hydroxyl side chains of 29Ser and 143Ser, rest of the protein is shown in the background.)*

3.1.4. Silk fibre spinning

Our current understanding of natural spidroin silk assembly is primarily based on studies performed on dragline (major ampullate) silk, with the protein fold before and after fibre formation being completely different. The atomic-level rearrangements initiating extension, alignment and packaging of individual protein molecules include conformational changes and phase separation, and are described by two primary hypotheses: namely, the 1) liquid crystal theory[281] and 2) micelle theory.[282] (***Fig. 3.4.)*** Although these are somewhat distinct in the postulated form of protein that is expected in the gland, both have experimental evidence to back them up and – as noted below – they need not be considered mutually exclusive mechanisms.[283, 284]

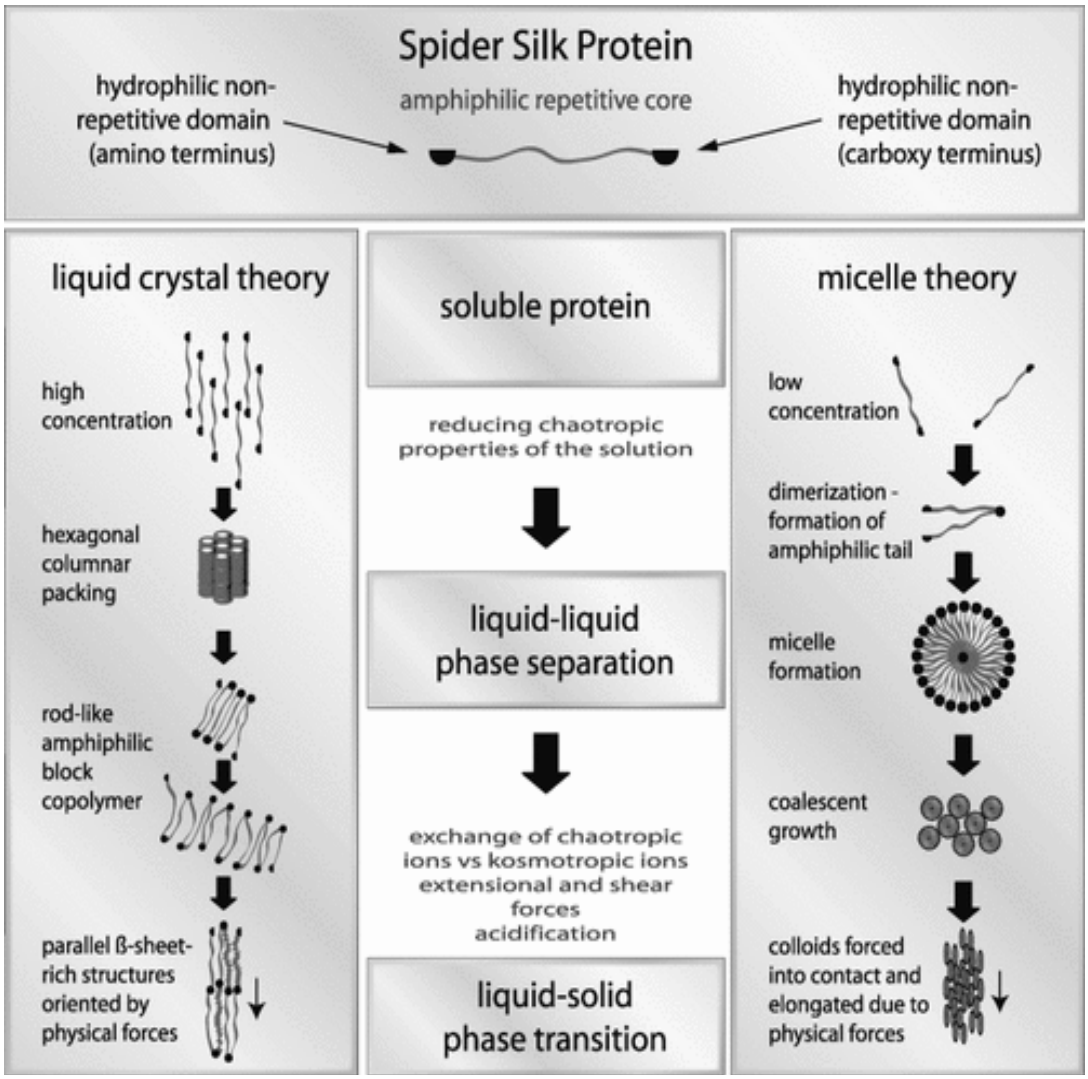


Fig. 3.4. *Theories describing spider silk thread formation (as adapted from Heim et. al.)*[6]

Spidroins are secreted and stored within a specialized spinning gland in the form of a highly concentrated solution, often referred to as the protein "dope". Navigation of this protein dope from the spinning gland through a tapering spinning duct involves changes in the chemical environment (pH, ionic strength etc.) and shear forces, which in turn initiates fibrillogenesis.[283] According to the first theory noted above, silk proteins demonstrate a liquid-crystalline behaviour. Namely, unidirectional alignment of protein molecules at high concentrations favours intermolecular interactions in the form of van der Waals forces and hydrogen bonding between the side chains

and solvent molecules, without evident conformational or solubility changes to the protein leading to liquid-liquid phase separation.[281] Alternatively, the second model advocates a micellar organization of the proteins. Owing to the amphiphilic nature of the silk proteins considered in the development of the hypothesis, nanoparticle assemblies (100-200 nm) are initially formed, whose association at higher concentrations is manifested in the form of micron-sized globules.[282] The formation of liquid crystals or globular intermediates involves liquid-liquid phase separation that occurs without evident conformational or solubility changes to the protein.[281, 285, 286] Accelerated elongational flow and shear forces towards the tapering end of the spinning duct cause mechanical drawing and/or unfolding of the protein to elongated shapes known as prefibrillar assemblies (i.e., a liquid-solid phase transition). Subsequent removal of solvent molecules (dehydration) instigates further conformational changes and rapid thread assembly leading to the formation of silk fibres.[287-289]

Underlying the attempt to mimic and replicate the materials produced through natural spinning processes, advancements in the areas of bioengineering and materials sciences along with the available knowledge of silk assembly have led to the introduction of various techniques for artificial spinning of fibres.[137, 152, 290-292] Two techniques are used to prepare recombinant protein fibres in the studies detailed in this chapter– a) hand drawing, used as a primary indicator of fibre formation ability,[293] and b) wet spinning, used to produce uniform fibres for comparative analyses.

3.1.4.1. Hand drawing of fibres

In the hand-drawing technique, fibres are pulled out of a droplet of protein-containing solution on a substrate with appropriate surface chemistry prepared at a sufficiently high concentration using forceps, tweezers, a glass rod, etc.[293, 294] This technique is often used as a preliminary indicator of

the fibre-forming ability of biomaterials like protein. However, this technique is limited in scope due to inherent difficulty in ensuring consistent experimental parameters such as extrusion speed, applied shear forces, etc. and the potential for human error or, at the least, differences in technique from one individual to another. This technique, hence, may not be optimal for preparing fibres that are both uniform and that have reproducible mechanical properties.

3.1.4.2. Wet spinning of fibres

3.1.4.2.1. Spinning dope

As with the protein dope solutions found in spinning glands, an artificial spinning dope is a highly concentrated protein solution. This is prepared by dissolving/resuspending the purified spidroin in a suitable solvent mixture optimized for fibre spinning. In the literature, organic (toluene, 1,1,1,3,3,3-hexafluoroisopropanol (HFIP)), inorganic (lithium bromide (LiBr), calcium nitrate ($Ca(NO_3)_2$), calcium chloride($CaCl_2$) etc.) and aqueous solvents (urea, buffered saline, etc.), and/or different combinations of these substances are reported for solubilizing various fibre-forming proteins, including recombinant spidroins. [144, 155, 157, 290, 291, 295-300] Each solvent type has distinct advantages and disadvantages for spinning dope preparation, with the choice of solvent mixture often based on protein compatibility. [155, 291, 295, 301, 302]

Acids are one of most the common dope constituents used for fibre-forming silk proteins. Although there is potential for sample degradation resulting from structural perturbation and modifications in peptide bonds in an acidic environment,[295, 301, 303, 304] this can be circumvented by reducing the time interval between the dissolution and spinning processes. Fluorinated organic solvents, specifically fluroalcohols (including HFIP, hexafluoro acetone (HFA), 2,2,2-trifluoroethanol (TFE) etc.) are known for their potential to disrupt the hydrogen bonds through preferential

solvation of the protein surface i.e., hydrating water molecules are replaced by fluroalcohols capable of forming relatively stronger H-bonds,[305 306] hence, changing the composition of protein hydration sphere without significant levels of protein degradation.[307] Studies on another protein prone to fibril formations, amyloid-β, have demonstarted that these solvents are capable of disrupting α-helices to collapsed coil conformations.[306] Since hydrogen bond rearrangements are crucial part of the major protein conformational changes that occur and for ensuring compactness during fibre formation,[308-310] fluorinated organic compounds are prevalent solvents (or components of solvent mixtures) used in solubilizing spidroin proteins.

Comprehensive studies carried out in the Rainey lab have explored various solvent mixtures with different compositions for aciniform silk spinning dope preparation and their effect on the fibres formed. These studies have highlighted the relevance of the choice of dope solvent in determining the ultimate outcome of the fibre spinning process. Namely, this initial step can impact not only the ability to form silk-like fibres but also the mechanical properties of the resulting fibres.[158, 291, 311, 312]

3.1.4.2.2. Wet spinning process

Silk fibres may be produced from the artificial spinning dope using a wet-spinning apparatus. The apparatus employed in my work for wet-spinning includes an extrusion syringe, coagulation bath and – optionally – a series of rollers allowing for post-spin stretching. (***Fig. 3.5***.) The dope solution in the syringe is extruded at a controlled speed under shearing flow into a solvent (mixture), called the coagulation bath, where the protein becomes amalgamated into the solid form – fibre. Often, solvents like methanol, ethanol or isopropyl alcohol are used in the coagulation bath.[292, 313, 314] The

fibre thus formed may be collected directly onto a roller and, at this stage, is referred to as an as-spun (AS) fibre.

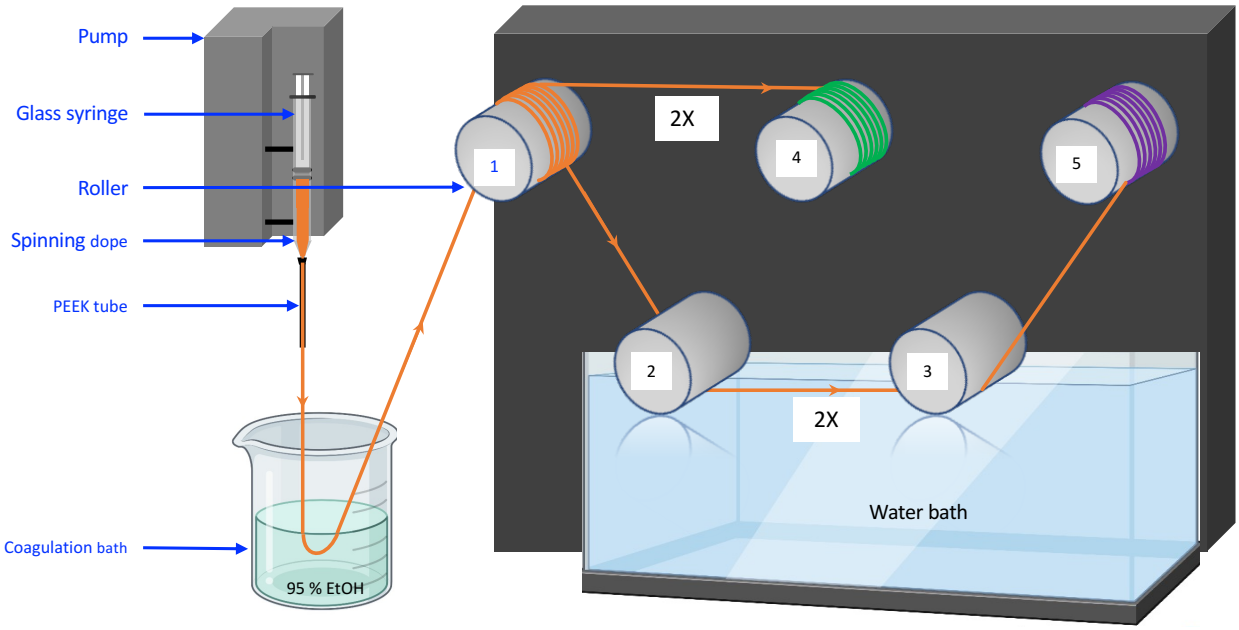


Fig. 3.5. *Schematic representation of the wet-spinning apparatus in the Rainey lab.*[7, 8] *As-spun (AS) fibres are collected on roller #1 (orange), 2X stretched in the air on #4 (green) and 2X stretched in water following 2X stretching in the air on # 5 (purple).*

Further molecular alignment and conformational changes through protein decompaction are often achieved with post-spin stretching treatments. This treatment is accomplished by stretching in air, one or more solvent baths, or both, with fibres, thereafter, referred to as post-spun (PS) fibres.[7, 278] Adding another variable that must be tested and controlled for in the wet-spun fibre processing, a variety of post-spin-stretching draw ratios may also be used at any of the steps. In general, the post-spin stretch process induces significant long-range ordered alignment and fosters hierarchical structuring within the fibre, consequentially improving mechanical properties like strength, extensibility, and toughness.[7, 158, 297, 299, 309, 315] With the apparatus mimicking natural silk assembly and each parameter optimized for the process, wet-spinning is one of the most extensively employed methods for reconstituted and recombinant silk fibre spinning.[8, 288, 295, 315]

3.1.5. Objectives

Understanding the complex process of the hierarchical transition of the aciniform silk from the primary to quaternary structure levels is crucial for engineering novel, advanced silk-based materials. With the aim to explore the unknown trigger and early stages in the transition from a highly soluble protein to insoluble aciniform silk fibre formation, the experimental work presented in this chapter focused on:

i. a comparison of the ability to wet-spin fibres from recombinantly produced Ser → Cys substituted W_2 proteins in the reduced and stapled forms;
ii. characterization of the morphology and mechanical behaviour of each type of fibre,
iii. analysis of the protein self-assembly in dope solution prior to fibre spinning; and,
iv. probing the solution-state behaviour of the protein.

To test for effects arising from both the change in side chain chemistry and upon the change in the conformational restriction upon disulphide locking, the functional performance of the wildtype, the reduced mutant, and the oxidized mutant are compared. For convenience, protein constructs will be annotated as:

i. WT W_1 or W_2- a single or double W unit construct of native *AcSp1* protein
ii. W_1M - W_1 with Cys29 and Cys 143 mutations; and,
iii. W_2M - W_2 with Cys29 and Cys 143 mutations in both W units (i.e., Cys29 and Cys 143 of the first and Cys229 and Cys343 of the second W units mutated).

While the reduced form of the mutant protein will be referred to as reduced W_1M or reduced W_2M, the stapled counterpart will be simply called W_1M or W_2M.

3.2. MATERIALS AND METHODS

3.2.1. Preparation of mutant protein

3.2.1.1. Plasmid construction and choice of constructs

Plasmids encoding W_1 and W_2 proteins with two Ser residues mutated to Cys on each W unit – Ser29 and Ser 143 - were obtained from Dr. Lingling Xu. In each instance, the W protein was fused to an N-terminal H_6-SUMO tag, following previous strategies for cloning and expression of W-based proteins.[5, 316] The construction of recombinant plasmids is described in *Xu, et. al.*[9] Since W_2 has been previously identified as the minimum required size of protein for recombinant aciniform silk fibre formation,[3, 275] throughout the course of my work, W_2 proteins have been employed for functional testing to check both fibre-forming ability and mechanical properties of the fibres formed henceforth, while the corresponding W_1 mutant proteins have been employed for NMR spectroscopic characterization at the atomic-level, discussed in ***Chapters 4*** and ***5***.

3.2.1.2. Recombinant protein expression and purification

Plasmids encoding either wildtype or mutant H_6-SUMO-W_2 were transformed into BL21 (*DE3*) *E. coli* cells and protein overexpression was achieved in LB medium[9] at 37 °C in the same manner as described in section ***2.2.2.*** Cells harvested at OD_{600} ~0.8-1.2 were centrifuged (6500 rcf /20 min/4 °C), resuspended in lysis buffer, and lysed using a French pressure cell press. The resulting lysate was centrifuged and the H_6-SUMO fusion proteins were purified from the soluble fraction using a Ni-NTA agarose column, as detailed in section ***2.2.2***. Protein-containing elution fractions were identified through absorption spectroscopy, collected and incubated for 1 h with SUMO protease (prepared *in-house*) in the presence of DTT to cleave the H_6-SUMO tag. Following O/N dialysis against 50 mM phosphate buffer (PB; pH 8.00 ± 0.05) at 4 °C, reverse-purification was

performed using a Ni-NTA affinity column incubated with ice-cold 50 mM PB. Upon loading the Ni-NTA column with the dialyzed solution, the H_6-SUMO fusion tag was retained by the column, while the WT W_2 protein was collected as elution fraction. Ni-NTA column was further treated with wash buffer, and fusion-tag removal was achieved at high imidazole concentrations as in elution buffer.

W_2M protein was expressed following the same protocol. Following reverse-purification, DTT was added to a final concentration of 10 μM to one half of the protein solution to keep the Cys residues reduced and, hence, obtain reduced W_2M, while the other half was used to prepare the oxidized sample.

3.2.1.3. Generating stapled mutant protein

Previous trials by Dr. Lingling Xu had established that the disulphide bond formed in W_1M to ~ completion through self-oxidation over a period of 2 weeks. To try to enhance the rate of this process, various techniques were tested through trial-and-error under basic conditions (i.e., pH 8)[317] in order to accelerate the process in W_2M (***Table 3.1.).*** The thiol-disulphide exchange method, discussed in section ***2.2.4.,*** and successfully employed to enhance oxidative folding in soricidin, was first tested with different GSH/GSSG ratios over different timescales and at various temperatures. The potential to apply oxygen pumped to the protein solution[318, 319] was also tested. Finally, W_2M protein solution obtained after reverse purification was subjected to 10× dilution with 50 mM PB and stored at 4 °C for self-oxidization. The oxidation status of the protein was qualitatively monitored as a function of time through 15% SDS-PAGE gel imaging with Coomassie Brilliant Blue R-250 as the staining agent and by NMR spectroscopy. Non-reducing conditions were used to check the oxidation status of the purified protein while reducing conditions

were used otherwise. All protein fractions were dialyzed against distilled water to remove any salts present, lyophilized following shell freezing by submersion in liquid N_2 and stored at -20 °C until use.

3.2.2. Functional testing

3.2.2.1. Hand drawn fibres

The hand drawing technique was tested on the W_2M protein fraction freshly obtained from reverse purification to verify the fibre forming ability of the reduced W_2M protein, and with the other fraction of W_2M after confirming complete oxidation.[3] DTT was added to the protein suspended in 50 mM PB at pH 8.0 and shear stress was induced by pulling from the solution with a pipette tip in the former case while DTT was avoided in the latter.

3.2.2.2. Preparing dope solution

Following the most recent studies from our group showing optimal wet-spun recombinant aciniform silk fibre formation,[278] a solvent mixture of trifluoroacetic acid (TFA)/ TFE/ type I distilled water (TFA/TFE/H_2O) in 8/1/1 (v/v/v) ratio was used. 30% (w/v) lyophilized proteins (WT W_2, reduced W_2M and W_2M) were individually resuspended into this mixture in glass vials. Protein suspensions were vortexed every 10 mins until homogeneous and transparent, subsequently centrifuged (18,000 rcf/30 min/ 4 °C) and the supernatant was transferred into new glass vials. 200 µL aliquots of each spinning dope were prepared at 1/100 dilution and the UV-Vis absorbance at 214 nm (molar extinction coefficient, ε_{214} = 555,716 M^{-1} cm^{-1}) was measured using 0.05 cm quartz cuvette (*Hellma; Müllheim, Germany*) to estimate the protein concentration as per ***Eq. 2.1.*** Vials containing the spinning dopes were closed to avoid solvent evaporation and

subsequent concentration change. Degradation of the protein sample over time was monitored using SDS-PAGE gel imaging.

3.2.2.3. Wet-spun fibres

The fibre-forming ability of the dope solution was initially tested in 100 % ethanol (EtOH) solvent before loading into the extrusion syringe for spinning. Fibres of WT W_2 and reduced W_2M were transferred to slides for imaging while the microfibrillar structures of oxidized W_2M were retained in the EtOH bath.

Once fibres formation in EtOH was confirmed, WT W_2 spinning dope was loaded into a 100 µL Hamilton reversible needle (RN) syringe (*Hamilton, Reno NV*), with a 6 cm long polyetheretherketone (PEEK) tube (outer diameter: 1/16 inch; inner diameter: 0.005 inch; *Sigma-Aldrich; Oakville, ON*) using RN compression fittings (1/16 inch; *Hamilton; Reno, N*V) attached to the syringe needle. The loaded syringe was then securely mounted onto a KD Scientific model 100 series syringe pump (*Holliston, MA*). Dope was extruded through the PEEK tube at a constant speed of 16 $\mu L \cdot min^{-1}$ into a 95% ethanol coagulation bath. Fibres were carefully picked up with tweezers in the coagulation bath and guided onto the first roller (#1 on ***Fig. 3.5.***) attached to an LKB 2232 MicroPerpeX S Pump (*GE Healthcare Life Sciences; Mississauga, ON*), that was spinning at a constant speed equivalent to the extrusion rate. This set of fibres is referred to as WT W_2 _AS. Following this, wet spinning was performed with reduced W_2M. Alongside collecting AS fibres, fibres after post-spin stretching treatment were also collected: 2X stretched in the air (W_2M_PS_2X_air; #2 on ***Fig. 3.5.***) and 2X stretched in a water bath (W_2M_PS_2X_water; #5 on ***Fig. 3.5.***). All fibres were collected at 28 ± 5 % relative humidity and 30.5 ± 0.5 °C, with the fibre ends secured by tape and stored on the roller at RT in a sealed container until used.

3.2.3. Characterization of wet-spun fibres

The fibres collected onto the roller (WT and reduced W_2M) and stapled W_2M microfibrillar structures in EtOH were characterized using various microscopic and biophysical techniques. Fibres collected on rollers were carefully removed, cut into pieces of a little more than 1 cm length, and mounted onto 3x2 cm rectangular testing cards with a 1 cm gap in the middle (***Fig. 3.6.***) for both diameter determination and morphological characterization using optical microscopy followed by mechanical testing. Microfibrillar structures of stapled W_2M were retained in EtOH until used.

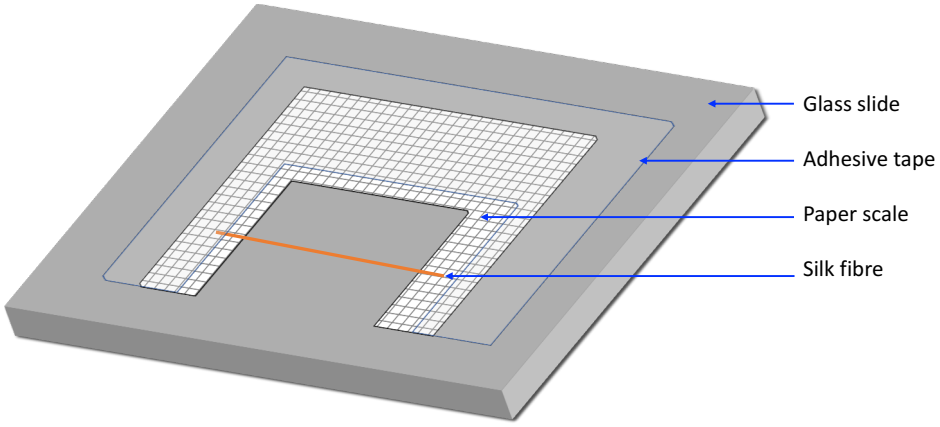


Fig. 3.6. *Schematic diagram showing the fibre mounted setup used for diameter estimation through optical microscopy.*

3.2.3.1. Optical microscopy of fibres

Mounted fibres were examined by optical microscopy for any apparent defects such as uneven thickness, damage, presence of multiple fibres etc. and discarded if there were any. For fibres without defects, utilizing an inverted optical microscope with a 10X ocular lens and a 10X objective lens (*Axio Observer A1; Carl Zeiss Canada, Toronto, ON*), three micrographs were taken along the long axis of each fibre, one near the middle and one near each end of the fibre. Using

ImageJ 1.48v (*National Institutes of Health; USA*), cross-sectional diameters were estimated at six separate places for each micrograph with a scale setting of 2.2003 pixel = 1 μm and averaged to determine fibre diameter ($\bar{d}$) and reported with standard deviation (SD).

3.2.3.2. Mechanical testing

Any material, when subjected to an applied force(s), undergoes deformation and the amount of deformation may be measured through mechanical testing. The deformation process is often represented in the form of a stress-strain curve (***Fig. 3.7.***), which may be analyzed to determine mechanical parameters including stress, strain, toughness, and elastic modulus (Young's modulus). Mounted fibres that cleared the quality check mentioned in section ***3.2.3.1.***, were characterized using mechanical testing. An *in-house* mechanical testing set-up in the Rainey lab is attached to a KD Scientific model100 series syringe pump (*Holliston, MA*) for constant pulling of the fibres (strain rate = 0.1 mm·s^{-1}).[278] Once a fibre was secured onto the clamp and syringe pump, the change in weight was determined as a function of displacement of the syringe pump using an X5105DU analytical balance (*Mettler Toledo*; *Greifensee, Switzerland*).

The applied force, F (mN) is calculated from the weight change, in accordance with Newton's second law of motion,

$$F = mg \tag{3.1.}$$

where m is the mass of the fibre (g), and g is the acceleration due to gravity (9.8 m·s^{-2}). Assuming a perfect cylindrical shape for the fibre, the cross-sectional area before the application of force, A_0, was estimated using the average fibre diameter as,

$$A_0 = \pi \left(\bar{d}/2\right)^2 \tag{3.2.}$$

The engineering stress, the maximum amount of stress a fibre can withstand before breakage may then be estimated at any given measured point on the stress-strain curve using the force and cross-sectional area estimated in ***Eqs. 3.1.*** and ***3.2.*** as,

$$\sigma = {}^{F}/_{A_0} \qquad \textbf{(3.3.)}$$

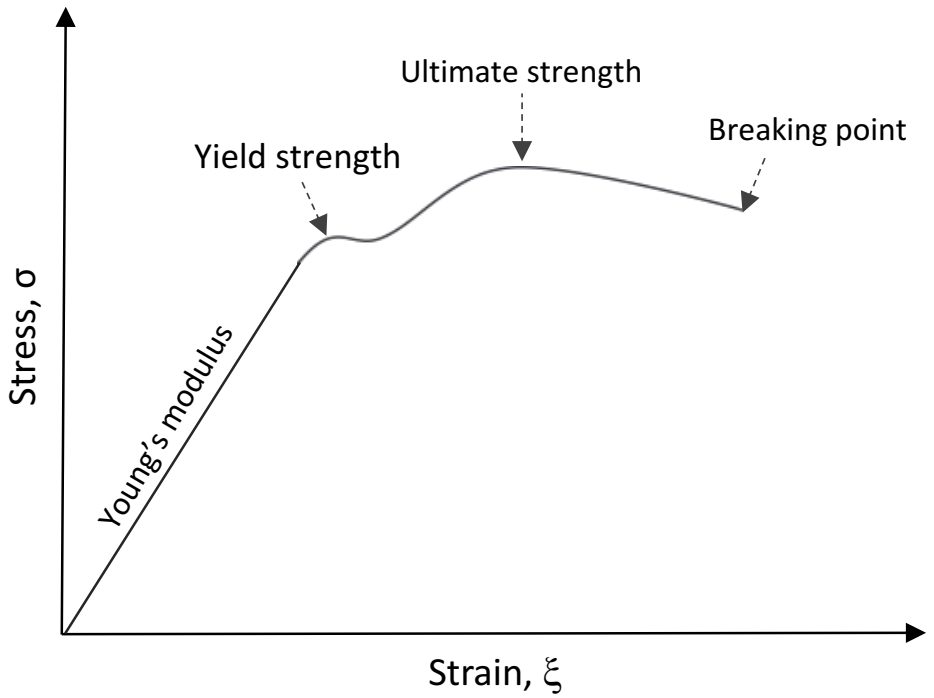


Fig. 3.7. *Generic stress-strain curve indicating major regimes of behaviour and key features.*

From the curve, the ultimate tensile strength (UTS) was extracted as the maximum stress observed prior to failure.[320] The extensibility of a fibre is determined by the maximum amount of strain it can withstand before breaking and is expressed in terms of ultimate engineering strain, ξ, given by

$$\xi = {}^{\Delta L}/_{L_0} \qquad \textbf{(3.4.)}$$

where ΔL is the total change in length and L_0 is the initial length of the fibre.[320]

The resistance to deformation or compressive stiffness of fibre is expressed in terms of Young's modulus, Y, given by: [320]

$$Y = {}^{\sigma}/_{\xi} \qquad \textbf{(3.5.)}$$

and is obtained by the slope of the stress-strain curve under elastic limits, where Hooke's law holds. The yield point corresponds to the point at which the maximum tensile stress a fibre can handle before the occurrence of permanent deformation is observed, i.e., a relatively large amount of deformation in response to a small increase in the stretching force. This is often used to characterize materials exhibiting an elastic behaviour. Finally, toughness is dependent on both the extensibility and strength of the fibre and is defined as the total energy per unit volume required to break the fibre, given by:

$$Energy/_{volume} = \int_0^{\xi_f} \sigma d\xi \quad \textbf{(3.6.)}$$

with ξ_f being the strain at breaking point, which is represented by the area under the stress-strain curve.

Each stress-strain curve dataset acquired using the Mettler Toledo instrument software LabX 2014® was exported to Microsoft Excel for further analyses. The cross-sectional area, applied force, engineering stress (MPa), and engineering strain (mm/mm or %) ($MJ \cdot m^{-3}$) were calculated for each fibre as per ***Eqs. 3.1. - 3.3.,*** while the corresponding Young's modulus (GPa) and toughness for each fibre were respectively determined from the slope of the elastic region and area of the plastic region in the stress-strain curve.

3.2.3.3. Surface morphology analysis

Scanning electron microscopy was employed for imaging the fibre surfaces, where an electron beam impinged upon the fibre at an angle interacts with the sample surface. Electrons interacting with surface atoms produce scattered signals in accordance with the surface topology of the material.[321] To facilitate this process, non-conductive samples (such as silks) are typically coated with metal particles to increase electron density and reduce charge buildup on the sample.

The electron microscopy imaging protocol from *Xu et. a.l*[278] was adapted, with modifications to perform SEM imaging on the fibre samples. Double-sided adhesive tape was placed on an SEM stub, from which thin parallel strips were carefully cut off using a fine blade, as shown in ***Fig. 3.8.*** Small lengthwise slices of fibres were gently removed from the collection roller and placed in between the adhesive tape-strips with ends pressed onto the tape at an angle of ~45° to suspend the fibres freely to avoid any morphological deformation/sample degradation resulting from contact with the adhesion tapes. Once secured onto the adhesive tape, fibres were coated with gold palladium particles using a Low Vacuum Coater Leica EM ACE200 (*Leica Microsystems Inc.; ON)* prior to acquiring SEM micrographs in secondary electron collection mode using a Hitachi Cold Field Emission S-4700 Scanning Electron Microscope.

High resolution SEM images were used to evaluate the morphological features and surface topography of the fibres. Apart from that, cross-sectional diameters at six separate locations were measured for each fibre using ImageJ 1.48v (*National Institutes of Health; USA*) and averaged to determine the fibre diameter under the assumption of cylindrical shape.

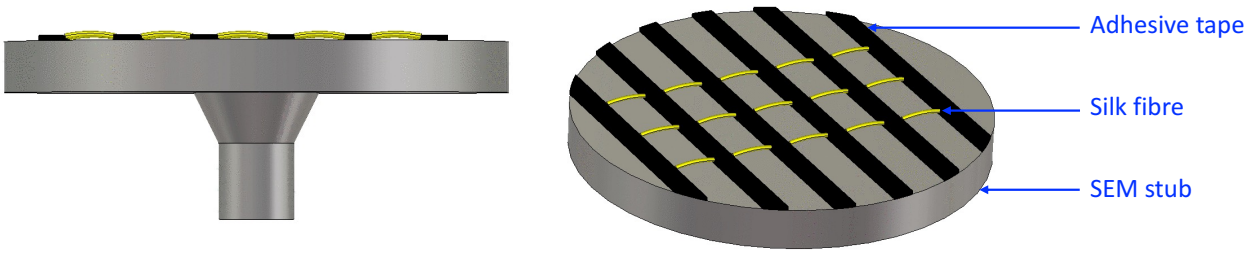


Fig 3.8. *Schematic representation of SEM stub preparation for imaging fibre samples.*

3.2.4. Characterization of dope solution of mutant silk

Based on our current understanding, the dope solvent is a critical precursor in the fibre forming process. The choice of dope solvent is key in deciding the surface morphology, packing and

mechanical properties of the fibres formed henceforth.[278] As a result, dope solutions of all the protein samples were subjected to a thorough examination. For characterization purposes, 1 mL dope solutions of the W_2 proteins (WT and mutants) at two concentrations - 30% (as in fibre spinning process) and 3% w/v were prepared, as detailed in section ***3.2.2.1***.

3.2.4.1. Estimating the size of self-assemblies in dope solution

3.2.4.1.1. Dynamic light scattering (DLS)

Dynamic light scattering (DLS) is a non-invasive technique used for measuring the particle size and size distribution of (colloidal) particles in a solution/suspension. Brownian motion of particles in solution leads to scattering of incident light with an intensity that fluctuates in time in response to particle diffusion and collision(s)[322, 323] Analyses of the resulting intensity fluctuations yield the velocity of Brownian motion and diffusion coefficient(s), from which, particle size(s) and – in the case of heterogeneous samples - particle size distributions for a given sample can be determined.[324-327] Assuming a spherical shape for the diffusing particle, the translational diffusion coefficient, D_C, is given by the Stokes-Einstein relationship, as:

$$D_C = \frac{k_B T}{6\pi\eta r} \tag{3.7.}$$

with k_B being the Boltzmann constant, T the absolute temperature, η the viscosity of the solution, and r the hydrodynamic radius of the diffusing molecule. The relative motion of particles is described in terms of a normalized electric field correlation function $G_1(\tau)$, that exponentially decays over time in response to the Brownian motion of the scattering particles, given by:[327, 328]

$$G_1(\tau) = e^{-\Gamma\tau} \tag{3.8.}$$

where Γ is the decay constant defined as

$$\Gamma = -D_C \mathbf{q}^2 \quad \textbf{(3.9.)}$$

with **q** being Bragg's vector. The resultant fluctuations detected for the scattered intensity are measured in terms of the correlation function, $G_2(\tau)$, given by [327, 328]:

$$G_2(\tau) = \frac{\langle I(t) I(t+\tau) \rangle}{\langle I(t) \rangle^2} \quad \textbf{(3.10.)}$$

with I(t) being the intensity at time t and I(t+τ) the intensity after t by a delay time τ.

DLS data are usually interpreted using fitting algorithms, with non-negative least squares (NNLS) being a well-established and common DLS fitting method. In this method, NNLS fitting is given by,[327, 329]

$$X^2 = \left[\sum_{j=1}^{N} g_1(\tau_j) - \sum_{i=1}^{M} b_i e^{-\Gamma_i \tau_j}\right]^2 \quad \textbf{(3.11.)}$$

where X is the intensity-weighted arithmetic and harmonic diameters, N is the number of data points, and M is the number of decay constants used, b_i is a series of coefficients set to fit the equation to the data. Since NNLS equally weights all data points, this fitting is susceptible to inaccuracies from factors like outliers, error and noise. An alternate fitting model, CONTIN, incorporates regularization constraints that are amendable as per the expected size distribution and enable rejection of oultiers etc., thus reducing fit sensitivity to noise and outliers.[330, 331] CONTIN fitting is given by:

$$X^2 = \sum_{j=1}^{N} \frac{1}{\sigma_i^2} \left[g_1(\tau_j) - \int G(\Gamma) e^{-\Gamma\tau} \right]^2 + \alpha^2 \|LG(\Gamma)\|^2 \quad \textbf{(3.12.)}$$

with α being the regularization parameter, L a regularizer operator, and σ the variance.

Sample heterogeneity is quantified in terms of polydispersity index (PDI) value which symbolizes the broadness of the particle size distribution. As nanoparticles within the spidroin dope are representative self-assemblies of a highly dynamic system whose size and size distribution

broadens over time, the PDI value is a valuable physical indicator of this dynamic nature of the protein in the dope state.

For DLS measurements, 3% (w/v) dope solutions of WT W_2, reduced and stapled W_2M were prepared as detailed in section ***3.2.2.1.*** Sample refractive indexes were measured at 25 °C using a Refracto 30GS refractometer (*METTLER TOLEDO; Columbus, OH, US*) and averaged over three samples.

700 µL of a given 3% dope solution was filtered with a 0.2 µm syringe filter and transferred to 1 cm quartz cell (*Hellma Analytics; Markham, ON, CA*) for DLS measurements using a BI-200SM instrument (*Brookhaven Instruments Corporation, Holtsville, NY*) at 25 °C with a wavelength of 637 nm. Through variation of scattering angle, 120 ° was determined to be the optimal angle at which to measure the DLS signal and the stability and behaviour of each W_2 protein dope solution was monitored as a function of time within 10 min of filtration, defining this as the 0 time-point, and again 30 mins later at 25 ºC using the BI-ISTW software with reference values for the viscosity (0.924 cP) for water at 25 ºC. Data analysis was performed using BI-ISTW software (*Brookhaven Instruments*) with CONTIN fitting.

3.2.4.2. Morphology of self-assemblies in dope solution

SEM imaging was applied for quantitative analysis of the morphology of self-assemblies, particle size estimation and distribution in dope solutions. 30% (as in wet spinning) and 3% (as in DLS measurements) dope solutions were produced as detailed in section ***3.2.2.1.*** 10-µL aliquots of each sample were deposited on EM grid and air dried, sputter coated with gold-palladium using a Low Vacuum Coater Leica EM ACE200 (*Leica Microsystems Inc.; ON)* at the Scientific Imaging Suite (*Sir Charles Tupper Medical Building, Dalhousie University*) with a diffuse coating method for

100 s and measured in secondary electron collection mode using a Hitachi Cold Field Emission S-4700 scanning electron microscope. Near spherical NPs were chosen while estimating the average particle size using ImageJ 1.48v (*National Institutes of Health; USA*). Two different locations with the same cross-sectional area were chosen in the SEM image, particles sizes estimated and those within the 10 nm size range were grouped, and the frequency of each size range was depicted in terms of size distribution histograms.

3.2.5. CD spectroscopic investigation of the secondary structure of the soluble protein

30 μM W_2 protein sample prepared in 20 mM sodium acetate (NaAc) buffer (pH 5 ± 0.05) were filtered with 0.2 μm syringe filters and the final protein concentration in each sample was determined using UV absorbance at 214 nm, as described in ***section 3.2.2.1***. CD spectra between 260 to 195 nm were recorded at 20 nm·min^{-1} in 0.1 nm intervals using a J-810 spectropolarimeter (*Jasco-Easton; MD, USA*) with a temperature control by Peltier system using 0.1 mm quartz cuvette (*Hellma; Müllheim, Germany*) as a preliminary probe of protein secondary structure. In order to determine the melting point of these proteins, spectra were collected from 20 °C to 84 °C in 10 ° increments from 20-60 ° and then in 2° increments. Three repeated scans were recorded for each spectral condition, averaged and blank corrected at each temperature, and data plotted using Microsoft Excel.

3.3. RESULTS

3.3.1. Recombinant protein production and purification

An open reading frame for the H_6-SUMO-W_2M fusion protein was constructed as a gene in a modified pET32 plasmid, transformed into BL21(DE3) *E. coli* cells, and expression was induced

using IPTG (lanes T, S and P in ***Fig. 3.9.***). The fusion protein recovered following Ni-NTA purification (lane E), was subsequently cleaved using SUMO protease (lane C). The desired W_2M protein was subsequently reverse purified using Ni-NTA affinity chromatography, as in the previously reported protocols for WT W_1.[5, 9] WT W_2 was produced in a similar manner. (***Fig. 3.10.***).

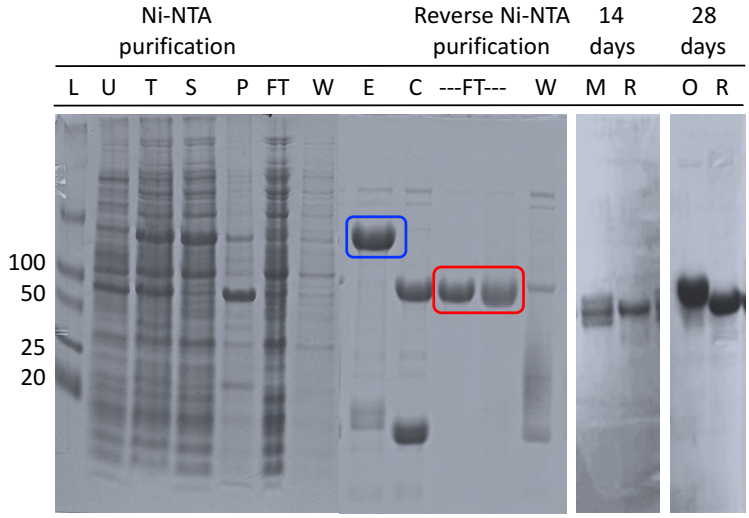


Fig. 3.9. *SDS-PAGE gel image showing the overexpression, representative purification (10% gel) and oxidation status of W_2M protein (15% gel). Lanes are defined as L: ladder, U: uninduced cell culture after 3 hours, T: total cell lysate, S and P: soluble and insoluble fractions of cell lysate, respectively, FT, W and E: flow-through, wash and elution from affinity purification, respectively, C: elution fraction after treatment with SUMO protease, M: W_2M with mixed sulphides, R and O: reduced and stapled forms of W_2M. the blue box highlights the fusion tagged W_2M recovered through affinity purification while the red box shows the tag-free W_2M recovered through reverse purification.*

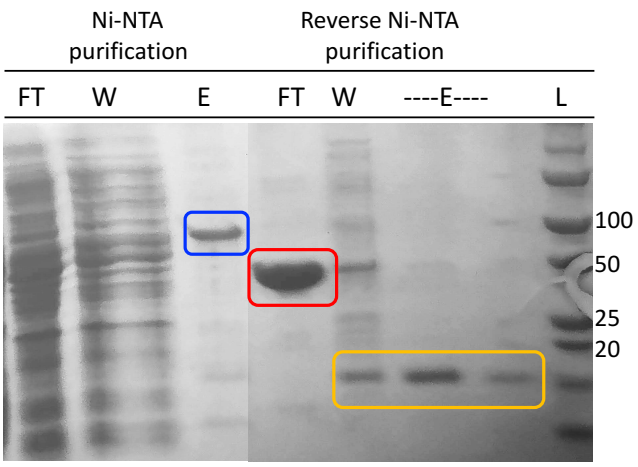


Fig. 3.10. *SDS-PAGE gel image showing representative Ni-NTA purification of WT W_2 protein (10% gel). Lanes are defined as FT, W and E: flow-through, wash and elution from affinity purification, respectively and L: ladder. Fusion tagged and tag-free WT W_2 are highlighted in blue and red respectively, and the H_6-SUMO fusion tag in yellow.*

Recovery of W_2M protein in high concentrations through reverse purification (RF), as is evident from the gel image (***Fig. 3.9.***), led to initial challenges in keeping the protein stable and ensuring the correct disulphide formation vs. incorrect, undesirable disulphides. 10× dilution with 50 mM PB (pH 8.0 ± 0.05) was employed to minimize the likelihood of intermolecular disulphide formation and, hence, overcome the potential obstacle of protein aggregation encountered at higher concentrations. Multiple attempts, tested over different timescales, were unsuccessful in accelerating the disulphide bonding in W_2M. Even though partial oxidation was observed in certain cases, prolonged incubation demonstrated elevated levels of protein degradation (***Table 3.1.***). I also tested the effect of oxygen pumped into the protein solution, which also failed to efficiently oxidize W_2M. While these strategies to accelerate the oxidation process failed, it was observed that the mutant protein was relatively more stable at pH 8.0 in 50 mM PB at 4 °C; hence, this condition was adopted for further oxidation trials. Self-oxidation was achieved in 28 days. The

absence of any other peak(s) in lane O indicates complete oxidation and electrophoretic migration of the stapled W_2M as a monomeric species.

Table 3.1.: TRIALS TO ESTABLISH DISULPHIDE BRIDGING IN W_2M PROTEIN				
Oxidation condition		**Incubation time (hrs)**		**Observation**
GSH:GSSG	1:1	24	48	partial oxidation; incomplete
	4:1	24	48	
	1:4	24	48	
Oxygen pumping		48	72	aggregation/degradation

3.3.2. Reduced mutant is highly prone to fibre formation

Hand pulling was successful with the W_2M protein fraction freshly obtained from reverse purification (***Fig. 3.11.),*** thus confirming the fibre forming ability of the mutant protein, following previous observations for WT W_2.[3] However, this technique was unsuccessful with the oxidized form of W_2M. This may have been a consequence of low protein concentration (as detailed in section ***3.2.2.***). To more robustly test for the capability of oxidized W_2M to form silk-like fibres, a 15 % dope solution was prepared with lyophilized recombinant proteins in an 8/1/1 solvent mixture (v/v/v) of TFA/TFE/H_2O to test for fibre formation at a higher concentration. Initial testing for fibre-forming ability of the dope solutions with 100% EtOH showed that fibres could be pulled out for both WT W_2 and reduced W_2M, while the oxidized W_2M failed to produce similar fibrillar structures. Instead, minuscule structures were observed at the bottom of the beaker containing EtOH (similar to those previously reported for WT W_1). The structures were very fragile to touch by tweezers and that, once taken out of EtOH, turned into fine powder. Hence these microfibrillar structures were retained in the EtOH bath, while fibres of WT W_2 and reduced W_2M were transferred to slides for imaging.

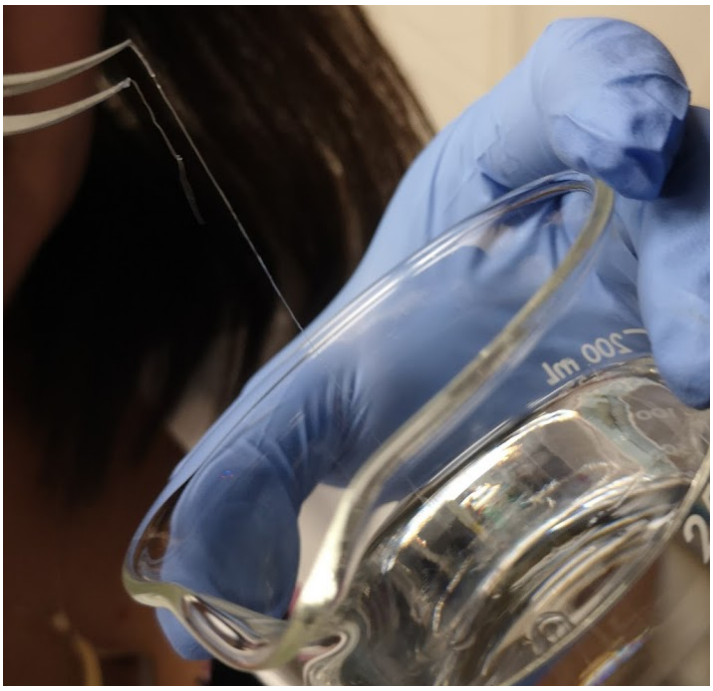


Fig. 3.11. *Fibre of reduced W_2M being hand-drawn technique from ethanol bath.*

Further uniform fibres were produced with 15% dope solution through wet spinning under controlled parameters (***Fig. 3.12.***) using the *in-house* fibre spinning apparatus (***Fig. 3.5.***). Reduced W_2M fibres were observed to be strong and smooth compared to WT W_2 fibres. As a result, post-spin stretching was carried out on the reduced W_2M, producing PS fibres under two stretching conditions: 2X stretched in air and 2X stretched in water. However, this could not be performed on WT fibres since they did not hold well for similar post-stretching.

However, it was very difficult to form fibres from oxidized W_2M even at 30% concentration in spinning dope as the structures formed in the coagulation bath was too thin to even have a shadow, and very brittle leading to breakage if touched with tweezers; hence, these could not be picked up with tweezers even when they could be observed. With this protein condition, minute species were deposited on the base of the EtOH bath which looked like microfibrils or linear aggregates, as was also observed for 15% dope (***Fig. 3.12.)***. Again, these were brittle to the touch and turned upon exposure to air. Multiple wet-spinning trials on different days produced the same outcome.

3.3.3. Characterization of mutant silk fibres

Fibres collected through wet-spinning (***Fig. 3.12**.)* were analyzed for their size, morphology and mechanical properties. 20 fibre samples prepared in each case (as in section ***3.2.3.1***.) were examined through optical microscopy and those with anomalies like breakage, multiple fibres, bulging etc. were omitted from mechanical testing *(*with some examples shown in ***Fig. 3.13**.)*. Independent samples were prepared for SEM imaging.

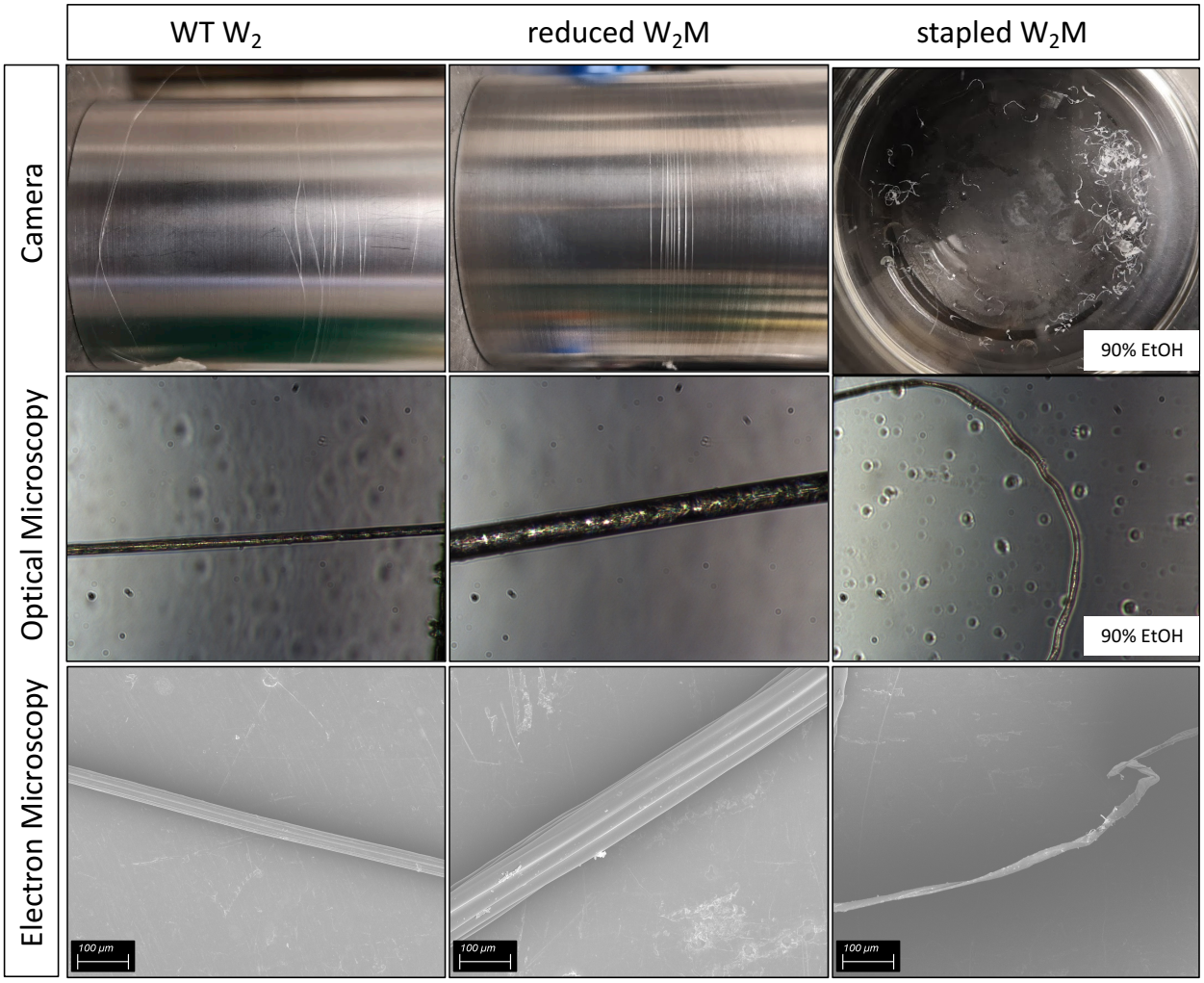


***Fig. 3.12**. Wet-spun silk fibres of WT W_2 and reduced W_2M and microfibrils of stapled W_2M imaged by camera (top row), optical microscope (middle row) and scanning electron microscope (bottom row; 100 X resolution; scale: 100 µm).*

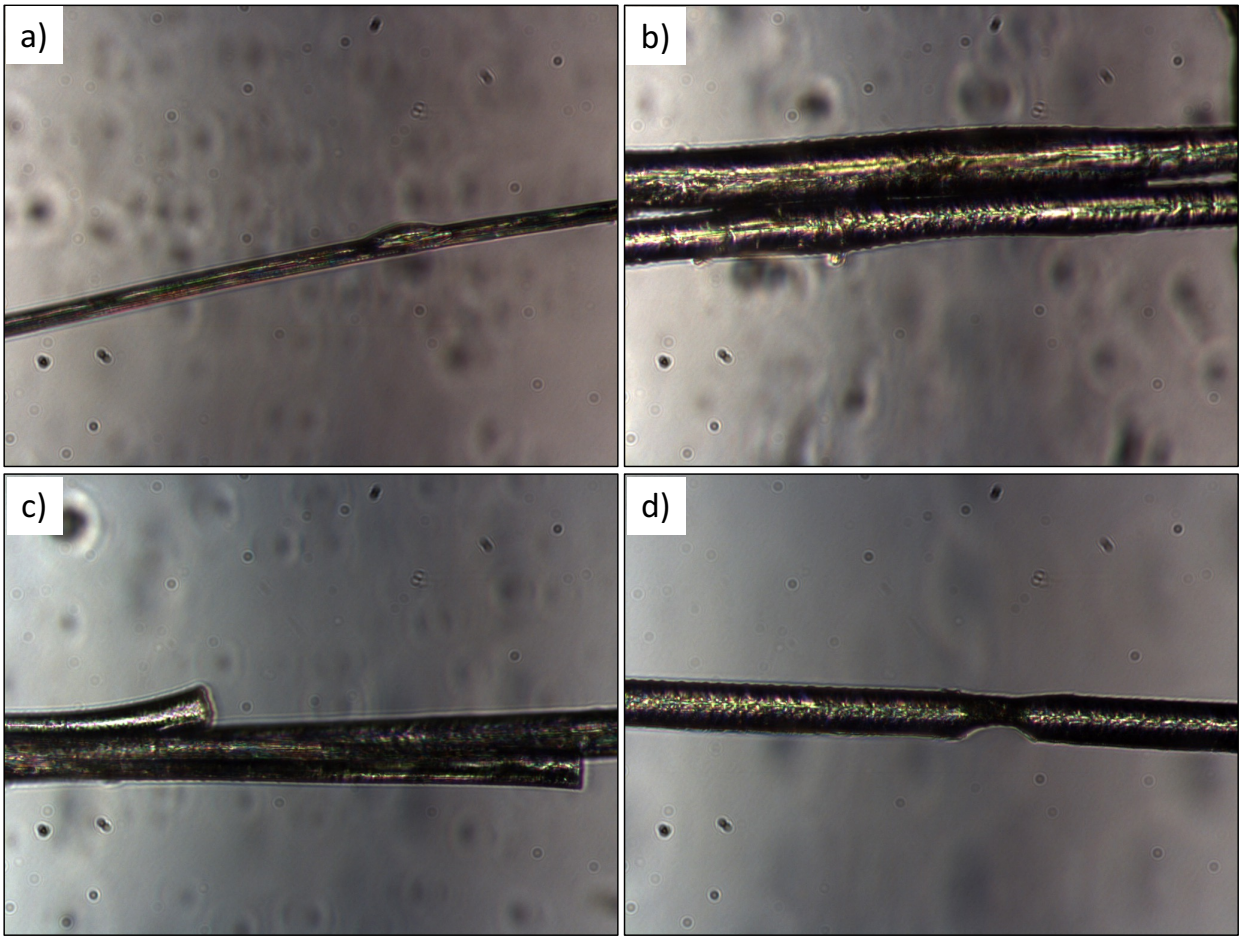


Fig. 3.13. *Examples of classes of anomalies detected in wet-spun fibres – a) bulging in WT W_2 b) double fibres, c) breakage and d) necking in reduced W_2M.*

3.3.3.1. Size of wet-spun fibres

Assuming a perfectly cylindrical shape, the average diameters of wet-spun fibres prepared under each spinning condition were determined using optical microscopy, as detailed in section ***3.2.3.1***. Comparing AS fibre state, it was found that reduced W_2M fibres were ~2× times thicker than the WT W_2 fibres (***Table 3.2.).*** The observed decrease in the diameter of the PS fibres of the reduced W_2M 2X stretched in the air following wet spinning is in accordance with the higher ordering and packing of microfibrillar assemblies through post-spin stretching treatments as reported for the W_2 and higher WT AcSp*1* protein. A further decrease in diameter is expected post stretching in water,

which was not evident in all cases (e.g., as in the case of the fibre of this class with the maximum observed). The second set of PS fibres – 2X stretched in water, annotated as PS_2X_H_2O stretched showed a broader range of values. Even though all fibres exhibit almost uniformly cylindrical shape, cross-sectional diameter varied from sample to sample, with fibre diameters greater than those for reduced AS also being reported (in ***Table 3.2.***, the maximum value for PS_2X_H_2O highlighted in red vs. minimum value for AS fibres of reduced W_2M). This anomaly in the size of PS_2X_H_2O fibres of reduced W_2M could be a consequence of the fibre swelling in response to water treatment i.e., water uptake by the fibre (detailed in ***chapter 6***). Whether this is a case of cyclic response to water or supercontraction needs to be further investigated.

Table 3.2.: DIAMETER OF RECOMBINANT *AcSp1* FIBRES WET-SPUN UNDER INDICATED CONDITION DETERMINED BY OPTICAL MICROSCOPY

Protein & Spinning Conditions		**Diameter (µm)***			**SD (µm)**	**# of fibres**
		Max	**Min**	**Average**		
WT W_2	AS	18.0	13.6	**16.1**	1.8	8
Reduced W_2M	AS	36.3	32.76	**35.1**	1.3	6
	PS_2X_Air	26.6	24.16	**25.5**	0.8	15
	PS_2X_H_2O	33.3	14.8	**26.0**	6.6	14

**Given the measurements are prone to human error, an error of ~1-2µm resolution/pixel should be considered for the reported values.*

3.3.3.2. Surface morphology of wet-spun fibres

Imaging through SEM revealed the morphological features of the wet-spun fibres with reduced W_2M being similar to its WT counterpart – cylindrical fibres with smooth and even surfaces were observed. The tiny microfibrils of stapled W_2M obtained from the EtOH bath turned powdery upon exposure to air and, hence, could not be analyzed with SEM. However, comparatively thicker microfibrils were also present in the bath, these were picked for SEM and imaged following the

standard protocol. Twisted microfibrils with flakey and uneven structural features similar to those reported by *Xu. et. al.*[278] were observed (***Fig. 3.14.).*** PS fibres of reduced W_2M imaged through SEM also showed smooth and even surface characteristics, similar to AS W_2M. (***Fig. 3.15.)*** and similar to previous images of W_3 post-spin stretched fibres spun from either the same spinning dope or from an HFIP: H_2O spinning dope.[278, 7]

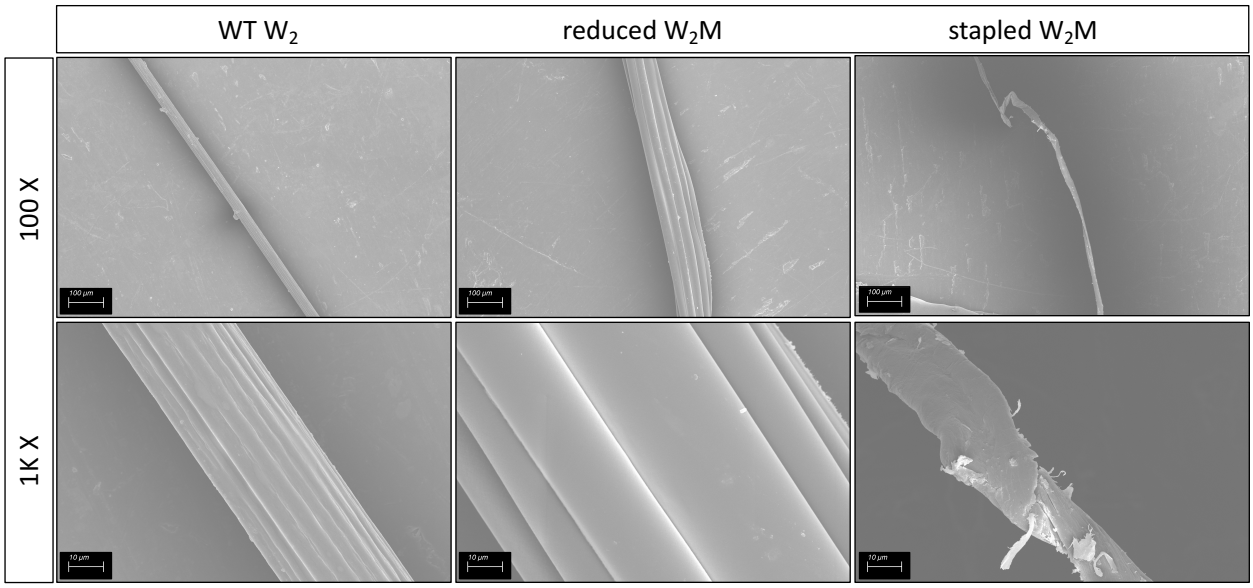


Fig. 3.14. *SEM images collected on silk fibres of a) WT W_2 and b) reduced W_2M, both wet-spun, and c) microfibrils of stapled W_2M at 100 X (top row; scale:100 μm) and 1K X (bottom row; scale:10 μm) resolution.*

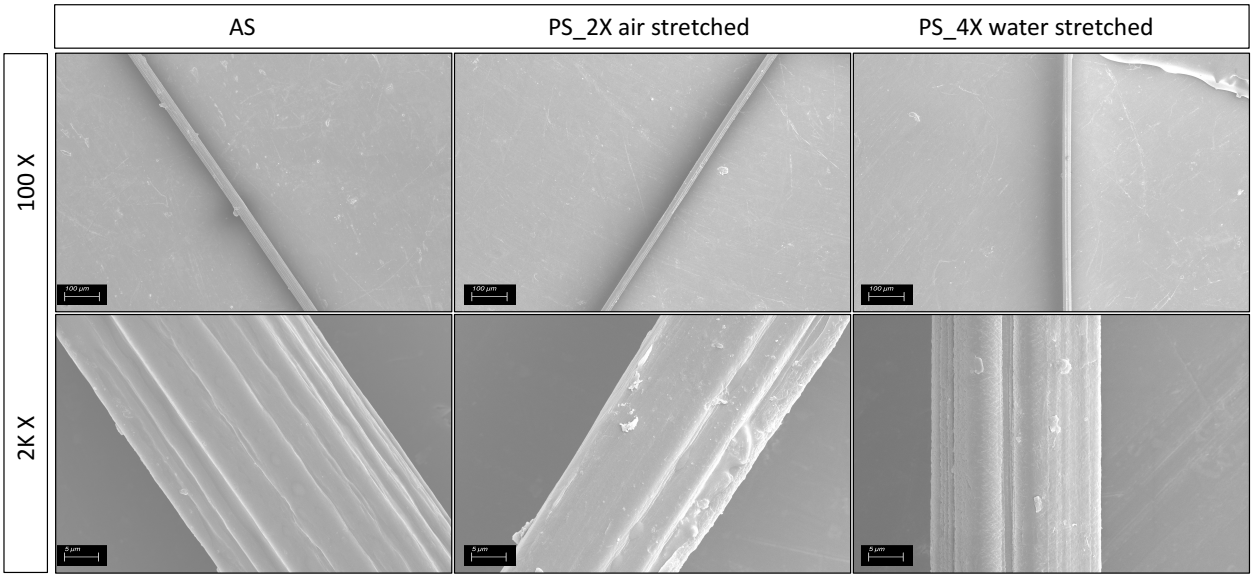


Fig. 3.15. *SEM images collected on a) AS b) PS_2X air stretched and c) PS_2X water stretched (2X stretching in the air followed by 2X stretching in water) fibres of reduced W_2M at 100 X (scale:100μm; top row) and 2K X(scale:5 μm; bottom row) resolution.*

3.3.3.3. Reduced mutant fibres are stronger than WT fibres

Mechanical testing and analysis of fibres without any visible anomalies was performed using an *in-house* setup and standard protocols.[7, 8] as detailed in section ***3.2.3.2.*** Stress-strain curves were thus determined **(*Fig. 3.16.*),** with an aggregate description of the results provided in ***Table 3.3.*** Comparing the AS fibres, the reduced W_2M fibres demonstrated a massive enhancement in their mechanical properties relative to WT W_2 fibres, with a ~15× increase in toughness and ~ 8× increase in strength.(***Fig. 3.16. a)*** and ***b)***).

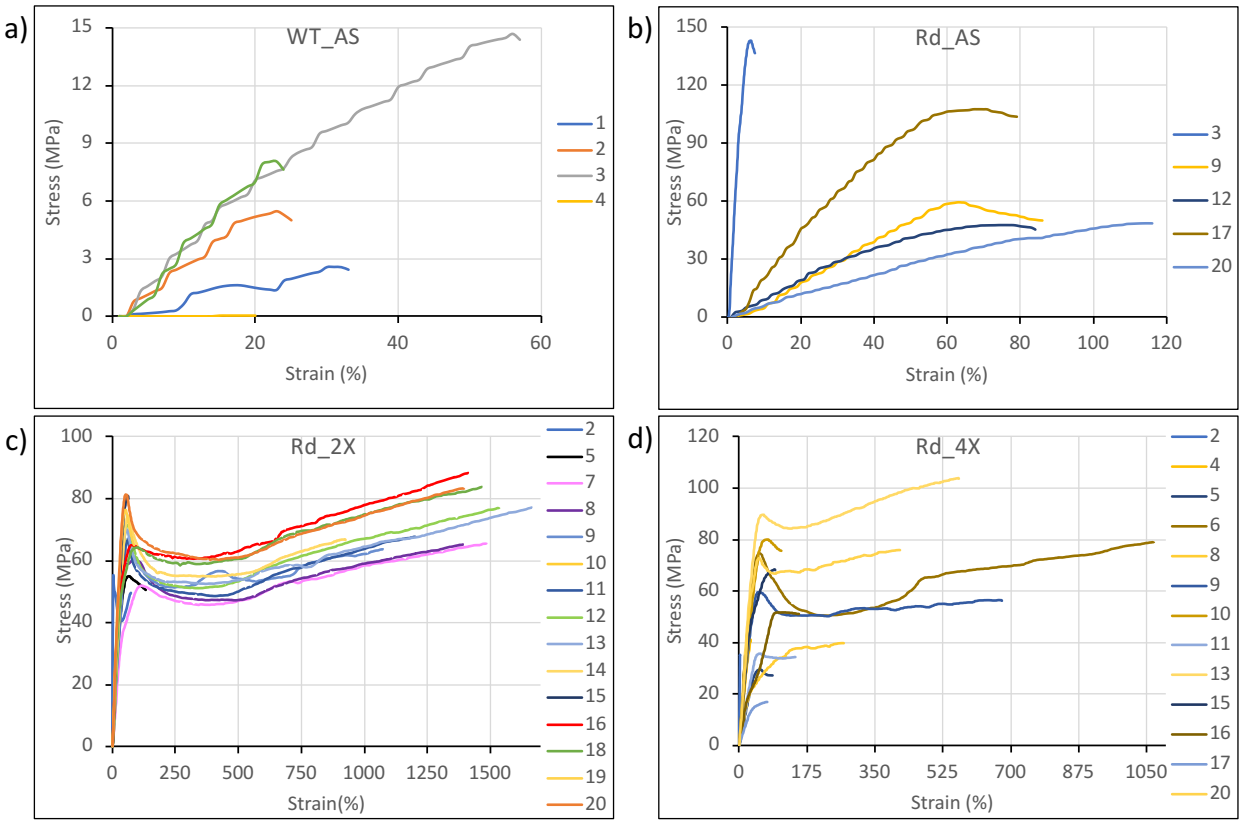


Fig. 3.16. *Stress-strain curves generated with the results obtained from mechanical testing of a) AS fibres of WT W_2 and b) AS, c) PS_2X air stretched and d) PS_2X water stretched fibres of the reduced W_2M.*

Testing of PS fibres of reduced W_2M in the same fashion demonstrated further improvement in mechanical properties through post-stretching treatments, following trends previously reported for *AcSp1* and pyriform silks.[7, 8,] While the 2× stretching in the air (i.e., reduced W_2M PS_2X_air fibre sample) enhanced fibre strength by ~15%, a ~30× increase in toughness was observed relative to AS reduced W_2M fibres (***Fig. 3.16.c)***). However, a peculiar scenario was witnessed in the case of PS_2X_H_2O stretched fibres, where a great deal of variability with respect to both stress and strain characteristics was observed (***Fig. 3.16.d)***). These discrepancies in mechanical properties could be arising from variability in individual fibre size;[332] that, in turn, may be indicative of underlying differences in the fibre state and processing.

Table 3.3. : MECHANICAL PROPERTIES OF W_2 FIBRES				
Sample ID	**Average**	**SD**	**Error**	**% Error**
a) WT_AS				
Engineering Stress (MPa)	7.4	5.1	0.7	69.2
UTS (MPa)	7.7	5.2	0.7	66.6
Strain ($mm \cdot mm^{-1}$)	0.03	0.02	0.44	44.2
Toughness ($MJ \cdot m^{-3}$)	0.2	0.2	1.2	117.1
Young's Modulus (GPa)	0.3	0.1	0.4	41.0
Diameter (µm)	16.6	1.8	0.11	10.9
b) Rd_AS				
Engineering Stress (MPa)	61.9	28.0	0.5	45.2
UTS (MPa)	65.8	28.4	0.4	43.2
Strain ($mm \cdot mm^{-1}$)	0.1	0.02	0.12	18.6
Toughness ($MJ \cdot m^{-3}$)	2.7	0.8	0.3	29.4
Young's Modulus (GPa)	1.1	0.5	0.4	43.4
Diameter (µm)	35.4	0.6	0.02	1.7
c) Rd_2X_Air				
Engineering Stress (MPa)	71.7	11.4	0.2	15.9
UTS (MPa)	73.2	9.9	0.13	13.5
Strain ($mm \cdot mm^{-1}$)	1.3	0.3	0.2	21.8
Toughness ($MJ \cdot m^{-3}$)	78.5	23.1	0.3	29.4
Young's Modulus (GPa)	1.9	0.4	0.2	20.9
Diameter (µm)	25.2	0.83	0.033	3.33
d) Rd_2X_Water				
Engineering Stress (MPa)	52.0	24.9	0.5	47.8
UTS (MPa)	53.1	25.1	0.5	47.2
Strain ($mm \cdot mm^{-1}$)	0.22	0.22	0.99	99.2
Toughness ($MJ \cdot m^{-3}$)	12.3	16.4	1.3	133.5
Young's Modulus (GPa)	1.44	0.6	0.4	42.6
Diameter (µm)	25.9	7.38	0.28	28.45

3.3.4. Disulphide lock triggers heterogeneity in silk nanoparticles

Based upon the dramatic differences observed in fibre "spinability" and mechanics, DLS measurements and SEM imaging were used for qualitative and quantitative investigation of the protein dope solution, respectively. This allowed me to test for differences in spinning dope pre-assembly that may be correlated to these differences in fibre formation and properties. In the case of a 30% dope, as employed in fibre spinning, the presence of too many scatterers made DLS infeasible. Hence, DLS measurements were performed with 3% dope solutions to avoid high concentration effects and, through systematic comparison, 120° was determined as the angle of observation. Data analysis was performed using CONTIN fitting.

At the 0 time point (defined at the start of the first DLS measurement period for a given freshly filtered dope solution), PDI values of 0.3 ± 0.05 were observed for both WT and reduced W_2M, while oxidized W_2M samples exhibited a higher PDI value of 0.7 ± 0.20, indicating a broader size distribution within the W_2M dope. Even though the formation of larger assemblies owing to the dynamic nature of the dope state and, hence, a time-dependent increase in mean particle diameter and PDI values was expected, the stapled W_2M dope demonstrated an enormous increase in particle size with the appearance of micron-sized particles without visible aggregation. These observations were consistently seen in multiple trials.

Based on these discrepancies between oxidized W_2M and reduced W_2M, SEM imaging was carried out with 30% dope solutions for detailed analysis. Particle sizes were determined over the same cross-sectional area and particle sizes within windows of 10 nm were grouped together to visualize size distribution histograms depicting the frequency of each size range (***Fig. 3.19.).*** Gaussian size distribution was observed for WT samples prepared either with or without DTT. Comparing the

NP size distribution in these samples, DTT does not seem to produce any significant perturbation to the self-assembly processes occurring within the dope solution of WT W_2 (***Fig. 3.17.).***

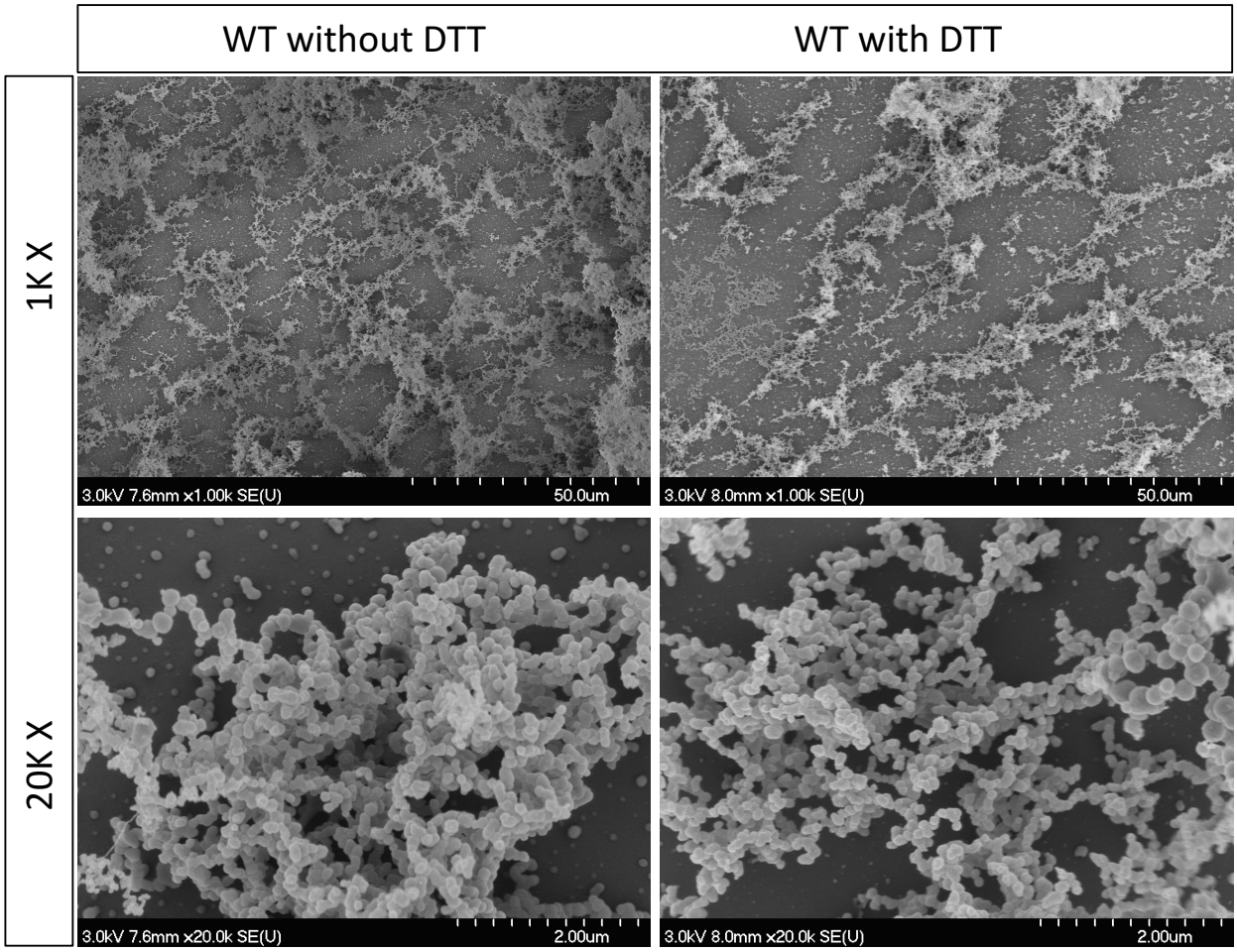


Fig. 3.17. *SEM images collected on 30% dope solution of W_2 proteins at a) 1K (top row) and 20 K (middle row) resolutions.*

The stapled W_2M exhibits a distinct pattern both in particle sizes and size distribution with a high frequency of particles in the range of 100-300 nm observed, consistent also with DLS measurements. (***Fig. 3.18.).*** However, the size distribution for the reduced W_2M is a positively skewed Gaussian with a higher frequency of smaller particles. Both the WT and reduced W_2M distribution patterns give a strong indication of the dynamic nature of the dope state, where higher-ordering is seeded by the initial NP-sized assemblies. Interestingly, a minor decrease in the

highest-frequency particle size was observed for the reduced W_2M relative to WT. Even though the addition of DTT was ineffective in the case of WT W_2, DTT retains the reduced state of Cys side chains in reduced W_2M by disulphide bonding with the -SH group. This side chain derivatization could have instigated local modifications within the monomeric protein that could manifest in the higher-order assembly into the NPs detected within the dope solution. Hence, I hypothesize that DTT is contributing to the minor changes observed in the size of reduced W_2M NPs relative to those of WT NPs, although I have no direct experimental evidence to demonstrate this. This requires further investigation.

In order to test for the comparability of these observations made at fibre-spinning concentrations to the DLS conditions, the same procedure was repeated at 10× dilution. Namely, SEM imaging was performed with the corresponding dope solutions at 3% (w/v) concentration. Similar particle sizes and size distribution patterns were observed at lower protein concentrations, thus confirming that the results detailed above are not affected by spinning dope concentration ***(bottom panel in Fig. 3.18.)***.

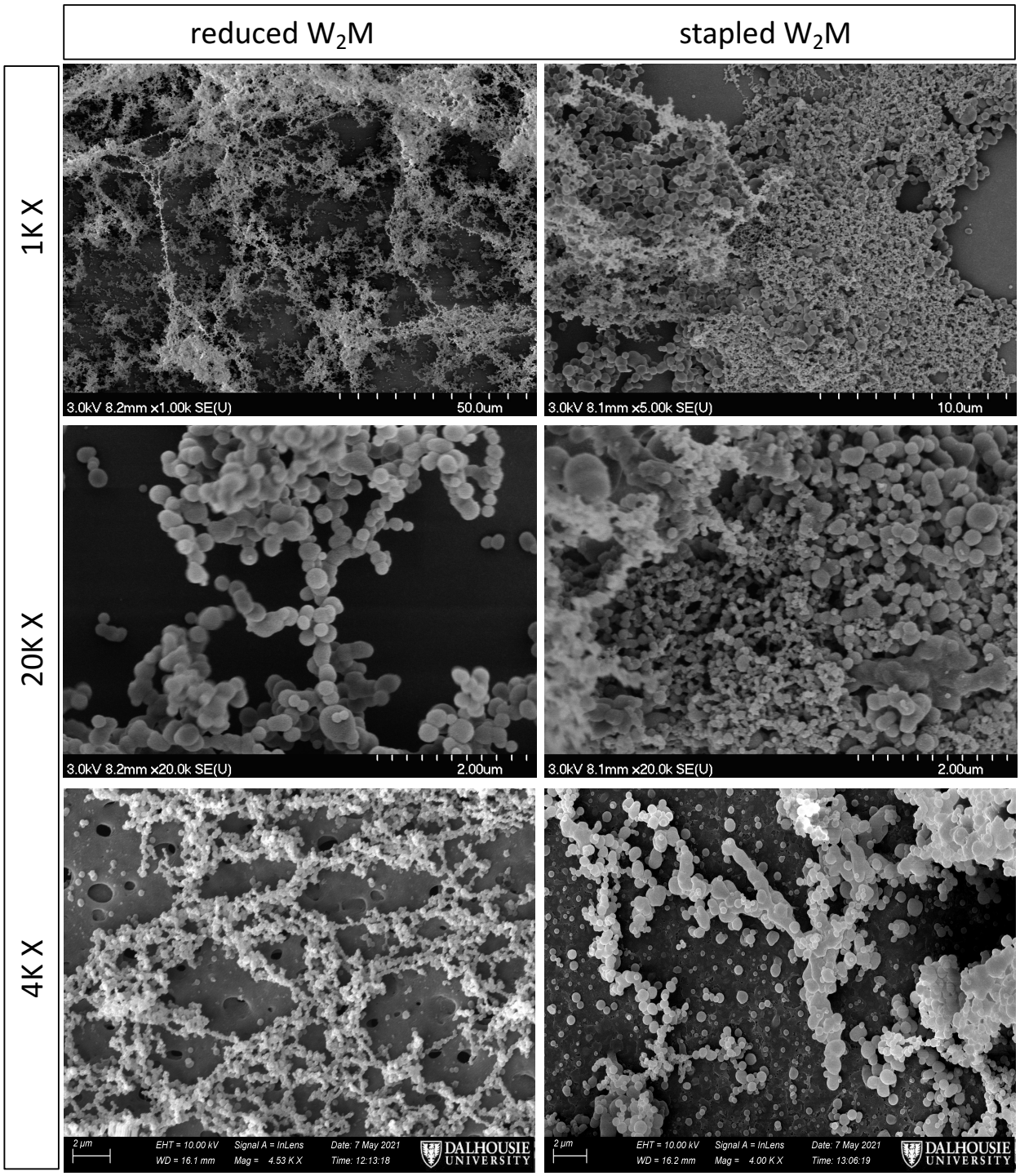


Fig. 3.18. *SEM images collected on 30% dope solution of reduced (left) and stapled (right) W_2M at 1K (top row) and 20K resolutions, and 3 % dope solution at 4K resolutions (bottom row).*

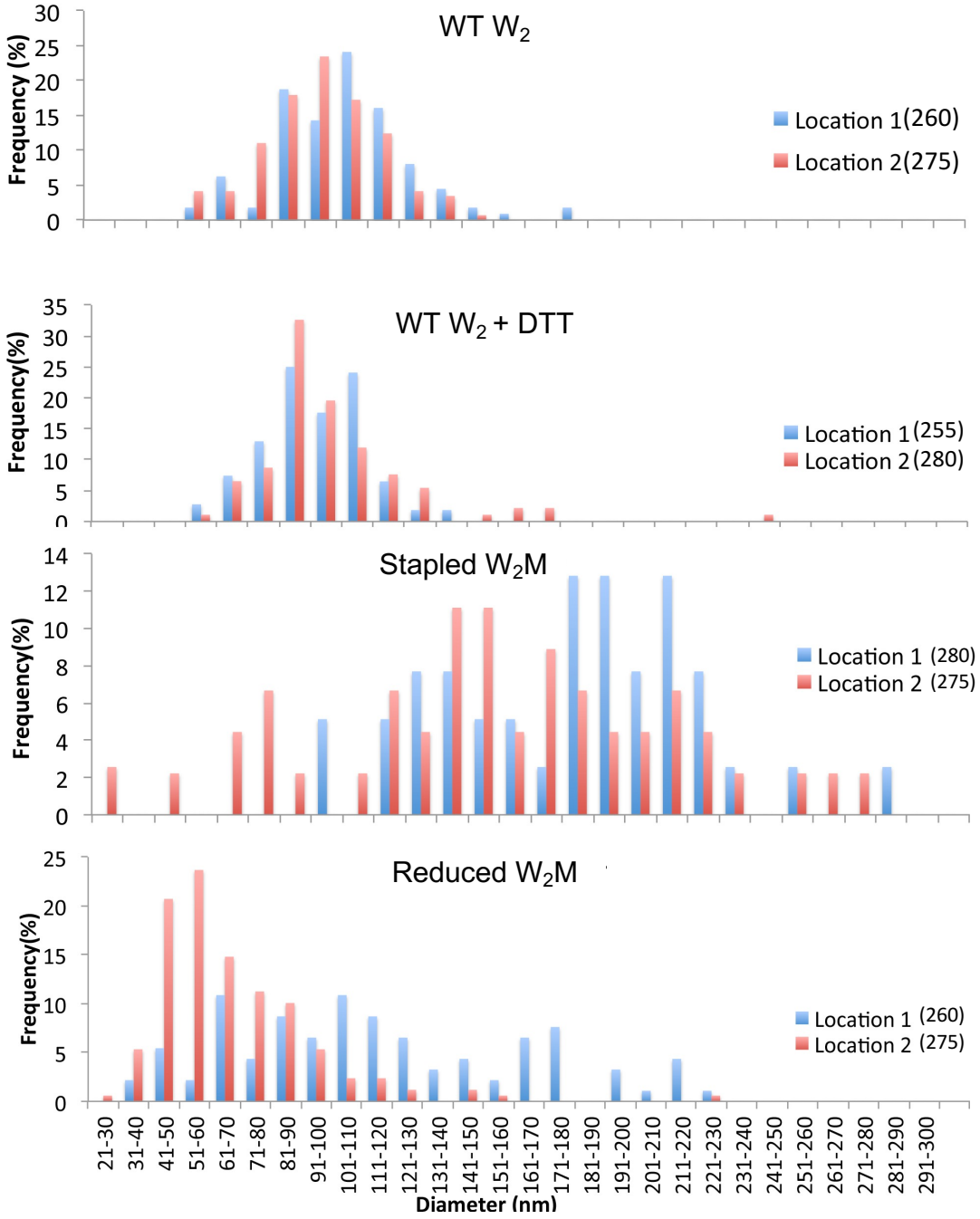


Fig. 3.19. *Particle size analysis from SEM images obtained on 30% dope solution of a) WT W_2 without DTT, b) WT W_2 (with DTT), c) stapled W_2M and d) reduced W_2M. Total number of near-to spherical particles measured in each instance is mentioned in brackets.*

3.3.5. Disulphide lock alters thermal stability of soluble protein

The differences that were observed in the NP assembly state in the dope directed me to question the behaviour of the mutant protein when in the solution state. Previous CD and NMR studies on the WT W unit proteins had established the solution-state structure of the protein and showed its modularity with respect to concatemers formed by increasing numbers of W units. In order to check for any deviations from the established secondary structure reported for WT, both forms of the mutant protein were investigated through far-UV CD spectroscopy.

In order to avoid interference of DTT in the CD signal and, subsequently probable spectral interference and inability to observe accurate protein CD spectra, 30 mM TCEP was employed to maintain the reduced state of W_2M. Preliminary CD spectroscopic studies carried out on the mutant samples prepared at uniform concentrations suggested minor differences in the secondary structure of stapled W_2M in comparison to the reduced form. An α-helix is, canonically, characterized by two negative bands of similar magnitude at 208 nm and 222 nm. Here, a relative decrease in the molar ellipticity is observed for the band at 222 nm in the case of stapled W_2M, suggesting a decrease in the α-helical content of the mutant protein when in the stapled form (***Fig. 3.20.)***.

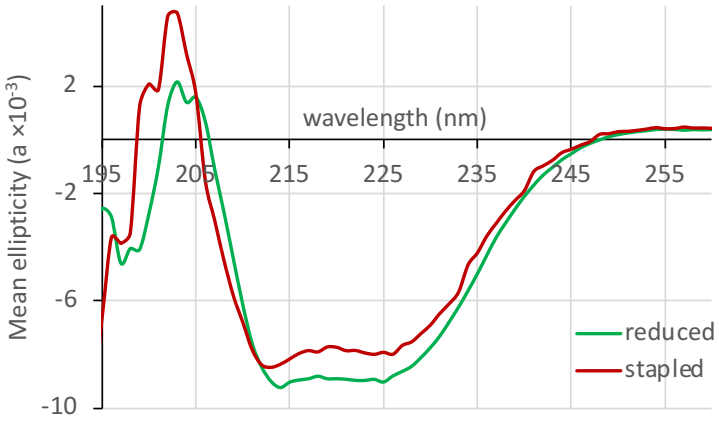


Fig. 3.20. *Far-UV CD spectroscopy collected on sasmples with same protein concentration shows secondary structural differences between the reduced and stapled forms of W_2M.*

Thermal denaturation studies carried out over the temperature range of 20-84°C on both forms of W_2M showed that the mutant protein had a melting point of ~74°C, slightly higher than the previously reported value of 71°C for WT W_2, indicating an improvement in thermal stability in both forms of the mutant protein. (***Fig. 3.21.)***.

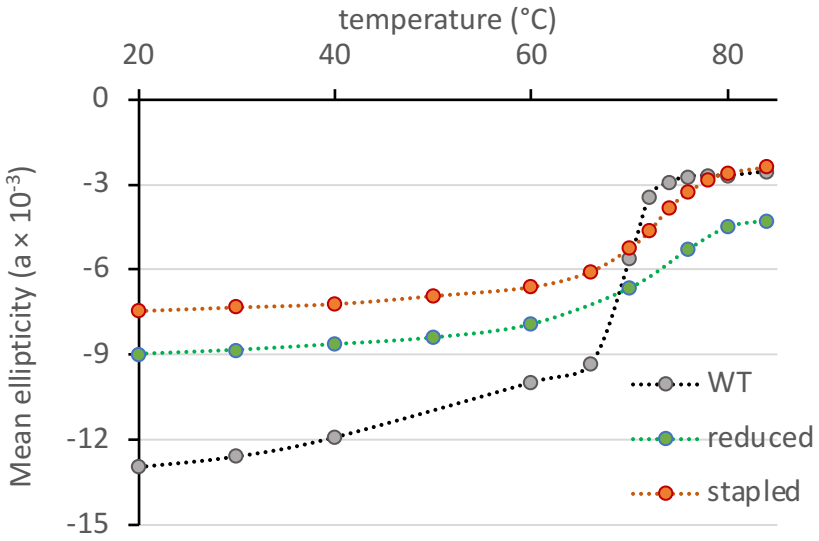


Fig. 3.21. *Thermal denaturation of mutant W_2 proteins depicts that both forms have a relatively higher melting temperature of 74 °C, with respect to the WT W_2 (71 °C).*

Notable aggregation (with micelle-like assemblies stuck to the cuvette walls that later degraded over time) was observed for the the stapled W_2M around 55 °C in contrast to the reduced W_2M, which remained soluble and stable throughout the experiment. Monitoring of CD spectral shape as a function of temperature while cooling to RT suggests the formation of two distinct conformational states for the stapled form when it is allowed to renature after thermal melting (***Fig. 3.22.)***. This implies a greater propensity for temperature-dependent secondary structural change and relative instability of the stapled form relative to reduced form. These results were confirmed with three different trials performed with a new batch of freshly prepared protein sample each time.

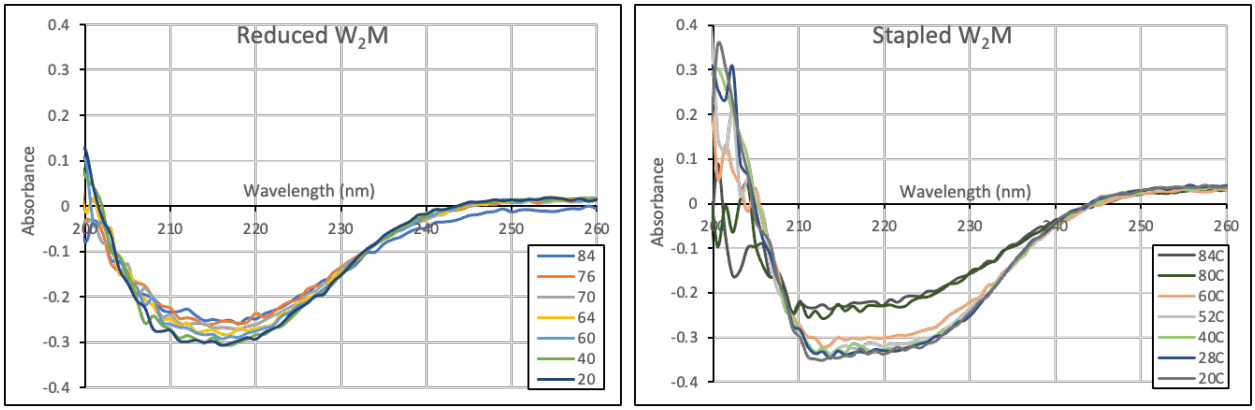


Fig. 3.22. *Far-UV CD spectra for refolding of reduced (left) and stapled (right) W2M during renaturation show at least two distinct secondary structures for the stapled form.*

3.4. SUMMARY

In studies on the W_3 protein, our group recently reported that there is a key role of the dope solvent in determining both self-assembled particle size and its homogeneity in the dope state, with corresponding changes to the morphology and mechanics of the fibres formed thereafter.[278] Here, I report similar observations for a single spinning dope condition through amino acid sequence modification that can be switched from a homogeneous, smaller particle state to a heterogeneous, larger particle state through disulphide bonding. Protein conformational changes were also implied at the ensemble-averaged level through the incorporation of one disulphide bond within each W unit.

Following the production of all proteins, attempts were made to produce fibres with appropriate and uniform protein concentrations. While the reduced form of the mutant (reduced W_2M) readily formed fibres, the stapled form (W_2M) was unable to form solid fibre-like assemblies. Instead, the formation of microscopic aggregates and nanofibrillar structures similar to those previously observed for W_1 - which also does not form silk-like fibres – were detected. My comprehensive probing of the dope state also revealed differences in the dope state precursor self-assembly. Not

only were these differences encountered in the size of the most probable self-assemblies but there were modifications in their size distribution patterns as well. Preliminary investigation with CD spectroscopy indicates minor secondary structural differences between these two conformational states.

W_2 is the smallest AcSp1 protein unit that displays fibre-forming ability. The observed drastic enhancement in mechanical properties, - toughness, engineering stress etc. – for reduced W_2M makes it the best reported fibre for a recombinant *AcSp1* protein with only two concatenated W units i.e., a W_2 protein. Though serendipitous, this intriguing discovery is promising for future bioengineering of novel and improved fibres. Improvements in mechanical properties following post-spin stretching treatment are observed for reduced W_2M (2X_Air stretch), (***Tables 3.2. - 3.4.***), which is consistent with our lab's previous reports for WT *AcSp1* and pyriform silk proteins.[7, 8,] To address the anomalies in fibre size and hence, mechanical properties of 2X_water stretched reduced W_2M, spinning and characterization of these fibres need to be repeated.

Pronounced differences between the two oxidation states of the mutant were consistently observed not only in the fibres and dope state but also in the soluble state of the protein with far-UV CD spectroscopy suggesting changes in the secondary structure. To further evaluate these changes, in the next two chapters I will evaluate the soluble protein at the atomic level to scrutinize the protein structure and dynamics at the residue level. These chapters report on a series of solution-state NMR studies carried out with this objective.

Chapter 4

EFFECT OF DISULPHIDE LOCK ON THE SECONDARY STRUCTURE OF THE AcSp1 W UNIT

4.1. INTRODUCTION

Inspired by the secondary structural difference observed between the oxidized and reduced W_2M proteins in CD spectroscopic studies discussed in the previous chapter, I decided to examine the soluble state at the molecular level - through to the atomic-level for more evidence to allow in the identification of factors affecting the fibre-forming ability of the -S-S- stapled mutant protein. Based on how the current understanding of native *AcSp1* proteins in the soluble form has been developed, as detailed in section ***3.1.***, I applied a similar approach to examine both the stapled and reduced forms of the mutant protein. Specifically, three aims were targeted in the course of the work described in this chapter:

i. test for changes in hydrodynamic size arising from the Ser →Cys mutations and the disulphide bond;
ii. probe the modularity of the W unit in both reduced and oxidized mutant proteins; and,
iii. identify the specific source(s) of secondary structural differences arising from the incorporation of a disulphide bond in the stapled form, if any.

Each of these aims was approached using solution-state NMR spectroscopy. Aim (i) relied upon hydrodynamics characterization through ^{1}H NMR spectroscopy-based experiments. Aims (ii) and (iii) employed triple-resonance NMR spectroscopy, in a similar approach to the soricidin studies detailed in ***Chapter 2***. For these latter aims, the mutant forms of W_1 and W_2 (i.e., W_1M and W_2M) had to first, therefore, be produced with ^{13}C- and/or ^{15}N-labelling, purified in sufficient quantity

(e.g., multi-milligram), and – for the stapled form - oxidized. The production of isotope-enriched proteins is first described, followed by each of the three aims.

4.2. MATERIAL & METHODS

4.2.1. Recombinant expression and purification of mutant proteins

[U-^{13}C,^{15}N] W_1M and W_2M were prepared as previously discussed in sections ***2.2.2***. and ***2.2.6.*** Reverse-purified WT W_1 was exchanged to 20 mM acetate buffer (20 mM NaAc, 5 mM DTT, pH 5.0 ± 0.05) through dialysis (3×) using 6.5 kDa cut-off dialysis tubing (*Bio Design Inc.*, *Carmel Hamlet, NY*) and stored at 4 °C. W_1M and W_2M proteins were expressed and purified in a similar fashion. Following protein recovery through reverse purification, reduced and oxidized conformers of W_2M were prepared as detailed in ***3.2.1***. The same protocol was adopted for W_1M - namely, production of the reduced mutant by the addition of DTT (***3.2.1.2)*** and the oxidized conformer through self-oxidation (***3.2.1.3)***, respectively. 15% SDS-PAGE gel was employed to qualitatively confirm disulphide bond formation in the oxidized mutants. All mutant proteins were then concentrated and exchanged to 20 mM acetate buffer (20 mM NaAc, pH 5.0±0.05), in the same manner as described for the WT W_1, and stored at 4 °C until used.

4.2.2. NMR sample preparation

WT and reduced mutant samples with ^{13}C- and/or ^{15}N-isotope labelling were exchanged to 20 mM acetate buffer (deuterated acetate (CD_3COO^-), 5 mM DTT, 1 mM DSS, 1 mM NaN_3) in 9/1 (v/v) H_2O/D_2O at pH 5 ± 0.05 (without correction for effects of deuterium) NMR samples of oxidized proteins were prepared in a buffer without DTT, but that was otherwise identical.

Protein solutions were decanted, filtered through a 0.45 μm syringe filter (*Millipore, Billerica, Massachusetts*) and concentrated using centrifugal filter units (*Ultracel-3K Amicon Ultra; Millipore, Ireland*). The final protein concentration in each instance was calculated as in ***Eq. 2.1.*** using the UV absorbance measured at λ = 210 nm, with ε values[187] for W_1 and W_2 proteins estimated as 270,858 $M^{-1} \cdot cm^{-1}$ and 543, 596 $M^{-1} \cdot cm^{-1}$ respectively. The sample was transferred into a 5 mm Shigemi susceptibility-matched NMR tube (*Bruker Biospin; Fällanden, Switzerland*) for data collection.

4.2.3. Pulsed field gradient (PFG) diffusion NMR spectroscopy

Pulsed field gradient (PFG)-based ^1H diffusion ordered spectroscopy (DOSY)[333] experiments were performed using freshly prepared ^{15}N-labelled reduced and oxidized W_1M proteins (20 mM CD_3COO^-, H_2O/D_2O v/v (9/1), 1 mM DSS, 1 mM NaN_3; pH 5.0 ± 0.05; 303.15 K). Both reduced and oxidized protein samples were specifically prepared at the same concentrations (560 ± 20 μM) to avoid potential discrepancies arising from concentration differences. Data were collected using an 18.8 T Avance III HD NMR spectrometer (*Bruker Canada; QANUC, Montreal, QC*).

PFG-based 2D ^1H-DOSY[333] (64 scans, sweep width 12 ppm, relaxation delay of 2 s incorporating pre-saturation) experiments were acquired, employing a stimulated echo and longitudinal eddy current delay (LED; 5 ms) with bipolar gradient pulses for diffusion and two spoil gradients (standard Bruker TopSpin pulse program ledbpgppr2s).[334] The pseudo 2D data collection for DOSY involved a series of 1D ^1H-^{15}N HSQC experiments carried out with increasing diffusion gradient amplitude (F2). The envelope of ^1H signals was attenuated by increasing the gradient strength from 2% to 95% in 16 linear steps. The observed peak intensity was fitted as a function of variable gradient strength was used to carry out a single component exponential fit and

determine the translational diffusion coefficients, D_c($m^2 \cdot s^{-1}$), for each diffusing species in a given DOSY experiment were determined from the fit using the Stejskal-Tanner formula[335] modified as appropriate for the chosen PFG experiment[334] using DynamicsCenter2.7.2 in TopSpin 4.0.6 (Bruker):

$$I = I_0 \cdot \exp[-D_c(g\gamma_H\delta)^2 \cdot (\Delta - \delta/3) \times 10^4] \qquad (\mathbf{4.1.})$$

where I is the observed peak intensity of a given resonance at a given value of gradient strength, I_0 is the intensity of the resonance at 2% gradient strength, D_c is the diffusion coefficient for the species giving rise to the given NMR signal, g is the gradient strength, γ_H is the gyromagnetic ratio of ^{1}H (4257.7 $Hz \cdot G^{-1}$), δ is the gradient length (3 ms), and Δ is the diffusion time (100 ms). Error estimation was carried out as RMS per spectrum (or trace/plane). Diffusion coefficients were obtained by integrating two to three separate regions of the amide backbone in the DOSY ^{1}H NMR spectrum. DOSY without water suppression was separately performed using two NMR buffers - with and without DTT to check for the effect of DTT and/or the resultant changes in viscosity if any. Effects on viscosity were minimal, and therefore used the inferred viscosity of 90/10 water/D2O at 303.15 K in all cases. Apart from that viscosity effects arising from the preferential hydrogen bonding differences of the sample, if any, were probed separately on both protein samples utilizing DOSY collected without water suppression.

4.2.4. Triple resonance solution-state NMR experiments

4.2.4.1. Data collection

NMR experiments were acquired on an 18.8 T Avance III HD NMR spectrometer equipped with TCI cryoprobe (*Bruker; QANUC, Montreal, QC*) at 303.15 K. Similar to the approach employed for soricidin (***Chapter 2***) and following previous studies carried out on the WT W unit,[9, 275,] a

series of triple-resonance 3D experiments were acquired on reduced and oxidized W_1M and W_2M proteins to allow for sequential backbone assignment (***Table 4.1.***). 2D ^{1}H-^{15}N HSQC experiments (employing ^{13}C-decoupling for double-labelled samples) were also routinely acquired, providing an anchor for 3D data assignment, a rapid means of sample-to-sample comparison, and a straightforward means of tracking sample stability and longevity.

4.2.4.2. Spectral processing and data analysis

^{1}H frequencies were referenced to DSS at 0 ppm while ^{13}C and ^{15}N were indirectly referenced to the ^{1}H zero-point DSS frequency.[214] All datasets were processed using NMRpipe[212] and the sequential assignment of the backbone chemical shifts[215](Cα, Cβ and C'), DANGLE predictions, and calculation of CSI and secondary chemical shifts (Δδ) were executed using CcpNmr Analysis 2.2.1.[213] CSD plots were generated to compare each mutant form to the WT, and identification of residue-level change(s) evoked in the protein backbone in response to the introduction of a disulphide bond in the stapled mutant.[336] CSD values were determined for the backbone amide and carbon resonances utilizing the assigned chemical shift values (in ppm).

Table 4.1: DETAILS OF PARAMETERS FOR 2D/3D NMR EXPERIMENTS						
Experiment *Pulse Program*	**recycling delay(s)**	**# of scans**	**Nus %; Amount**	**Dimension**	**# of complex points**	**Sweep width (ppm)**
a) WT W_1 (200 µM)						
^{1}H-^{15}N HSQC *fhsqcf3gpph*	1	4	50; 50	^{1}H ^{15}N	2048 200	16.020 24.000
HNCO *hncogpwg3d*	1	16	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.020 24.000 11.000
HNCACO *hncacogpwg3d*	1	64	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.020 24.000 11.000
HNCA *hncagpwg3d*	1	16	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.020 24.000 30.000
HNCOCA *hncocagpwg3d*	1	16	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.020 24.000 30.000
CBCACONH *cbcaconhgpwg3d*	1	16	12.5; 613	^{1}H ^{15}N ^{13}C	2048 90 218	16.020 24.000 80.000
HNCACB *hncacbgpwg3d*	1	16	12.5; 720	^{1}H ^{15}N ^{13}C	2048 90 256	16.020 24.000 80.000
b) Oxidized W_1M (400 µM)						
^{1}H-^{15}N HSQC *fhsqcf3gpph*	1	4	50; 50	^{1}H ^{15}N	2048 200	16.020 23.000
HNCO *hncogpwg3d*	1	16	12.5; 352	^{1}H ^{15}N ^{13}C	2048 88 128	16.020 23.000 11.000
HNCACO *hncacogpwg3d*	1	64	12.5; 352	^{1}H ^{15}N ^{13}C	2048 88 128	16.020 23.000 11.000
HNCA *hncagpwg3d*	1	16	12.5; 352	^{1}H ^{15}N ^{13}C	2048 88 128	16.020 23.000 30.000

Experiment *Pulse Program*	**recycling delay(s)**	**# of scans**	**NUS %; Amount**	**Dimension**	**# of complex points**	**Sweep width (ppm)**
HNCOCA *hncocagpwg3d*	1	16	12.5; 352	^{1}H ^{15}N ^{13}C	2048 88 128	16.020 23.000 30.000
CBCACONH *cbcaconhgpwg3d*	1	32	12.5; 523	^{1}H ^{15}N ^{13}C	2048 88 190	16.020 23.000 70.000
HNCACB *hncacbgpwg3d*	1	32	12.5; 704	^{1}H ^{15}N ^{13}C	2048 88 256	16.020 23.000 70.000
c) **Reduced W_2M (250 µM)**						
^{1}H-^{15}N HSQC *fhsqcf3gpph*	1	4	50; 50	^{1}H ^{15}N	2048 200	16.020 24.000
HNCO *hncogpwg3d*	1	16	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.020 24.000 11.000
HNCACO *hncacogpwg3d*	1	32	25; 5	^{1}H ^{15}N ^{13}C	2048 1 40	16.020 24.000 11.000
HNCA *hncagpwg3d*	1	16	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.020 24.000 30.000
HNCOCA *hncocagpwg3d*	1	16	12.5; 360	^{1}H ^{15}N ^{13}C	2048 90 128	16.020 24.000 30.000
CBCACONH *cbcaconhgpwg3d*	1	32	12.5; 613	^{1}H ^{15}N ^{13}C	2048 90 218	16.020 24.000 80.000
CBCANH *cbcanhgpwg3d*	1	16	25; 16	^{1}H ^{15}N ^{13}C	2048 1 218	16.020 24.000 80.000
HNCACB *hncacbgpwg3d*	1	16	25; 720	^{15}N ^{13}C ^{13}C	90 218 128	24.000 80.000 80.000

d) **Oxidized W_2M (200 µM)**						
Experiment *Pulse Program*	**recycling delay(s)**	**# of scans**	**Nus %; Amount**	**Dimension**	**# of complex points**	**Sweep width (ppm)**
^{1}H-^{15}N HSQC *fhsqcf3gpph*	1	8	50; 50	^{1}H ^{15}N	2048 200	16.020 23.000
HNCO *hncogpwg3d*	1	32	12.5; 352	^{1}H ^{15}N ^{13}C	2048 88 128	16.020 23.000 11.000
HNCA *hncagpwg3d*	1	32	12.5; 352	^{1}H ^{15}N ^{13}C	2048 88 128	16.020 23.000 30.000
HNCOCA *hncocagpwg3d*	1	32	12.5; 352	^{1}H ^{15}N ^{13}C	2048 88 128	16.020 23.000 30.000

All experiments were performed at 303.15 K.

Combined CSD values for the backbone amide (HN) resonances were calculated as

$$\overline{d_{NH}} = \sqrt{[\Delta(\delta H)]^2 + 0.2[\Delta(\delta N)]^2 / 2} \qquad \textbf{(4.2.)}$$

with $(\delta H) = \delta H_m - \delta H_{wt}$, and

$$(\delta N) = \delta N_m - \delta N_{wt}$$

where $\delta H_m$and δH_{wt} are chemical shift values (in ppm) of amide protons, and δN_m and δN_{wt} are chemical shift values (in ppm) of amide nitrogens, with the subscript indicating mutant (m) or WT (wt) protein.CSD values for the backbone and Cβ carbons were calculated to demonstrate changes in secondary structuring, as

$$d_C = \sqrt{(\delta C_m - \delta C_{wt})^2} \qquad \textbf{(4.3.)}$$

where, δC_m and δC_{wt} correspond to the chemical shift (CS) values of Cs (C**α**, Cβ or C') in mutant and WT proteins, respectively.

4.3. RESULTS

4.3.1. Recombinant protein production, purification and oxidation

All three proteins – WT W_1, W_1M and W_2M with H_6-SUMO fusion tags – were successfully overexpressed in *E. coli* in M9 media and ^{13}C and/or ^{15}N-isotope labelled (confirmed through NMR). Each H_6-SUMO-W fusion protein was purified using separate Ni-NTA affinity columns (to avoid cross-contamination) and the H_6-SUMO fusion tags were efficiently removed through SUMO protease-mediated proteolysis (lane C), allowing reverse purification of the desired W unit (lane FP). (***Fig. 4.1.a.***).

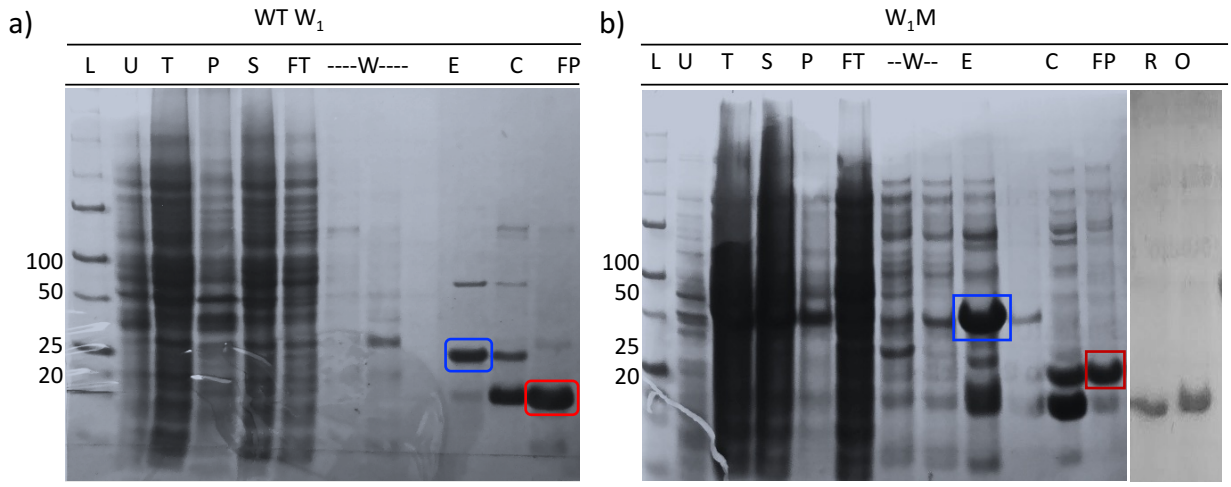


Fig. 4.1. *SDS-PAGE gel image showing representative purification of a) WT W_1, and b) W_1M proteins (10% gel) and oxidation status of W_1M (15% gel). Lanes are defined as L: ladder, U: uninduced cell culture after 3 hours, T: total cell lysate, S and P: soluble and insoluble fractions of cell lysate, respectively; FT, W and E: flow-through, wash and elution fractions from affinity purification, respectively, C: elution fraction after treatment with SUMO protease, FP: flow-through fraction from reverse purification, M: W_1M with mixed sulphides, R and O: reduced and stapled forms of W_1M. The blue box highlights the fusion-tagged W_1M while the red box shows the tag-free W_1M.*

Following reverse purification, half of each sample of mutant W protein was treated with DTT while half was allowed to undergo self-oxidation. Specifically, to maintain the reduced state, 0.5 mM DTT was added to prevent disulphide bonding, probable at high protein concentration as in the flow-through and, thus, allowed maintenance of the reduced state of the Cys residues over extended periods (lane R). Self-oxidation was achieved following sample dilution and incubation at 4 °C at pH 8.0 (discussed in section ***3.2.1.3.).*** Unlike W_2M which took 28 days (***Fig. 3.5.)***, oxidation was achieved in 14 days for the W_1M (***Fig. 4.1.b)***. The absence of any other peak(s) in lane O indicates complete oxidation and electrophoretic migration of stapled W_1M as a monomeric species, as in the case of stapled W_2M. (***Fig. 3.5.)***

4.3.2. Hydrodynamic size of mutant proteins

Pulsed field gradient-based diffusion NMR experiments independently collected on two NMR buffers devoid of proteins - one with and the other without DTT - and water DOSY (i.e., DOSY without water suppression and optimized to observe the water signal) performed on both the reduced and oxidized W_1M samples, prepared in their respective NMR buffers, showed similar values. This served to confirm that DTT was not modulating solution viscosity and demonstrated a lack of difference in sample viscosity between the two states of the mutant protein. DOSY performed on the individual W_1M samples demonstrated that, irrespective of the state of disulphide bridging, the mutant protein assumes a similar hydrodynamic size in both conformational states with a translational diffusion coefficient of ~ 0.83×10^{-10} $m^2\cdot s^{-1}$ (RMS error: ±0.029). (***Fig. 4.2.)*** For comparison, a translational diffusion coefficient of 0.983×10^{-10} $m^2\cdot s^{-1}$ was previously reported for WT W_1.[276] Neither the reduced nor oxidized W1M, thus, exhibits hydrodynamic behaviour that largely deviate from WT. This provides strong evidence that the stapled mutant is stable in the

monomeric state, with an intramolecular disulphide bond (as also implied by non-reducing SDS-PAGE in ***Fig. 4.1.***), rather than as a dimer, oligomer or as a heterogeneous mixture of monomers, dimers and/or oligomers. The observed decrease in translational diffusion coefficient, consistent with a larger hydrodynamic diameter relative to WT, implies slower diffusion of the mutant unit in either oxidation states.

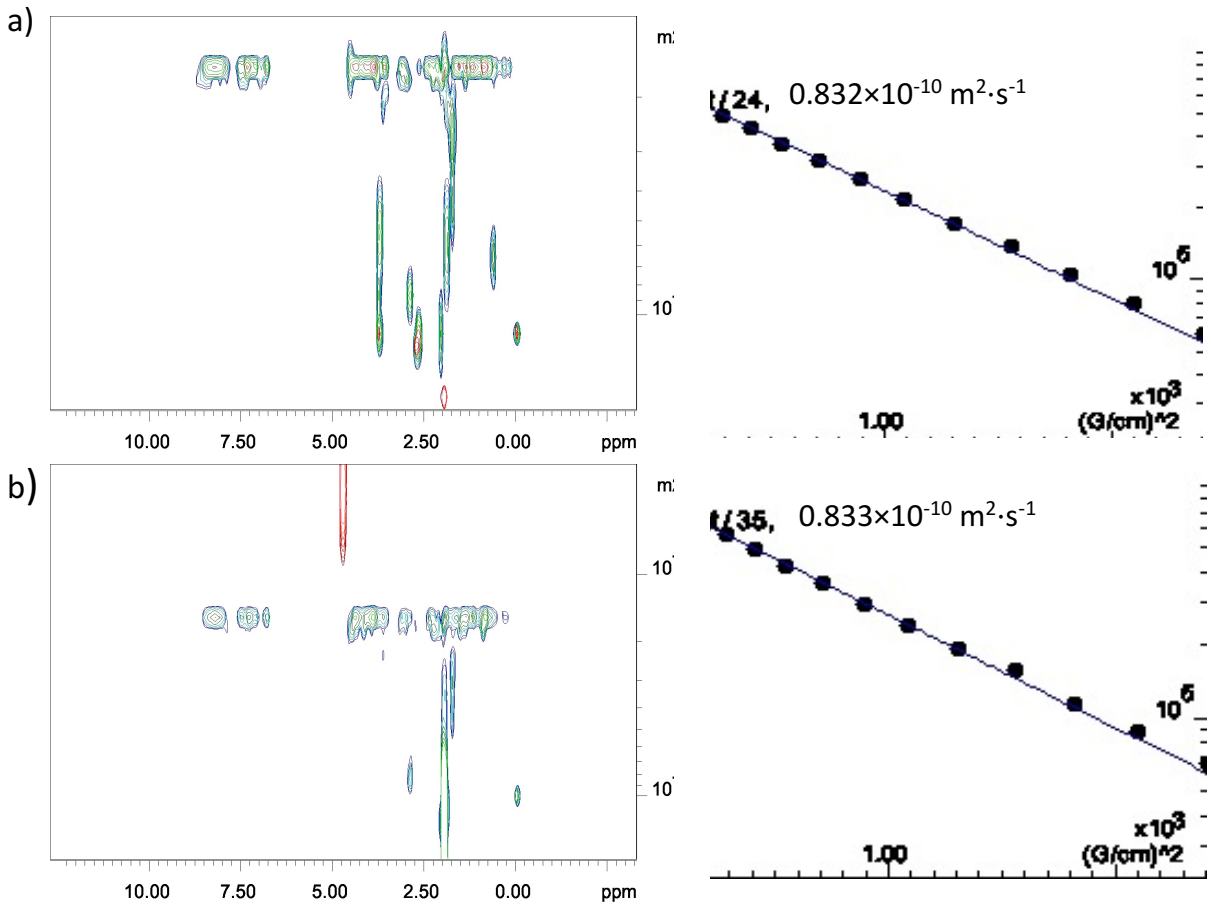


Fig. 4.2. *DOSY spectra (left) and the representative exponential fitting (right; amide resonance at 8.226 ppm) for the a) reduced and b) stapled W_1M. (Complete result from the DOSY analyses is provided in* ***Appendix C****).*

4.3.3. DTT does not affect the HSQC peak pattern of the native protein

Events like site-specific binding, complex formation, conformational switching, denaturation or oligomerization frequently involve changes in surface charge, solvation and/or bonding

interactions, resulting in variations in the secondary structure of a protein.[337-348] Protein chemical shifts, being an extremely sensitive parameter to the environment, are indicative of such structural transitions. Changes in NMR peak positions are thus often used to track these changes. Hence, backbone amide peak patterns in 2D spectra like the ^{1}H-^{15}N HSQC are reflective of modifications in the secondary and tertiary structure of a protein.

0.2 mM W_1 with DTT (W_1+DTT) was prepared to serve as a control for the reduced W_1M. The ^{1}H-^{15}N HSQC peak patterns observed for the W_1+DTT sample are indistinguishable from those previously reported by Tremblay *et. al.* in the absence of DTT (***Fig. 4.3.***)*.*[9] From the near perfect overlay of ^{1}H-^{15}N HSQC spectra of WT samples prepared with and without DTT, it was concluded that DTT does not alter the amide peak distribution and, hence, does not perturb the structuring of the WT W unit in the solution state. Minor differences observed in the side chain amide peak patterns could be attributed to different sample preparation and experiemntal setting. From here on W_1+DTT (referred to as WT W_1) was used for comparing the chemical shifts and structuring of the reduced mutant proteins, which will then be compared to their oxidized counterparts.

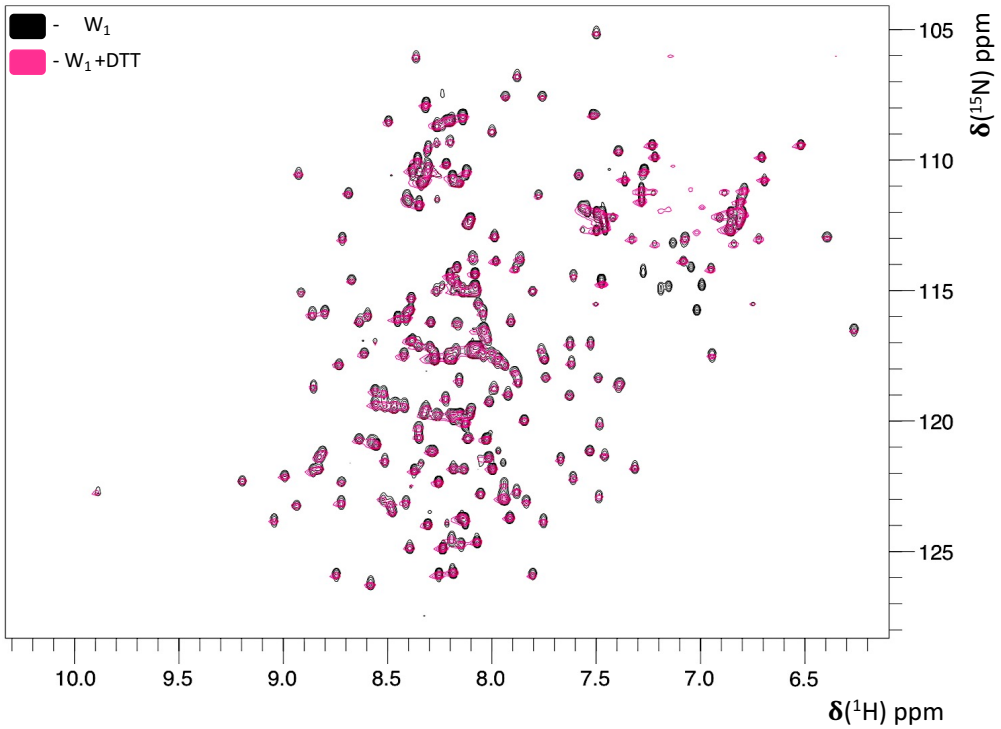


Fig. 4.3. *Overlayed ^{1}H-^{15}N HSQC spectra of WT W_1 in the absence (black, 700 MHz) and presence of DTT (magenta; 800 MHz) show the ineffectiveness of DTT on the amide peak patterns. Spectrum of WT W_1 without DTT was reproduced from Tremblay et. al.[9] for direct comparison.*

4.3.4. Mutant W unit acts as a modular unit

Highly similar backbone and side chain amide ^{1}H and ^{15}N resonance patterns were observed for the W_1M and W_2M proteins in both the reduced and oxidized states, as is apparent through the direct overlay of ^{1}H-^{15}N HSQC spectra (***Fig. 4.4.)***. Following the DOSY results that indicate a monomeric nature of W_1M in both states. An alternate case of random or intermolecular disulphide bridges would have caused protein aggregation over time or resulted in alterations in peak width/patterns/intensities/numbers**.** This is consistent with the previously reported modular nature of the WT W unit, thus providing a clear demonstration that the structure is retained by the mutant W unit in either oxidation state. Based upon this, W_1 proteins were employed for all solution-state structural studies (detailed in this chapter) and dynamics investigations (elaborated in ***Chapter 5***).

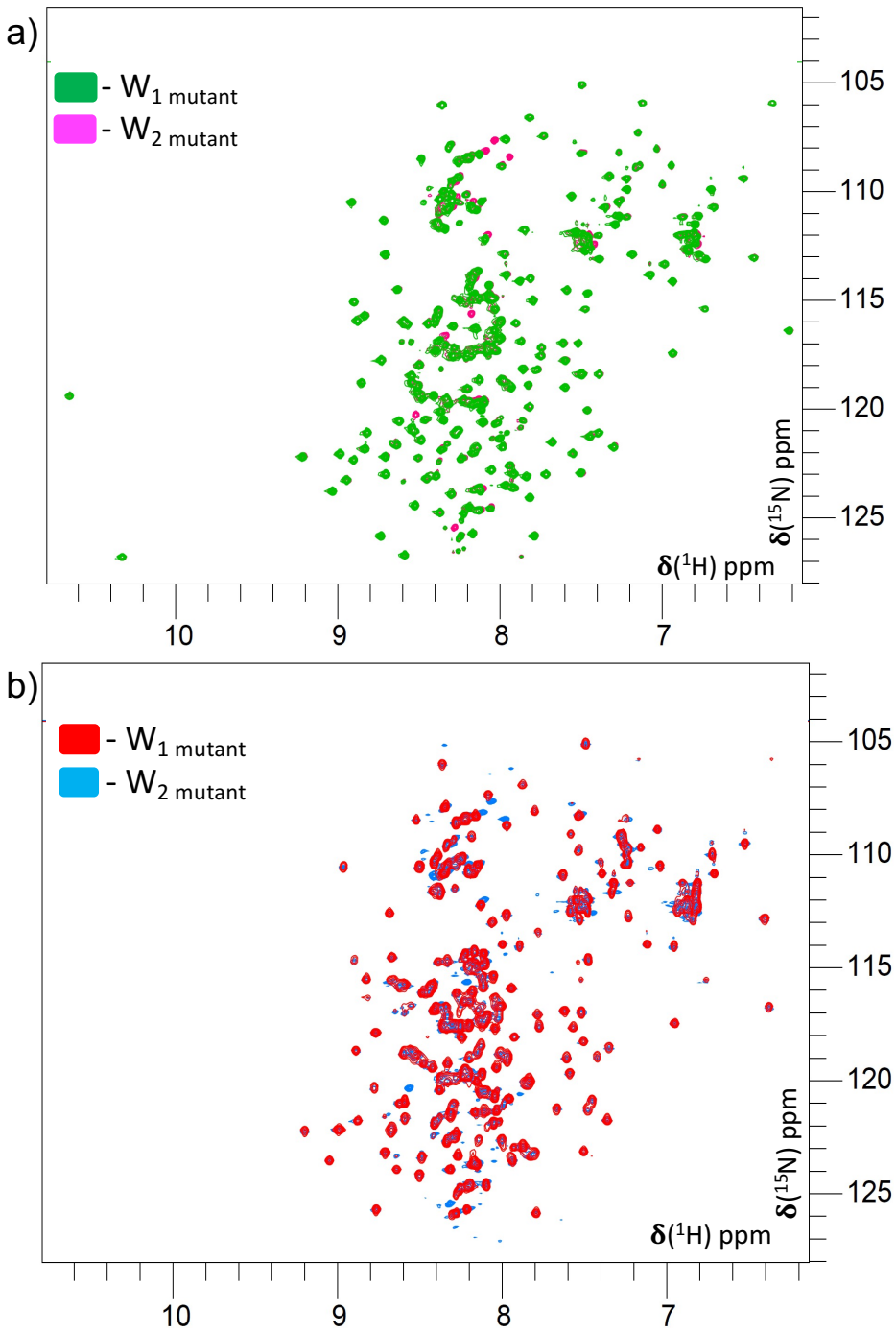


Fig. 4.4. *Overlayed ^{1}H-^{15}N HSQC spectra of W_1M and W_2M proteins demonstrate the modular behaviour of the W unit in both the a) reduced and b) oxidized conformational states.*

4.3.5. Backbone chemical shift assignment

Significant differences were observed in backbone amide peak positions for the ^{1}H-^{1}N HSQC spectral overlay of the WT W_1 with the stapled mutant while the reduced mutant had a very similar peak pattern to WT (***Fig. 4.5.***). In order to track down both the differences and similarities in structuring and chemical shifts, triple resonance 3D NMR experiments were performed to allow sequential backbone assignment (following the same strategy detailed in ***section 2.2.2.8.2.***) Chemical shifts corresponding to Cα, Cβ, C'), N and H_N resonances were independently assigned for both stapled and reduced W_1M proteins (***Appendix D.***), following which assignment quality was verified using the PANAV server[217] ***(Appendix E.)***. ^{1}H-^{15}N HSQC spectra with the assignments directly annotated are shown in ***Fig. 4.6.*** The presence of cystines in the stapled mutant and cysteines in the reduced mutant was unambiguously confirmed from the Cβ values of the Cys residues assigned for each state of W_1M (***Fig. 4.7.)***, in the same manner as discussed for soricidin in ***section 2.3.1.2.***

Table 4.2: % CHEMICAL SHIFT ASSIGNMENT FOR W_1M

Atom Type	% assignment	
	Reduced	Stapled
backbone	97.8	97.3
N	95.0	94.0
C'	99.5	99.0
Cα	99.0	99.0
Cβ	98.2	97.0
H	99.5	98.4

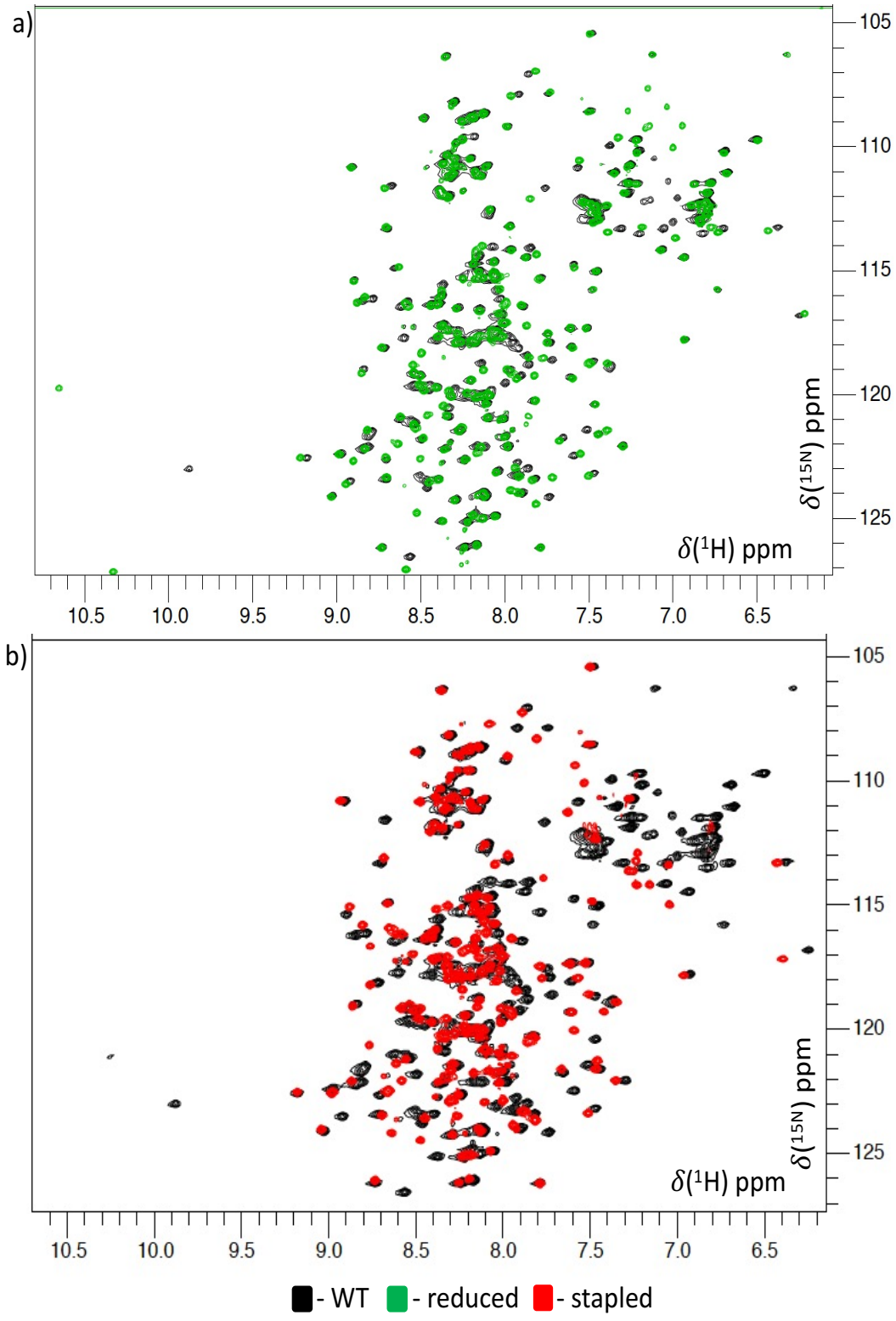


Fig. 4.5. *^{1}H-^{15}N HSQC spectral overlay of a) reduced and b) stapled W_1M with WT W_1.*

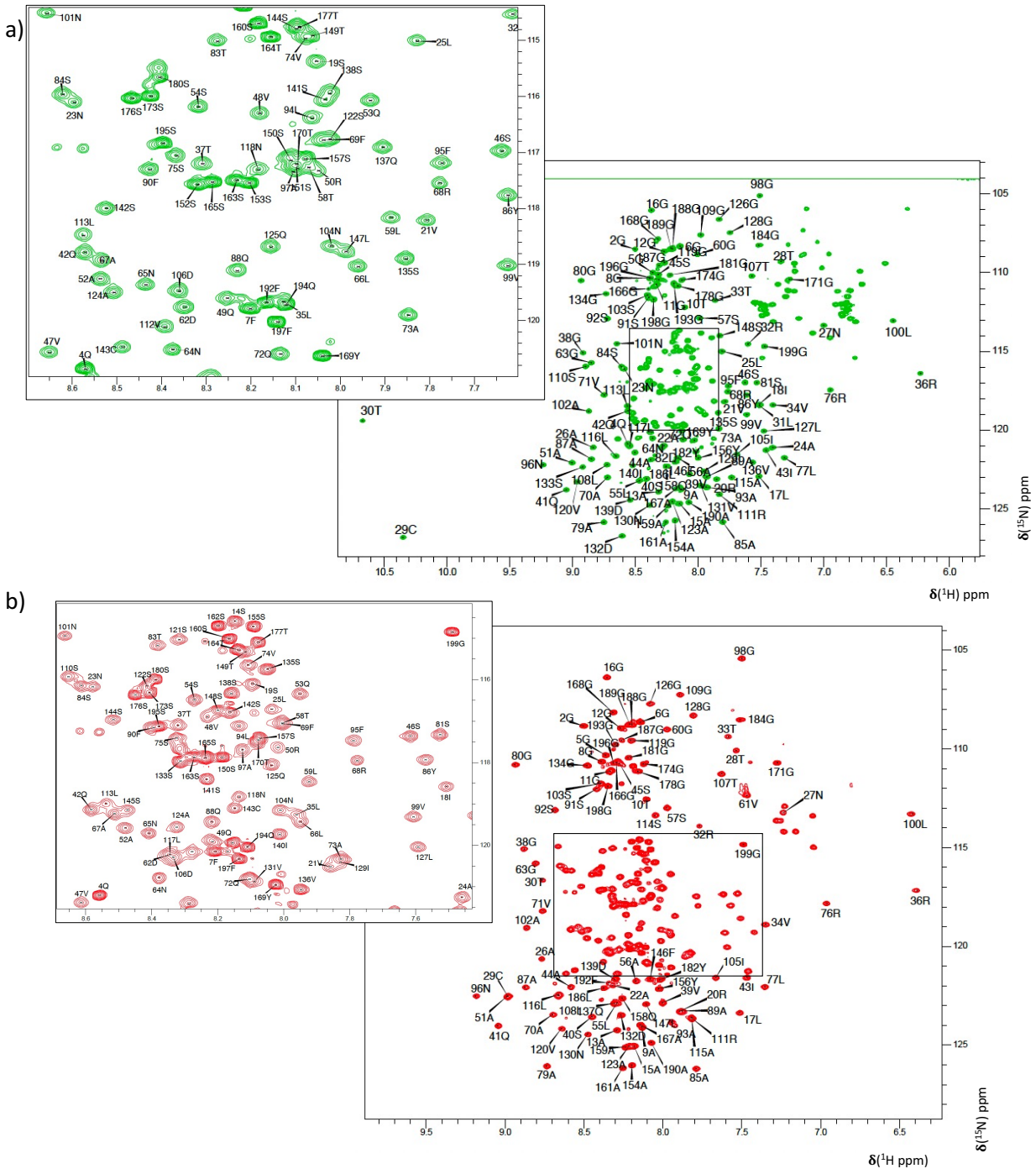


Fig. 4.6. *Assigned HSQC spectra of a) reduced and b) stapled W_1M mutants.*

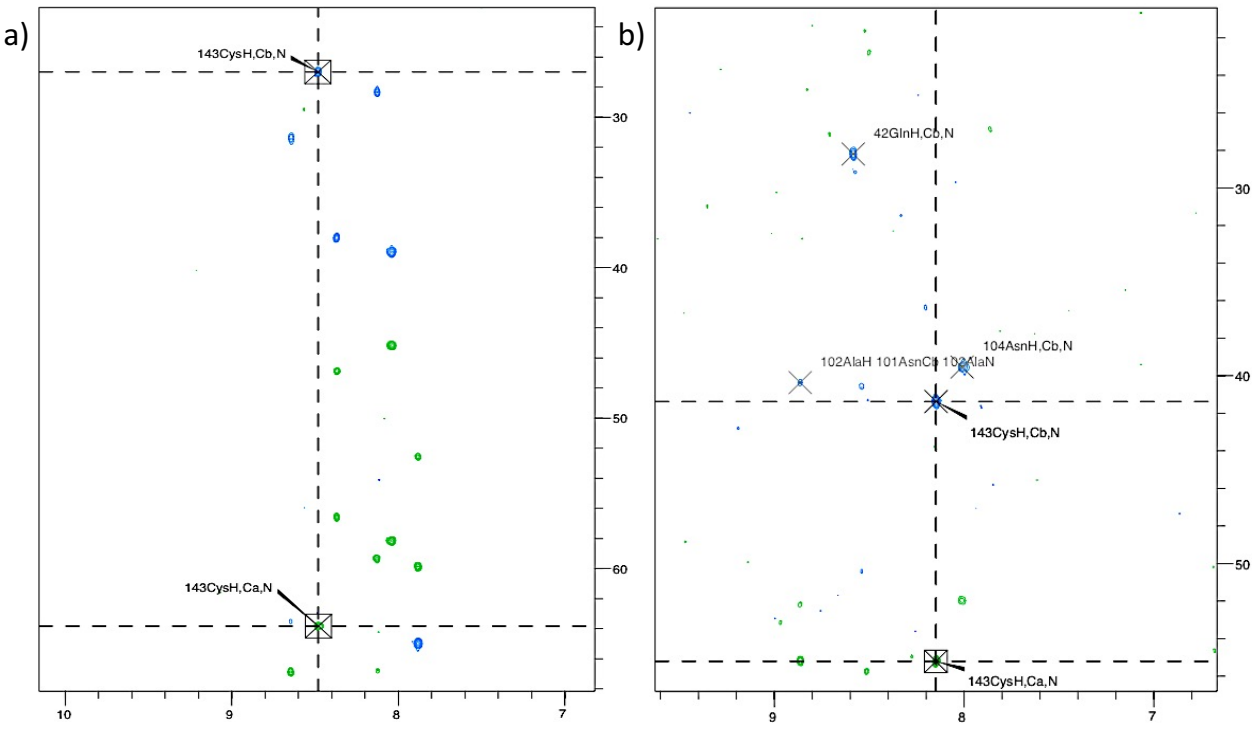


Fig. 4.7. *Spectral slices from CBCANH spectra collected on the a) reduced and b) stapled W_1M proteins show a pronounced difference in the Cβ value for 143Cys, indicative of two different oxidation states (^{15}N: 119.109 ppm).*

Comparable peak patterns were observed for the Cys residues across all the backbone NMR experiments collected on stapled W_1M and W_2M (as in ***Fig. 4.4.b***), as previously reported for WT[5] and reduced mutants being reported herein. This, in turn, validates complete oxidation and the success in directing disulphide bridge formation within each W unit in both W_1M and W_2M proteins in the stapled form, thus ruling out the possibility of crosslinking between W units.

4.3.6. Disulphide bond alters W unit domain architecture

Since chemical shifts are highly dependent on the chemical environment,[349] following backbone chemical shift assignment, the chemical shifts of the reduced and stapled mutants were compared. The oxidized mutant showed significant variance in comparison to the reduced W_1M, with major

deviations observed around the mutated residues suggesting alterations in the domain architecture in response to disulphide locking.[349] CSD plots, (***Fig. 4.8.***) generated as per ***Eqs. 4.2.*** and ***4.3.,*** quantitatively represent these differences.

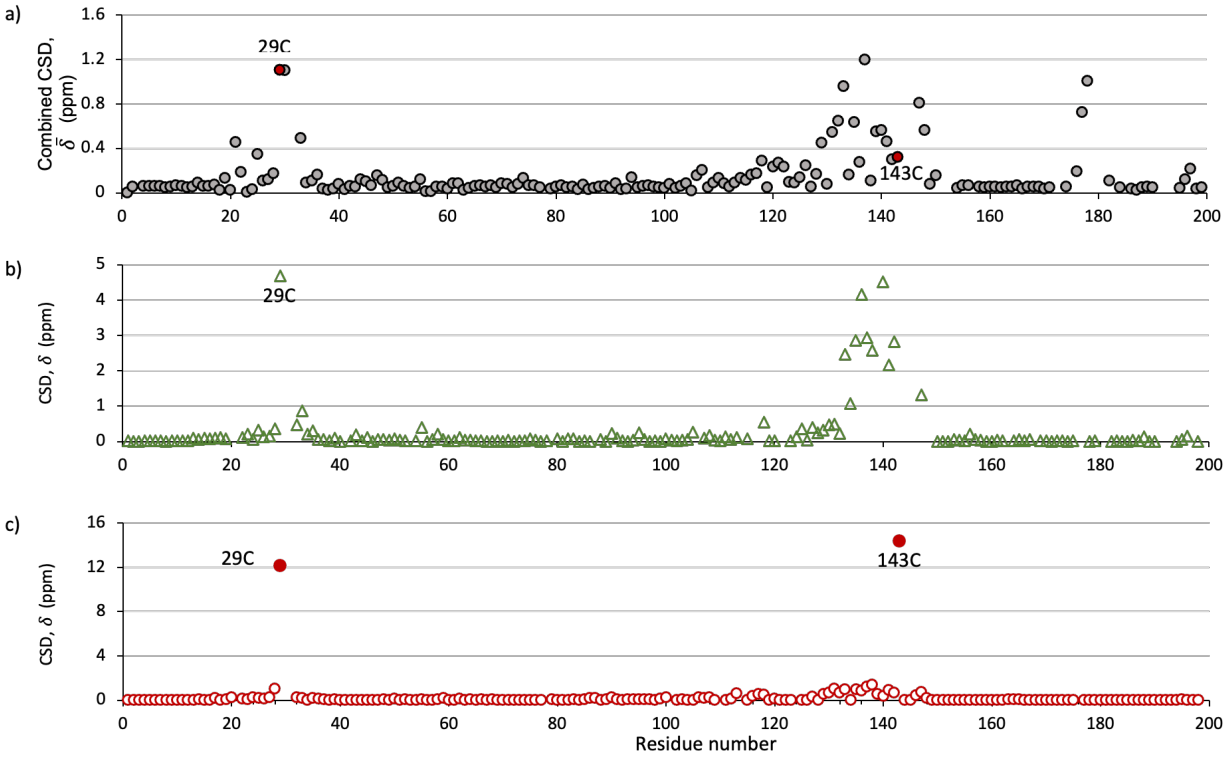


Fig. 4.8. *Variations in backbone resonance patterns of oxidized vs. reduced W_1M are quantitatively depicted as chemical shift displacements (CSD) plotted against residue number. Shown here are a) combined CSD ($\bar{\delta}$) for the backbone amides and CSD (δ) for the b) Cα and c) Cβ carbons.*

4.3.7. Secondary structure comparison of native and mutant proteins

Alongside the secondary chemical shift plots, DANGLE[221]- predicted secondary structure (***Fig. 4.9.)*** indicate that the reduced mutant is structurally comparable to WT protein while the oxidized mutant showed significant differences, with loss of/altered helicity suspected for the H5 region which is covalently linked to the protein core region. Positions of a small number of resonances were observed to be perturbed in each case, which was identified as residues corresponding to the

unstructured regions comprising the linkers between helices and the C-tail based on the sequential assignment.

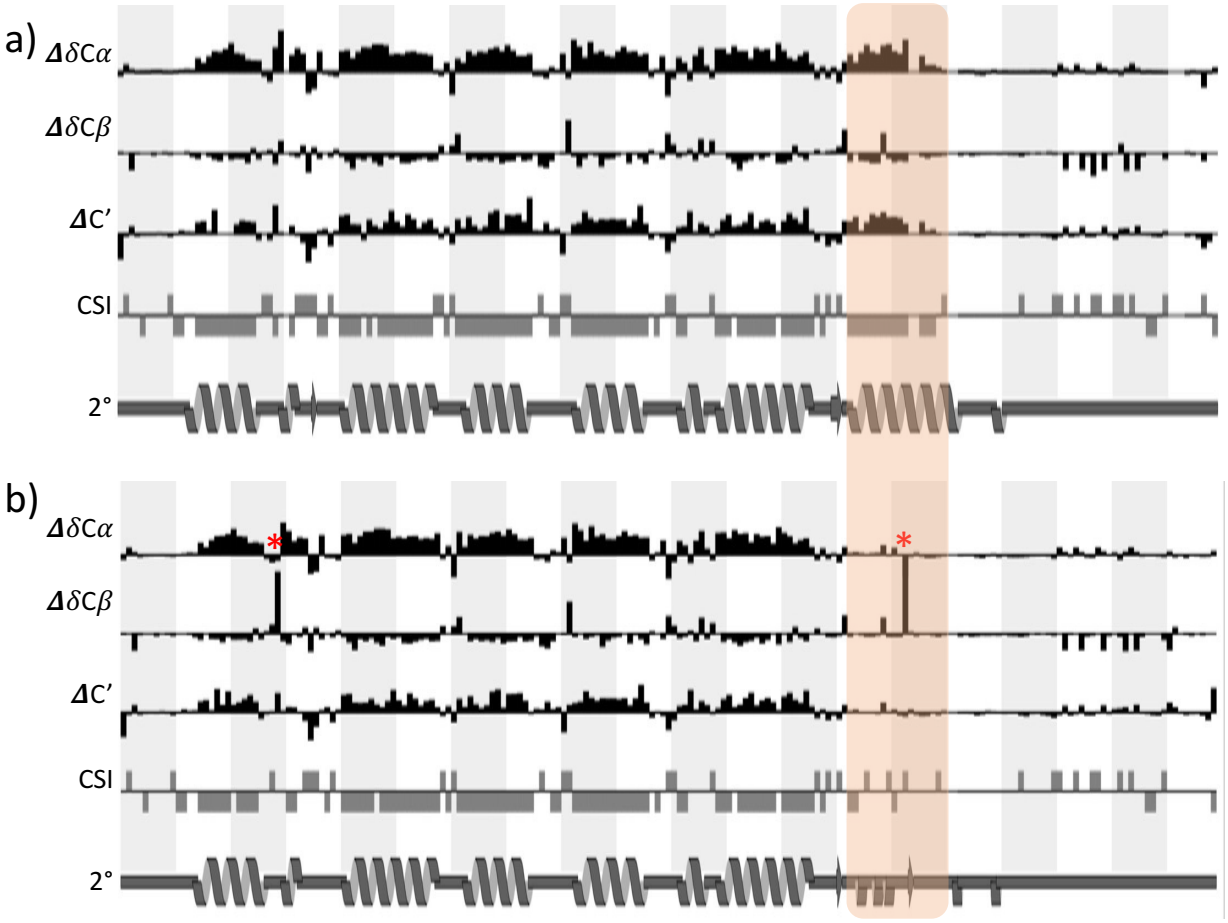


Fig. 4.9. *DANGLE predicted secondary structure for the W units of a) reduced, and b) stapled W_1M. demonstrate significant alterations in the H5 region (highlighted in orange box). Asterisks highlight the large deviations in Cβ values, demonstrating the presence of cystine while the CSI index indicates structural alterations at H5.*

4.4. SUMMARY

Here, an atomic-level investigation of the protein backbone using solution-state triple resonance NMR spectroscopy was carried out on the soluble mutant *AcSp1* proteins to test for the modularity of the mutant W unit in both the reduced and disulphide-stapled forms. Through the demonstration

that the reduced mutant and the WT protein with DTT exhibit NMR chemical shifts consistent with WT W_1 under the conditions employed in previous structural studies, the potential for DTT to modify the secondary structural features in the solution-state can be ruled out. The major discovery from the studies carried out in this chapter is the revelation that the disulphide bond present in the stapled form leads to a significant secondary structural alteration, even though the hydrodynamic size remains comparable to that of the reduced form. Some structural change was suggested by far-UV CD spectroscopy in ***Chapter 3***, but these data allow me to unambiguously demonstrate that the perturbation is in the region localized around the introduced disulphide bond. Based on these structural insights about the stapled mutant, I decided to carry out a detailed comparison of the protein backbone dynamics and to probe for the possibility of observable conformational exchange in each form of the protein. Given the premise of the engineered disulphide as a means of limiting protein dynamics and, particularly, the unfolding and decompaction of H5 from the core, I felt that a comparison of dynamic behaviour would be essential to understand the fundamental differences between the two forms of W_1M protein, in order to relate back to the differences in functionality demonstrated in ***Chapter 3***. These dynamics studies carried out through ^{15}N nuclear spin relaxation measurements, employing reduced spectral density mapping, and through ^{15}N relaxation dispersion experiments are discussed and detailed in the next chapter.

Chapter 5

THE EFFECT OF THE DISULPHIDE LOCK ON THE DYNAMICS AND CONFORMATIONAL SAMPLING OF THE AcSp1 W UNIT

5.1. INTRODUCTION

Based on the differences observed between the reduced and stapled mutants relative to WT in their protein backbone discussed in ***Chapter 4***, I decided to expand my NMR spectroscopy-based studies to probe the effects of disulphide stapling upon the dynamics of motion within the W unit. Specifically, I carried out a series of studies evaluating ^{15}N spin relaxation and, as discussed in this chapter, use the resulting analyses to show differences in the backbone dynamics for the stapled relative to reduced mutant W unit.

5.1.1. Dynamics in proteins

A pictorial illustration of the tertiary structure of any protein ensemble is often deceiving since this does not portray all aspects of the flexible and dynamic nature of molecules when in solution. Namely, these molecules may undergo vibrational, rotational, and translational motions or conformational rearrangements over a variety of timescales (ns-s) and length scales (Å-nm) as shown in ***Fig. 5.1***. These motions may be modulated in response to a variety of environmental factors (e.g., change in temperature, pH, ionic strength, chemical modification, binding etc.), making them a sensitive probe for functionally relevant processes. Through a variety of NMR spectroscopy experiments, intramolecular and global molecular motions over this entire timescale may be probed (***Fig. 5.1.***). In the experiments carried out here, ^{15}N nuclear spin relaxation

phenomena (the R_1, R_2, NOE, and CPMG relaxation dispersion experiments noted in ***Fig. 5.1.***, specifically) are quantified and evaluated to probe W unit dynamics over the ps-ms timescales.

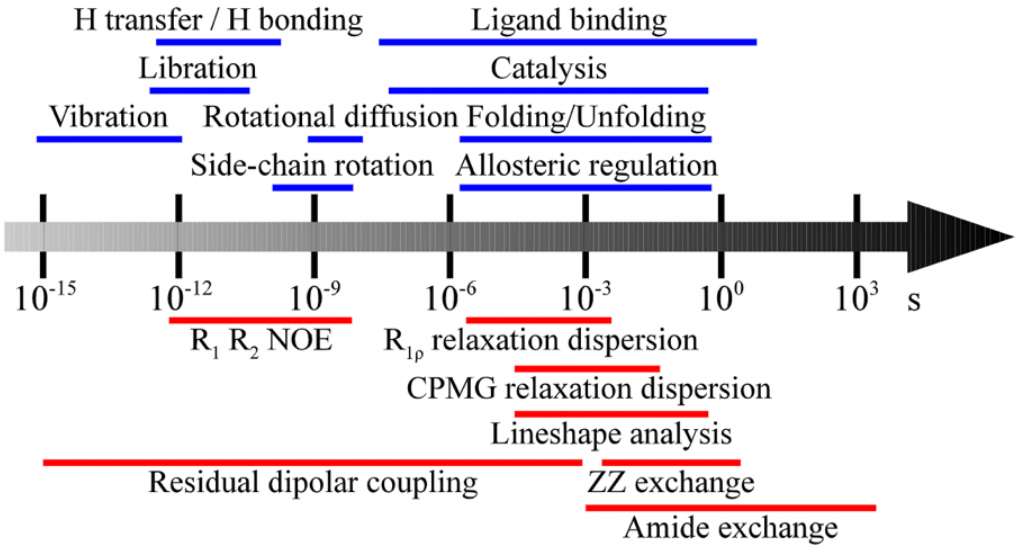


Fig. 5.1*. Diagram representing timescales of protein dynamics (blue) and timescales probed by NMR experiments (red) (adapted from Palmer et. al.[10])*

5.1.2. Nuclear spin relaxation

At equilibrium, the bulk averaged nuclear spin magnetization only has a longitudinal component (i.e., along the z-axis of the NMR instrument, as defined by the direction of the static magnetic field, $\mathbf{B_0}$) and no net transverse magnetization in the x-y plane (plane perpendicular to the direction of $\mathbf{B_0}$). The application of radiation at the appropriate frequency (i.e., the Larmor frequency of the nucleus in question at magnetic field strength, $\mathbf{B_0}$), applied as an RF pulse in an NMR experiment, allows this equilibrium state to be disturbed with the introduction of net transverse magnetization. It should also be noted that in the typical NMR spectrometer configuration, it is only transverse magnetization that is detectable, not longitudinal relaxation. With the spin population at each energy level being specifically defined by the Boltzmann distribution at thermal equilibrium, the process of re-establishing the Boltzmann equilibrium (i.e., net magnetization along z and no x-y

magnetization) occurs through time-dependent magnetization transfer (i.e., energy exchange) processes known as nuclear spin relaxation.[209]

5.1.3. NMR as a tool to probe protein dynamics

Nuclear spin relaxation processes are relatively slow in comparison to most other molecular energy relaxation processes. The rate of nuclear spin relaxation depends upon the physical environment, the nature of the individual nuclei relaxing, and the motion of and within the molecule. With chemical shift being an observable measure of changes in the electronic environment(s) of different nuclei within a molecule, high-field solution NMR spectroscopy thus can serve as a tool for assessing protein dynamics.[10, 350-353] Given the sensitivity of the NMR observable to an extensive range of time scales and magnitudes of motion, both the type of motion of and within an NMR-active molecule like a protein occurring at different timescales (***Fig. 5.1.)*** and the corresponding relaxation rate constants for any NMR active nuclei can be estimated or quantified at the atomic-level. [209, 354, 355] Recent advancement of NMR techniques also enables the extraction of kinetic and thermodynamic information to probe for chemical exchange, thus revealing transient and sparsely- or multiply-populated states, if these exist.[356]

Although relaxation NMR experiments can be performed on any NMR active nucleus, the study of N-H bond motions enables investigation of the peptide bonds for all residues except Pro and, thus, provides the ability to carry out the backbone-level characterization of localized motion that covers the entire length of the protein. Since ^{1}H-^{15}N spin pairs in peptide bonds have been extensively characterized, and serve as a good model for describing motions within proteins,[10, 355, 357] only ^{15}N-relaxation in the context of J-coupled ^{1}H-^{15}N spin pairs within peptide bonds is presented herein. The ^{15}N-relaxation experiments used for studying protein dynamics rely

primarily upon the measurement of three phenomena: longitudinal relaxation, transverse relaxation, and the steady-state heteronuclear NOE. These are introduced in turn.

5.1.3.1. **Longitudinal relaxation** (synonym: spin-lattice/thermal relaxation)

Longitudinal relaxation refers to the process of restoring the magnetization along the z-direction when the system returns to the net equilibrium defined by the Boltzmann distribution following some perturbation (***Fig. 5.2.; red).*** Mechanistically, molecular motions of nearby nuclei cause fluctuations in the transverse field which, in turn, stimulates longitudinal magnetization transfer. Hence, the interaction of spins with their environment (often referred to as the lattice), including changes in the local magnetic field due to molecular motions (like collision, rotation etc.), localized chemical shift anisotropy, J-coupling or dipolar coupling between different nuclei, all facilitate energy loss. The rate of this relaxation process is given by,

$$I_t = I_0 \cdot e^{-R_1 t} \qquad \textbf{(5.1.)}$$

where I_t is the peak intensity observed at a delay time t, I_0 is the original peak intensity and R_1 is the longitudinal relaxation decay constant. The corresponding time constant for this relaxation process is often also employed and discussed, as defined by

$$T_1 = 1/R_1 \qquad \textbf{(5.2.)}$$

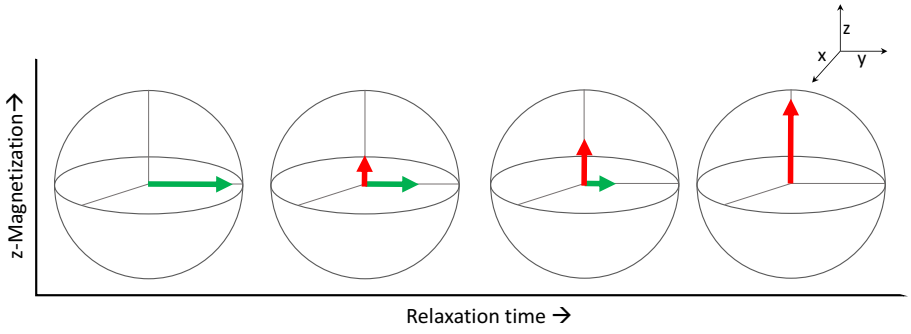


Fig. 5.2. *Spin diagram depicting longitudinal (red) and transverse(green) relaxation.*

5.1.3.2. **Transverse relaxation** (synonym: spin-spin relaxation)

Application of an RF pulse during an NMR experiment creates coherence among the spins, which independently evolve through stochastic magnetization transfer between spins as they precess around B_0. Local magnetic field fluctuations initiate processes like dipole-dipole interactions, chemical exchange etc. which affect the Larmor frequency of individual spins, thus resulting in a gradual loss of coherence. The decay of magnetization in the x-y plane to restore the equilibrium state where there is no net x-y magnetization is known as transverse relaxation (***Fig. 5.2.; green***) and is characterized by the rate constant, R_2 which is given by[358]

$$I_t = I_0 \cdot e^{-R_2 t} \quad \textbf{(5.3.)}$$

with R_2 - transverse relaxation decay rate constant. As with longitudinal relaxation, the time constant is obtained from the inverse of the decay rate constant.

5.1.3.3. **Steady-state heteronuclear NOE**

The steady-state [^{1}H]-^{15}N heteronuclear NOE, previously introduced in section ***2.3.4.8.b.***, provides a quantitative measure of ^{1}H-^{15}N cross-relaxation processes operating as a function of position in the protein backbone. The identity of residues/domains exhibiting disparities are identified in terms of the NOE enhancement, σ, given by ***Eq. 2.5***.[353]

5.1.4. **Reduced spectral density mapping**

The molecular motion may be depicted by a time-dependent rotational correlation function (Cτ), which accounts for the conformational arrangement of spins within a system. Fourier

transformation of this correlation function generates a spectral density function $J(\omega)$, which is a measure of the relative magnitude of motion as a function of frequency.[359, 360]

The ^{15}N-relaxation parameters ^{15}N-R_1, ^{15}N-R_2 and [^{1}H]-^{15}N NOE may be related to specific linear combinations of the spectral density function at five frequencies, ω_0, ω_H, ω_N, $\omega_H + \omega_N$ and $\omega_H - \omega_N$, as[18, 359]

$$R_1 = \frac{d^2}{4}[3 \cdot J(\omega_N) + J(\omega_H - \omega_N) + 6 \cdot J\,(\omega_H + \omega_N)] + \; c^2 \cdot J(\omega_N) \tag{5.4.}$$

$$R_2 = \frac{d^2}{8}[4 \cdot J(0) + 3 \cdot J(\omega_N) + J(\omega_H - \omega_N) + 6 \cdot J(\omega_H) + \; 6 \cdot J\,(\omega_H + \omega_N)] + \frac{c^2}{6}[4 \cdot J(0)] + 3 \cdot J(\omega_N)] + R_{ex} \tag{5.5.}$$

$$NOE \;=\; 1 + \frac{d^2}{4R_1}\left(\frac{\gamma_H}{\gamma_N}\right)[6 \cdot J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \tag{5.6.}$$

where $J(\omega)$ is the spectral density at the frequency indicated by the subscript (0, ^{1}H, ^{15}N, or the indicated combination), γ_H and γ_N are the gyromagnetic ratios of $_1$H and ^{15}N, respectively,R_{ex} is a contribution accounting for slow exchange processes occurring on the µs-ms timescale, and d_c and c are dipolar and chemical shift anisotropy coefficients, respectively, defined specifically as:

$$d_c = \left(\frac{\mu_0 h \gamma_H \gamma_N}{8\pi^2}\right)\left(\frac{1}{r_{NH}^3}\right) \tag{5.7.}$$

$$c = \frac{\omega_N}{\sqrt{3}}(\sigma_{\parallel} - \sigma_{\perp}) \tag{5.8.}$$

where μ_0 is the permeability of free space, h is Planck's constant, r_{NH} is the vibrationally-averaged effective N-H bond length, and, $\sigma_{\parallel}$ and $\sigma_{\perp}$ are the parallel and perpendicular components of the anisotropic chemical shift tensor of ^{15}N in the context of an N-H bond.[355]

Complete spectral density mapping enables the estimation of the spectral densities at the five frequencies mentioned above. However, this requires the determination of three additional relaxation rates: 1) the ^{1}H R_1, 2) the $R_{NH}(2H_z^N N_z)$ longitudinal two-spin relaxation rate, and 3) the $R_{NH}(2H_z^N N_{x,y})$ antiphase ^{15}N single quantum coherence relaxation rate. The dependence of these relaxation parameters on ^{1}H-^{1}H dipolar interactions and the local magnetic field can lead to deviation from mono-exponential decay. As a result of this, precise extraction of these relaxation rates becomes difficult.[361] The reduced spectral density mapping approach[361] provides a simplified procedure to analyze spin relaxation data using only the three ^{15}N relaxation parameters, each of which can be reliably determined, which are noted above. This replaces the correlation function with a reduced correlation function such that the reduced spectral density, $j(\omega)$, is independent of the local magnetic field and is given by the Lorentzian function

$$j(\omega) = \frac{2\tau_c}{1 + (\omega\tau_c)^2}$$

(5.9.)

where τ_c is the rotational correlation time that defines the rate of molecular tumbling motion. Based on the Stokes-Einstein -Debye relationship for a sphere, correlation time, τ_c is given by [362]

$$\tau_c = \frac{4\pi\eta r^3}{3k_B T} = \frac{V\eta}{k_B T}$$

(5.10.)

where r and V are respectively the radius and volume of the sphere, η the viscosity of the solution, k_B is the Boltzmann constant and T is the absolute temperature. ***Eq. 5.10.*** shows that correlation time is directly dependent on the molecular size and viscosity while inversely related to the temperature.

Eq. 5.9. shows that the reduced spectral density function steadily decreases with an increase in frequency and that the rate of decrease is determined by the correlation time. It should also be noted that reduced spectral density has a maximum value at zero frequency i.e., $j(0) = 2\tau_c$ which proportionally increases with an increase in correlation time.[363]

In this approach, owing to the order of magnitude difference in the gyromagnetic ratios of ^{1}H and ^{15}N (($\gamma_H = 26.7 \times 10^7$ rad·s^{-1}·T^{-1}, $\gamma_N = -27.2 \times 10^6$ rad·s^{-1}·T^{-1}), it is assumed that spectral density values for $J(\omega_H)$, $J(\omega_H - \omega_N)$ and $J(\omega_H + \omega_N)$ are centred around $J(\omega_H)$ and approximated to $\langle j(\omega_H)\rangle$,[364] the spectral density at the *effective* proton frequency defined as[365]

$$\langle j(\omega_H)\rangle = J(0.87\omega_H) \qquad \textbf{(5.10.)}$$

and the reduced spectral density mapping modifies ***Eqs. 5.4.-5.6*** at three frequencies to:

$$j(0) = \frac{1}{3d^2 + 4c^2}\left(6R_2 - R_1\left(3 + \frac{18}{5}\left(\frac{\gamma_N}{\gamma_H}\right)(NOE - 1)\right)\right) \qquad \textbf{(5.11.)}$$

$$j(\omega_N) = \frac{4}{3d^2 + 4c^2}\left(R_1\left(1 - \frac{7}{5}\left(\frac{\gamma_N}{\gamma_H}\right)(NOE - 1)\right)\right) \qquad \textbf{(5.12.)}$$

$$\langle j(\omega_H)\rangle = \frac{4}{5d^2}\left(R_1\left(\left(\frac{\gamma_N}{\gamma_H}\right)(NOE - 1)\right)\right) \qquad \textbf{(5.13.)}$$

The dependence of the three reduced spectral density functions on the measured relaxation parameters - R_1, R_2, and NOE - is clear from ***Eqs. 5.11.- 5.13.***, with the effect of field strength on R_1 being more pronounced,[3] as in ***Fig. 5.3.*** The amplitudes of these density functions indicate the total degree of freedom of the ^{1}H-^{15}N bond and indicate the type and extent of motion, along with the corresponding timescale.[209, 359, 361, 366]

The rate of longitudinal relaxation is determined from the spectral density at the Larmor frequency, $j(\omega_0)$ whose value is maximum when $\tau_c = 1/\omega_0$, implying that all frequencies other than ω_0 will

experience a slower relaxation. Two motional regimes are defined by the limits set for ***Eq. 5.9.*** Smaller τ_c values are observed for fast tumbling motions, as in the case of small molecules, which would mean that $\omega_0\tau_c << 1$, such that $1 + (\omega\tau_c)^2 \approx 1$ and hence $j(\omega_0) = 2\tau_c$, implying that at higher frequencies, $j(0)$ is independent of Larmor frequency. Slower tumbling motions with large values of τ_c would mean that $\omega_0\tau_c >> 1$, such that $1 + (\omega\tau_c)^2 \approx (\omega\tau_c)^2$ and, hence, $j(\omega_0) = 2/(\omega_0{}^2\tau_c)$, implying that $j(\omega_0) << j(0)$ for slower tumbling motions (***Fig. 5.3.***).

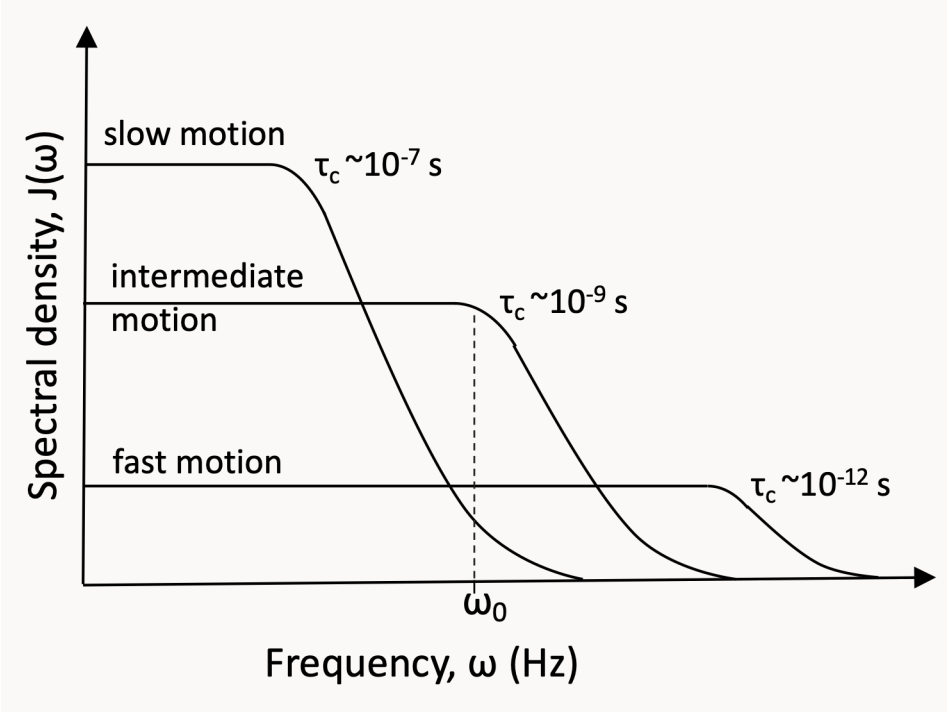


Fig. 5.3. *Diagram demonstrating the relationship between spectral density function J(ω) and frequency (ω) and their dependence on the effective correlation time(τ_C).*

5.1.5. Conformational exchange

The intrinsic flexibility and dynamic nature of a protein in solution results in motions of the protein chain which would result in conformations insignificantly distinct from the average structure. Such changes could be a response to environmental factors like pH, ionic concentration, temperature,

pressure etc. However, these motion(s) become relevant when such substates sampling alter protein function through intrinsic internal motions (different conformational state) or interaction with other molecules (oligomerization, complex formation etc.) is known as conformational exchange (***Fig. 5.4***). In a two-state exchange process, each residue will give rise to two distinct resonance patterns (in terms of peak number/position/intensity etc.) in NMR spectra – one corresponding to each conformation. Based on the differences in CS value, peak intensity etc. parameters detailing the process of conformational exchange can be calculated [60, 355, 367-369]

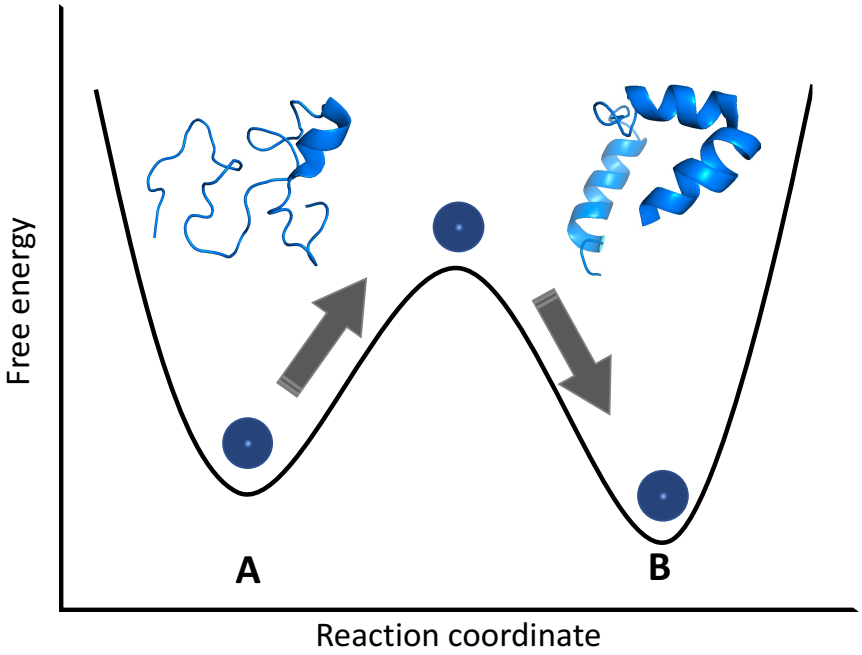


Fig. 5.4. *Schematic representation for two-state conformational exchange in proteins.*

The constant-time Carr–Purcell–Meiboom–Gill (CT-CPMG) relaxation dispersion experiment was used in my work to test for conformational exchange and characterize the conformational equilibrium and kinetics of exchange between conformational states on the ms-µs timescale.[370, 371] The ^{15}N R_2 is measured as a function of the effective CPMG field strength and the rate constant, $R_{2,\,eff}$ is defined as[370]

$$R_{2,eff} = \frac{-1}{d} \ln\left(\frac{I}{I_0}\right)$$

(5.14.)

where d is the constant relaxation delay time.

5.1.6. Objectives

Given the demonstration that the engineered disulphide bridge alters the tertiary structure in the W unit, this has the potential to modulate local dynamics, particularly in the H5 region. Comparing the dynamics behaviour of the stapled to the reduced/WT forms will thus be highly informative, given that the postulated decompaction of this region from the globular core would be rendered impossible in the stapled form. This chapter thus details ^{15}N spin relaxation-based studies performed with the WT W_1 and W_1M (in both reduced and stapled forms) proteins with the following objectives:

i. Measure ^{15}N-R_1 and ^{15}N-R_2 relaxation rates and the [^{1}H]-^{15}N NOE enhancement factors as a function of residue and as a function of protein form.

ii. Perform reduced spectral density mapping to better characterize molecular motions and identify any disparities.

iii. Evaluate the effect of magnetic field strength upon these phenomena to test for data consistency and to account for/rule out discrepancies arising from changes in magnetic field strength; and,

iv. Test for evidence of conformational exchange occurring in any of the forms of the protein.

5.2. MATERIALS AND METHODS

5.2.1. Sample preparation

Two sets of samples were separately used to perform ^{15}N-relaxation experiments. For the first set of experiments, the [U-^{13}C, ^{15}N] WT W_1 (125 μM) and the reduced (500 μM) and stapled (400 μM) W_1M samples used for backbone NMR studies (discussed in *section **4.2.2.***) were used. The second set of experiments was performed with the batch of U-^{15}N reduced and stapled W_1M proteins used for DOSY measurements (*section **4.2.2.)***.

5.2.2. ^{15}N-NMR experiments to determine relaxation parameters

5.2.2.1. Measurement of R_1

Relaxation recovery (***Fig. 5.4.; red)*** implemented in a 2D ^{1}H-^{15}N HSQC pulse sequence was employed to estimate ^{15}N-R_1 (or ^{15}N-R_1^z) decay resulting from spin-lattice relaxation.[215] Following the initial INEPT step that generates ^{15}N enhancement, the ^{15}N magnetization allowed to relax for a variable delay time (d) under conditions allowing longitudinal relaxation. During T, the ^{15}N will undergo longitudinal relaxation towards its equilibrium magnetization (i.e., a much smaller magnitude than the initial INEPT-enhanced state). Ultimately, following a variable ^{15}N chemical shift evolution time, the residual ^{15}N magnetization – that was modulated by R_1 relaxation during T – will be transferred back to the associated covalently attached ^{1}H for observation. ^{1}H-^{15}N HSQC spectra acquired at various delay times thus can be analyzed with respect to the peak intensity observed as a function of T, allowing quantification of the rate R_1 according to ***Eq. 5.1.***

5.2.2.2. Measurement of R_2

A 2D ^{15}N-HSQC pulse sequence with a series of spin-echo pulses[215] (***Fig. 5.4., green)*** was employed to perform collect data for ^{15}N-R_2 (or ^{15}N-R_1^{xy}) estimation. In this experiment, the INEPT-enhanced ^{15}N magnetization is stored in the transverse plane and allowed to relax under transverse relaxation mechanisms. Spin evolution is allowed to proceed for a variable delay time, during which a series of 180° pulses is applied to refocus the magnetization and remove the effects of inhomogeneous broadening. Ultimately, loss of magnetization during the variable delay is thus strictly due to transverse relaxation with the resulting signal intensity measured as a function of delay time being attenuated on the basis of transverse relaxation.[372]

Long R_2 relaxation delay times often require a CPMG pulse train with many ^{15}N 180° pulses. This requires a reduction of the power used for the ^{15}N pulses of the CPMG pulse train to avoid damaging instrumentation. The corresponding reduced span of frequencies being optimally refocused lead causes off-resonance effects for ^{15}N peaks away from the carrier frequency.[373, 374] To circumvent such off-resonance effects of ^{15}N 180° pulses encountered during the direct measurement of R_2, spin-lattice relaxation in the rotating frame ($R_{1\rho}$), was alternatively used in my experiments at the higher field strength of 18.8 T.

The variation of the delay time affects the signal amplitude based upon transverse relaxation, which is observed as peak intensity, and hence allows the quantification of ^{15}N- R_2 decay. Peak intensity (height) values from ^{1}H-^{15}N HSQC spectra acquired at various delay times were exported for quantification of R_2 values (or $R_{1\rho}$ at higher field strength) in accordance with ***Eq. 5.3.***

5.2.2.3. Measurement of heteronuclear-NOE

Steady-state [^{1}H]-^{15}N heteronuclear-NOE experiments were performed as detailed in section ***2.3.4.8*.c).** Namely, two spectra were recorded in an interleaved manner and the NOE enhancement factor was calculated from peak intensities observed in spectra with and without proton saturation and reported with error as in ***Eqs. 2.2***. and ***2.5.***, respectively.

5.2.2.4. Detection of conformational exchange/ Measurement of R_2^{eff}

CPMG relaxation dispersion experiments were acquired similarly to T_2-modulated ^{1}H-^{15}N experiments, with the exception that the CPMG spin-echo pulse train period was maintained for a constant time (CT) with variation in the frequency of the CPMG 180° pulses.[375] Peak intensity is then evaluated as a function of the CPMG pulse frequency. Following the CT-CPMG relaxation dispersion experiment, data fitting was performed on the GUI version of *Relax5.0.0.* R_2^{eff} values were obtained by fitting to a two-parameter exponential function as in ***Eq. 2.12.*** and plotted as a function of residue number. The error was estimated according to the equation,

$$\sigma_{R_2} = \frac{1}{T_{relax}} \sqrt{\left(\frac{\sigma_I}{I}\right)^2 + \left(\frac{\sigma_{I_0}}{I_0}\right)^2}$$

(5.15.)

5.2.3. Data collection and spectral processing

Experiments were carried out at 30.0 °C using three different NMR spectrometers - 1) a Bruker Avance III operating at 18.8 T equipped with a 5 mm indirect detection TCI cryoprobe (*Bruker, Canada; Kay laboratory, University of Toronto, ON)*, 2) a Varian Inova NMR spectrometer at 11.7 T with a triple resonance HCN probe (*QANUC, Montreal, QC*) and 3) an Avance III HD

NMR spectrometer at 18.8 T with a TCI cryoprobe (*QANUC, Montreal, QC)*. ^{1}H frequencies were referenced to DSS at 0 ppm and ^{15}N frequencies indirectly referenced to the ^{1}H zero-point DSS frequency[214] and all datasets processed using NMRpipe.[212]

5.2.4. ^{15}N Relaxation Analysis

The chemical shift assignments detailed in ***Chapter 4*** were used to assign the N-H resonances in each relaxation-modulated HSQC experimental data set using CcpNmr Analysis 2.2.1.[213] Peak intensities at various delay times for a given experimental data series were exported from the respective experiments. The baseline RMSD error for each spectrum was individually estimated using the following NMRpipe command at Unix prompt:

showApod -in t1_001.ft2 -noverb

where t1_001.ft2 corresponds to the first plane (in the case of a pseudo-3D experiment) or the first ^{1}H-^{15}N spectrum (in the interleaved experiment). *In-house* python scripts were used to input experimental parameters, resonance assignments, peak intensities and omit resonances with overlapping peak intensities to be omitted from the analysis (***Table 5.1.***). The graphical user interface in *Relax 5.0.0*.[376] was used to perform data fitting by a two-parameter mono-exponential decay:

$$I_t = I_0 \cdot e^{-Rt} \qquad \textbf{(5.16.)}$$

where I_t is the peak intensity observed at a delay time t, I_0 is the original peak intensity and R is the appropriate relaxation decay constant.

Table 5.1.: LIST OF RESONANCES WITH OVERLAPPING INTENSITIES	
Protein	**Resonances with overlapping peak intensities**
WT W_1	11, 22, 31, 34, 35, 74, 82, 89, 91, 97, 103, 106, 114, 117, 124, 131, 137, 138, 141, 143, 145, 149, 150, 153, 157, 160, 162, 165, 166, 167, 170, 177, 188, 189
Reduced W_1M	18, 31, 35, 50, 58, 69, 74, 91, 97, 103, 122, 138, 141, 144, 149, 150, 151, 152, 153, 157, 162, 163, 164, 165, 170, 177, 188, 189, 194
Stapled W_1M	9, 20, 21, 22, 35, 45, 48, 51, 55, 58, 62, 69, 72, 73, 89, 90, 106, 108, 111, 115, 116, 117, 122, 129, 131, 149, 157, 164, 167, 170, 188, 189, 195

The relaxation time T, (T_1, T_2 or $T_1\rho$) was calculated as the reciprocal of the corresponding delay constants[358] as in

$$T = \frac{1}{R} \tag{5.17.}$$

with the error for T estimated as

$$\sigma_T = \frac{\sigma_R}{R} \cdot T \tag{5.18.}$$

where σ_R is the error obtained for R from the *RelaxGUI* analysis.[376] Following T_1 and $T_1\rho$ estimation as in ***Eq. 5.15.***, T_2 was calculated as

$$\frac{1}{T_1\rho} = \frac{\cos\beta}{T_1} + \frac{\sin\beta}{T_2} \tag{5.19.}$$

with

$$\tan\beta = \frac{Spinlock\ field}{offset}$$

(**5.20.**)

Once the exponential decay fitting for T_1 and T_2 was completed, reduced spectral density mapping was subsequently performed on *RelaxGUI 5.0.0,*[376] as detailed in section ***5.1.4.*** In the case of CT-CPMG experiments, data fitting was performed in a similar manner on *RelaxGUI 5.0.0.*[376] and R_2^{eff} obtained henceforth was plotted as a function of residue number.

5.3. RESULTS

To check for discrepancies arising from sample - preparation and quality (i.e., whether this became compromised over time with a lag between triple-resonance backbone assignment data collection and relaxation experiment data collection), sample concentrations and field strength-dependent variations,[276, 377] ^{15}N relaxation experiments were repeated. Specifically, these were repeated at 1) the same field strength (18.8 T) on a different magnet and 2) at a different field strength (11.7 T), with independently prepared samples. All three of these data sets are analyzed and discussed below.

5.3.1. Alteration in local dynamics of protein backbone

[^{1}H]-^{15}N het-NOE measurements[353] carried out on WT W_1 (with DTT) showed positive enhancement throughout the folded protein core and a reduction in positive enhancement or a negative enhancement in the linker and terminal regions (***Fig. 5.5.a).***). This supports the localization of folded vs. disordered domains and follows previously reported studies for wildtype W_1 and W_2 by *Tremblay et. al.,*[5, 276] at 16.4 T in the absence of DTT and as observed from a chemical shift-based standpoint for WT W_1 with DTT in ***Chapter 4***.

The similarity in the NOE enhancement factors of reduced W_1M and WT W_1 over the folded domains, i.e., the helices (***Fig. 5.5.b))***, provides further support for the combined chemical shift analyses (***Fig 4.8.***) discussed in section ***4.3.6.*** that suggests comparable secondary structuring in these proteins. Interestingly, enhancement factors showed a minor increase in H5 and an evident decrease in the C-tail region for the reduced W_1M relative to WT, suggesting changes in backbone dynamics in the ps-ns timescale. This could be attributed to the substituted Cys side chains which are being maintained in the reduced form with DTT and resultant steric effects in proximal regions relative to the WT with less hindering Ser residues.

In contrast to this, the stapled mutant exhibited a different behaviour. Namely, there was a relative decrease in the NOE for residues in the H5 region alongside an evident increase in the C-tail region evident. Even though the NOE trends observed over the folded core are similar between the stapled mutant and the WT, the observed localized decreases in magnitude are suggestive of a decline in the overall structural rigidity, especially for the loops interconnecting the helices. (***Fig. 5.5.***).

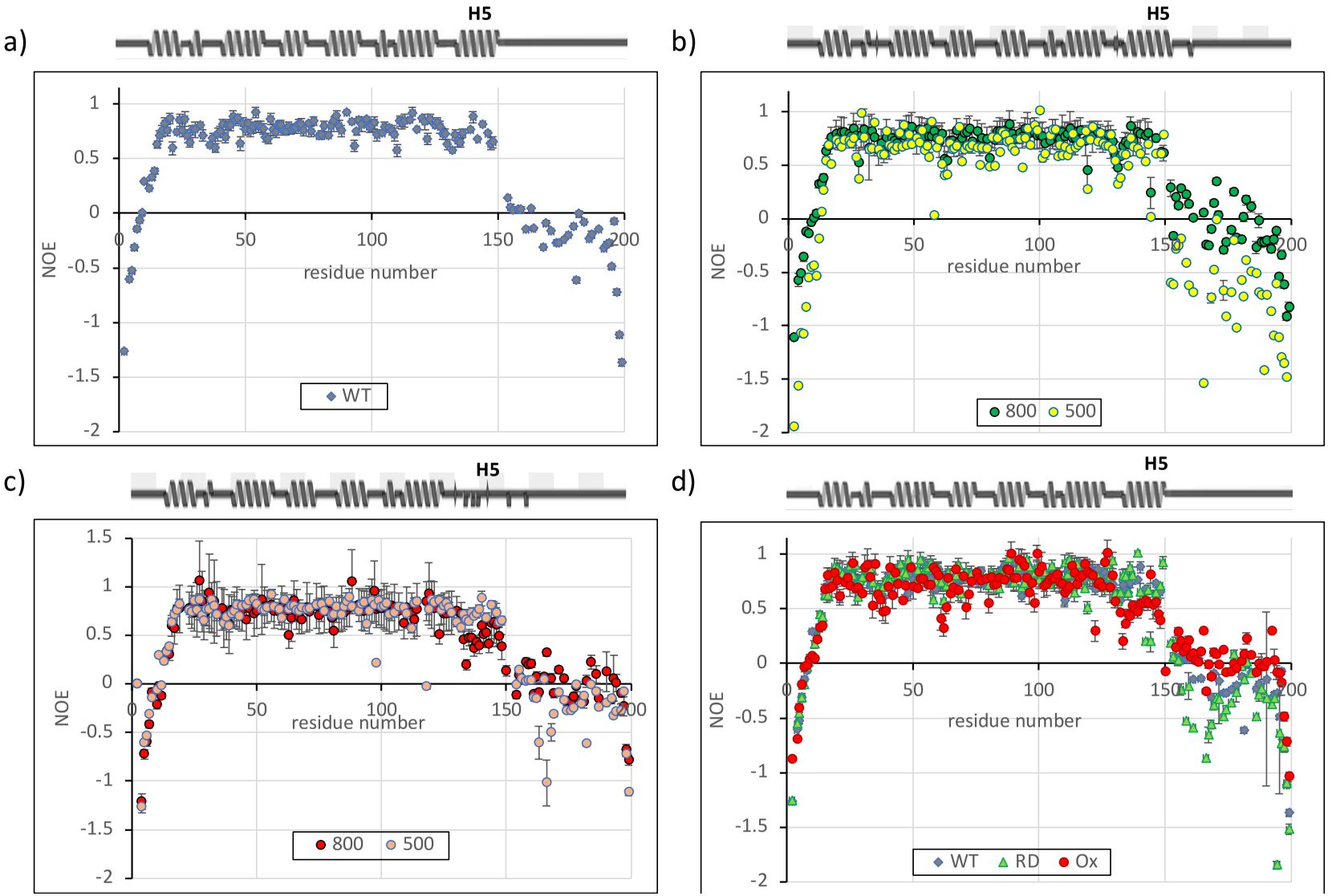


Fig. 5.5.*[^{1}H]-^{15}N heteronuclear NOE enhancement factors as a function of residue for a) WT W_1 at 18.8 T; b) reduced and c) stapled W_1M collected at 11.7 T and 18.8 T, and d) overlayed plots for all three proteins at 18.8 T.*

5.3.2. Variation in longitudinal and transverse relaxation time constants

The ^{15}N longitudinal and transverse relaxation rates, T_1 and T_2, for WT W_1 exhibit contrasting trends for the structured vs. the unstructured regions (***Figs. 5.6.a)*** and ***5.7.a))***. As with the het-NOE enhancement, this follows the same trend as was previously reported by *Tremblay et. al.*[5, 276] for WT W_1 and W_2 in the absence of DTT at the distinct field strength of 16.4 T.[5, 276] Specifically, slightly lower T_1 and dramatically higher T_2 values are observed for the N- and C-termini (N: 1-11 and C: 150-199), implying a slight increase in longitudinal relaxation rate and a much slower rate of transverse relaxation in these regions. Residues 12-149, corresponding to the folded core of the protein, display trends opposite to that of terminal regions. Although both T_1 and T_2 demonstrate these contrasting trends over the structured vs. unstructured regions, the difference is much pronounced in T_2 owing to the larger differences in magnitude.

Both the reduced and stapled forms of W_1M generally maintain the above discussed disparity over structured and unstructured regions. However substantial differences are observed in the magnitude of T_1 and T_2 values. (***Figs. 5.6.*** and ***5.7.).*** The observed T_1 trends are also distinct at 11.7 T relative to 18.8 T, whereas T_2 exhibits very similar behaviour at both field strengths. The trends at 18.8 T are discussed first. Namely, at 18.8 T for the reduced mutant, resides in the protein core experiences faster longitudinal relaxation and slower transverse relaxation relative to WT. Notably, in this mutant, both longitudinal and transverse relaxation mechanisms are much slower over the C-terminal regions at 18.8 T. Quite different trends are observed in the stapled protein. In this instance, the folded core experiences slower longitudinal relaxation while H5 experiences faster longitudinal relaxation, with a general decrease in T_2 values implying faster transverse relaxation over the protein as a whole. Although H5 ^{15}N spin relaxation behaviour was consistent with the remainder of the protein core in the reduced form, it is interesting to note its distinctive

nature in the stapled form – specifically, lower T_1 and higher T_2 values were exhibited in H5 at 18.8 T in comparison to the rest of the protein core. This resemblance between H5 and the dynamic and unstructured C-tail region (***Fig. 5.8.)*** supports the alteration or loss of the helical nature in this region of the protein in response to disulphide locking, as was inferred from the corresponding DANGLE-predicted secondary structure (***Fig. 4.9.).***

At 11.7 T, conversely, the globular protein core for both W_1M proteins exhibited a lower T_1 relative to the values observed in both the N- and C-terminal regions. Distinct from this, T_2 at 11.7 T follows almost the trend as at 18.8 T for both W_1M proteins, with slightly higher T_2 in the C-tail region at 11.7 T than at 18.8 T for the reduced mutant and slightly lower T_2 in the C-tail at 11.7 T than at 18.8 T for the oxidized mutant. At 11.7 T field strength, the reduced and oxidized forms of the protein show little distinction in their T_1 and T_2 behaviour relative to each other. Thus, although this trend of segregated relaxation behaviour for structured vs. unstructured regions was consistently observed at different magnetic fields, clear differences in T_1 relaxation were evident at 11.7 T and 18.8 T, which is a consequence of the direct dependence of T_1 on the magnetic field strength of acquisition,[209] as in ***Fig. 5.3.*** However, similar field strength-dependent variations were not that evident in NOE and T_2. Individual trends in T_1, T_2 and NOE as a function of field strength will not be considered further on their own, with the reduced spectral density mapping approach which considers all of these simultaneously instead employed (***section 5.3.2)***.

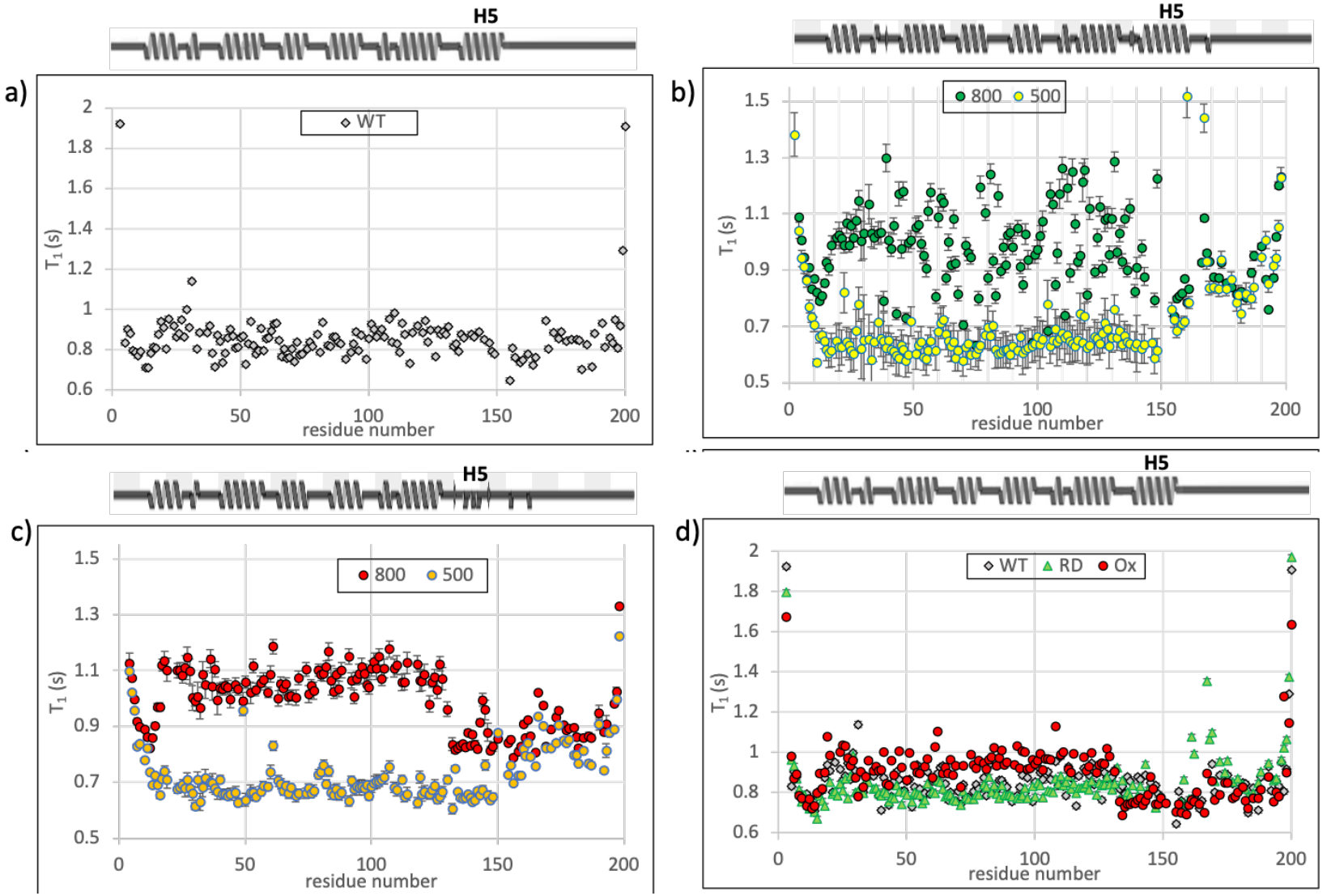


Fig. 5.6. *^{15}N T_1 relaxation time constants as a function of amino acid for a) WT W_1 collected at 18.8 T; b) reduced and c) stapled W_1M collected at 11.7 T and 18.8 T, and d) overlayed plots for all three proteins at 18.8 T.*

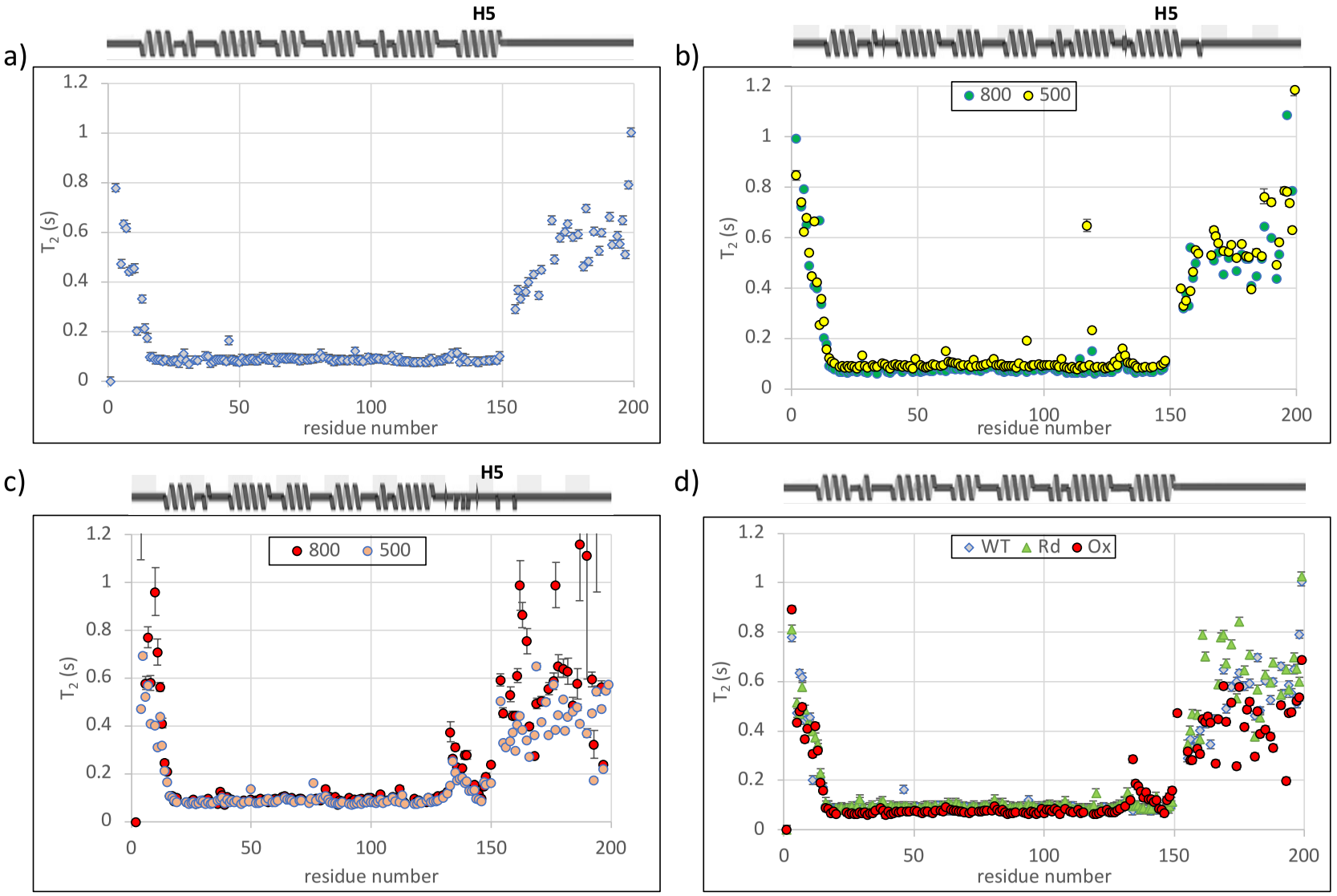


Fig. 5.7. *^{15}N T_2 relaxation time constants as a function of amino acid for a) WT W_1 collected at 18.8 T; b) reduced and c) stapled W_1M collected at 11.7 T and 18.8 T, and d) overlayed plots for all three proteins at 18.8 T.*

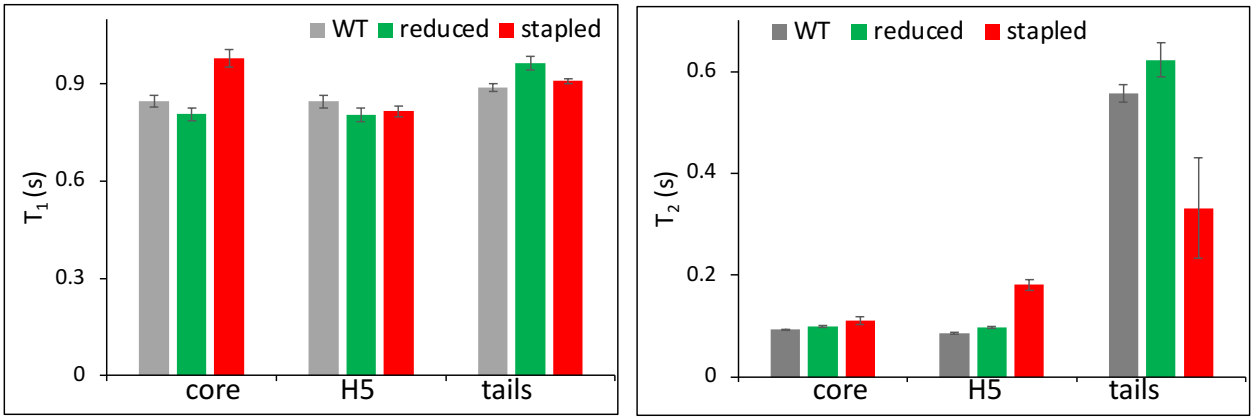


Fig. 5.8. *Bar graphs representing the mean T_1 and T_2 (with SD) for strategic regions in the W unit - core: 12–149, H5:135–149 and tails: 1–11 and 150–199.*

5.3.3. Rotational correlation time

Before moving into reduced spectral density mapping, the rotational tumbling behaviour as reflected in the T_1 and T_2 values measured for both mutants will be compared. Specifically, the ratio of T_1/T_2 was also evaluated for each protein and at both field strengths. ***(Fig. 5.9.)*** For regions of a protein with damped dynamics in the ps-ns regime (i.e., those residues exhibiting an NOE of > 0.65), this ratio is directly proportional to the rotational correlation time, τ_c. From the average ratios of T_1/T_2 for residues with an NOE > 0.65 in each protein, τ_c can be estimated as [378]

$$\tau_c = \frac{1}{4\pi\nu_N}\left(6\frac{T_1}{T_2} - 7\right)^{-(1/2)} \quad \textbf{(5.21.)}$$

where ν_N is the resonance frequency of ^{15}N (in Hz).

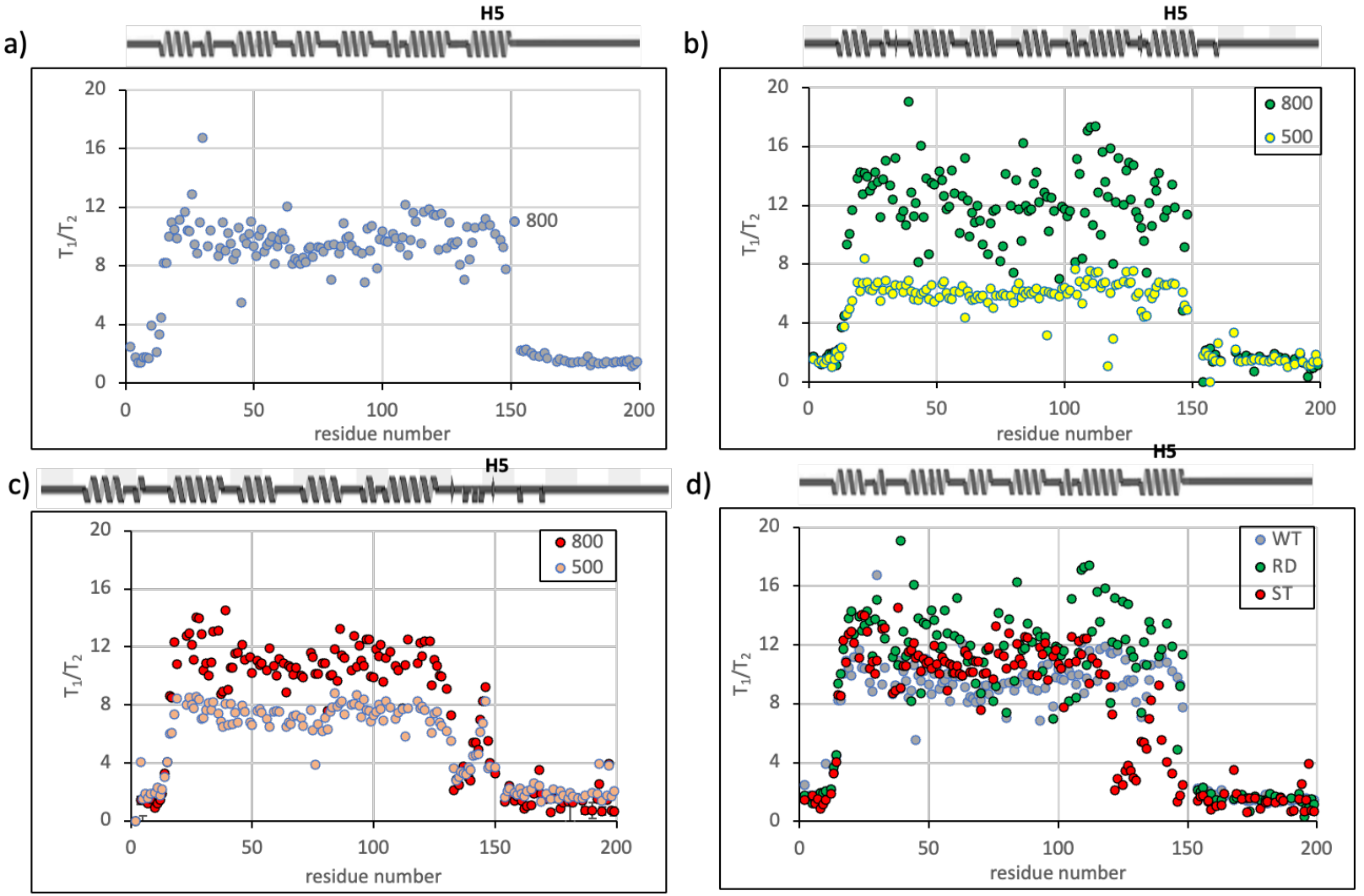


Fig. 5.9. *T1/T2 as a function of amino acid for a) WT W1 collected at 18.8 T; b) reduced and c) stapled W1M collected at 11.7 T and 18.8 T, and d) overlayed plots at 18.8 T.*

Table 5.2.: ROTATIONAL CORRELATION TIME FOR W_1 PROTEINS			
Protein ***(Concentration in μM)***	**T_1/T_2**	**Rotational correlation time*,τ_c (ns)**	**Translational diffusion coefficient from DOSY, D_c ($\times 10^{-9}$ $m^2 \cdot s^{-1}$)**
WT (125)	9.57	9.73	9.83†
Reduced W_1M (560)	11.99	7.55	8.32 ± 0.03
Stapled W_1M (580)	9.55	9.75	8.33 ± 0.03

**ν(^{15}N) at 18.8 T = 81.107 × 10^6 Hz. †DOSY value adopted from Tremblay et. al.[276]*

5.3.4. Reduced spectral density mapping

Although the model-free approach for analysis of protein backbone relaxation,[379, 380] is popular, this is not suitable in proteins with significant segments of intrinsic disorder.[381-384] Since W proteins contain long intrinsically disordered regions, reduced spectral density mapping was instead used for interpreting the relaxation parameters describing the ^{1}H -^{15}N bond motions for W_1 and W_2 proteins in the past[357, 359, 360, 385] and I have adopted this approach to analyze both of the mutant protein forms as well. Based upon the T_1, T_2, and NOE data collected and analyzed at both 11.7 and 18.8 T, I carried out reduced spectral density mapping analysis as a function of amino acid residue. The resulting reduced spectral densities mapped at 0.87(^{1}H),^{15}N and zero frequencies are shown in ***Figs. 5.10. - 5.12.***

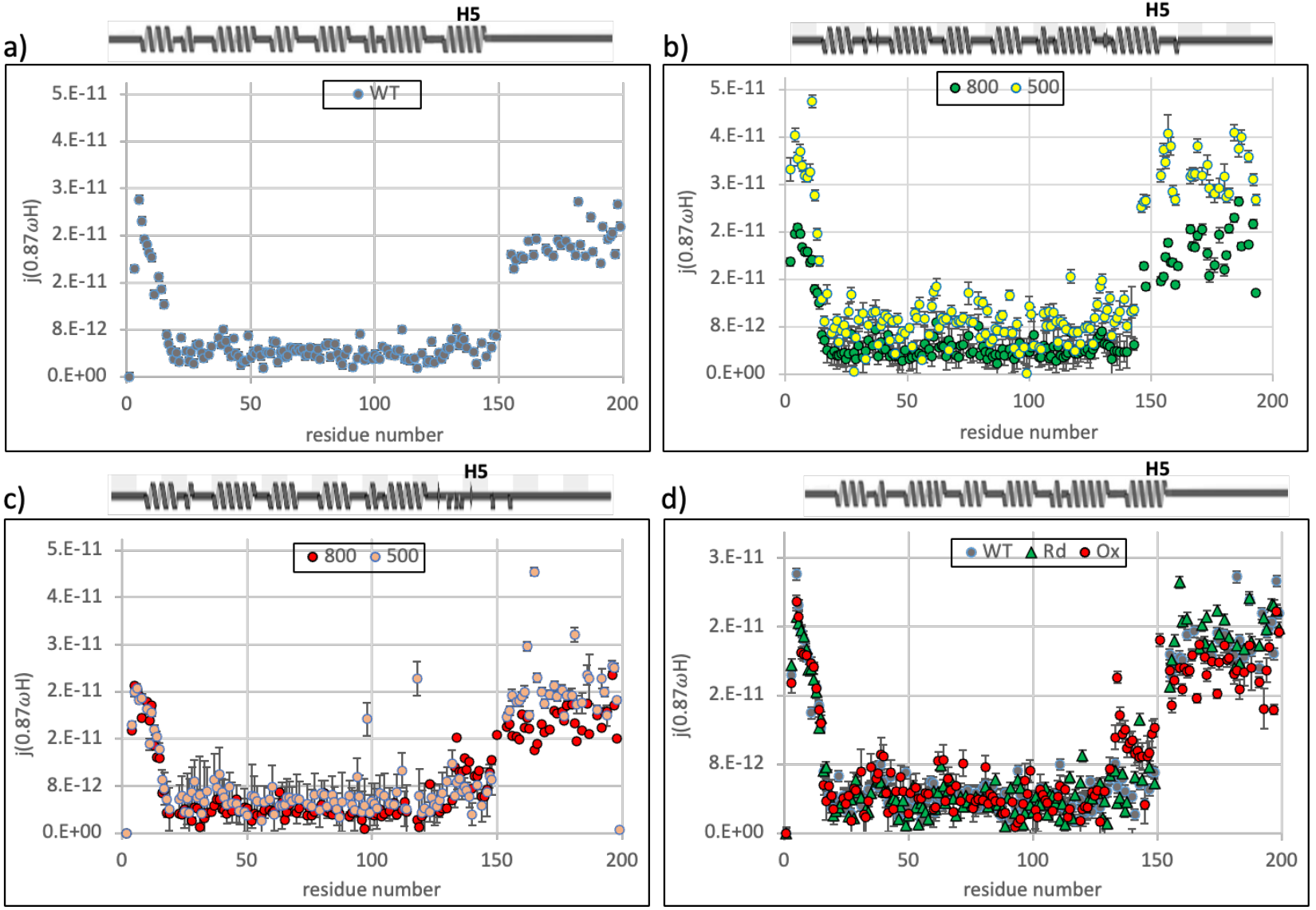


Fig. 5.10. *Reduced spectral density at $0.87\omega_H$ frequency plotted for a) WT W_1 collected at 18.8 T, b) reduced and c) stapled W_1M collected at 11.7 T and 18.8 T, and d) overlayed plots for all three proteins at 18.8 T.*

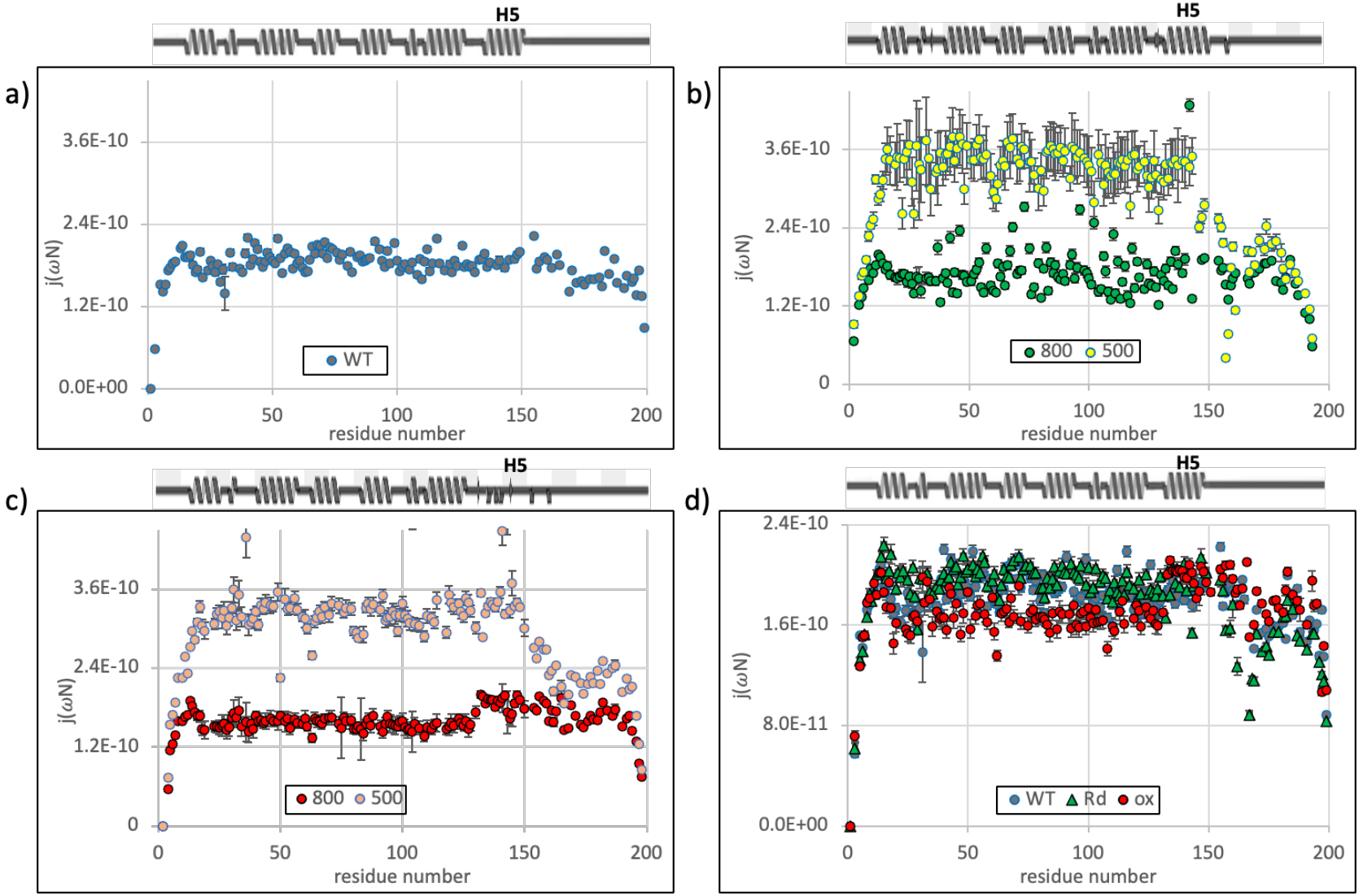


Fig. 5.11. *Reduced spectral density at ω_N frequency plotted for a) WT W_1 collected at 18.8 T; b) reduced and c) stapled W_1M collected at 11.7 T and 18.8 T, and d) overlayed plots for all three proteins at 18.8 T.*

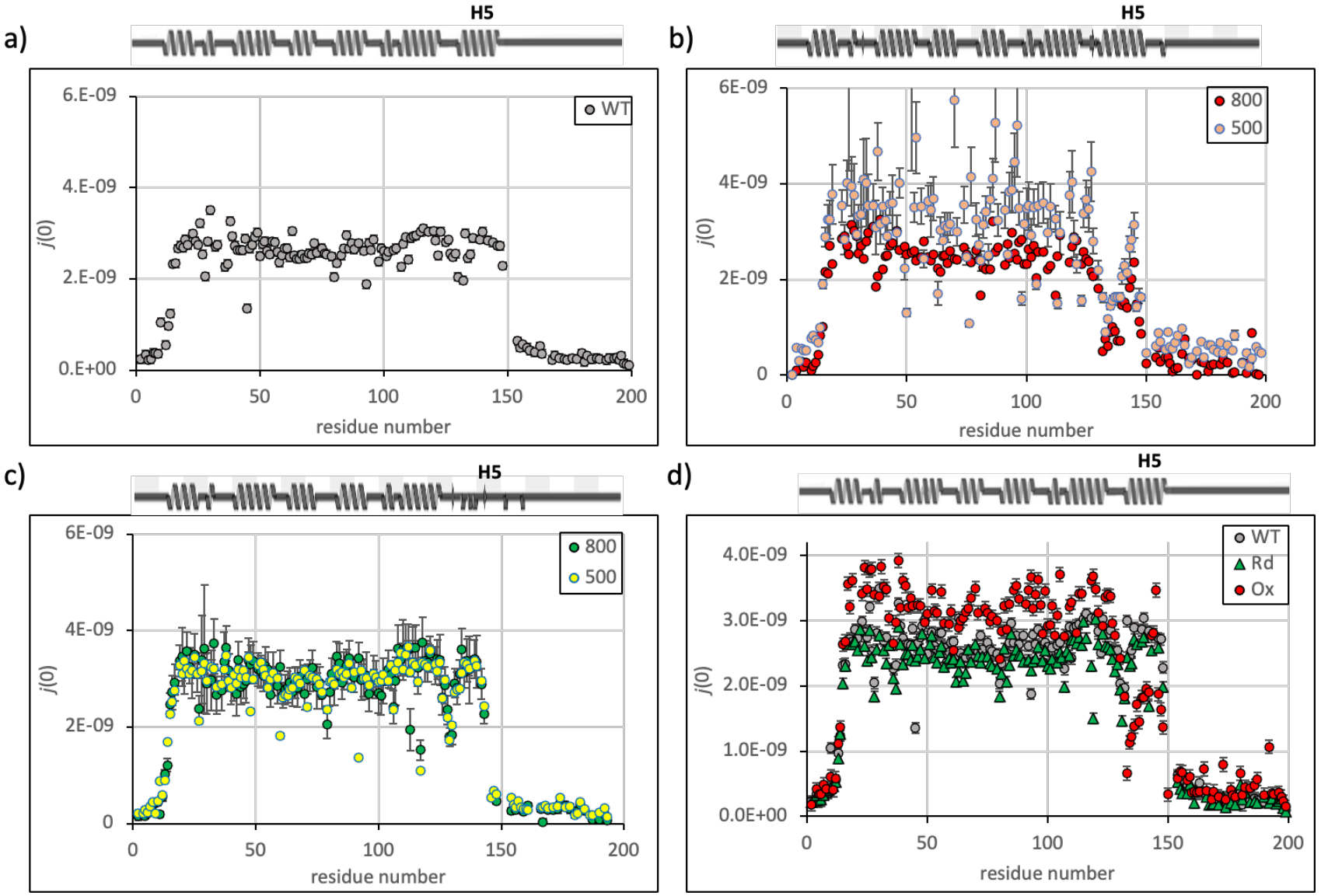


Fig. 5.12. *Reduced spectral density at zero frequency plotted for a) WT W_1 collected at 18.8 T; b) reduced and c) stapled W_1M collected at 11.7 T and 18.8 T, and d) overlayed plots for all three proteins at 18.8 T.*

Both the highest frequency component, $\langle j(\omega_H)\rangle$ and the zero-frequency component, $j(0)$, exhibit clear segregation between the structured core and N- and C-terminal tails (***Figs. 5.10.*** and ***5.12.)***. Specifically, relative to the globular core, $\langle j(\omega_H)\rangle$ is elevated in both termini while $j(0)$ is decreased. This trend indicates that higher frequency motions are more favourable in these terminal regions of the W unit regardless of the field strength employed for measurement and regardless of whether the protein is WT or either form of W_1M. Putting this in context, this echoes the trends observed in $j(\omega)$ in ***Fig. 5.3.*** as a function of slower or faster tumbling. $j(0)$ particularly strongly illustrates the differential dynamics between the folded core and the termini, echoing the inferences from the T_1, T_2 and het-NOE parameters detailed in ***sections 5.3.1*** and ***5.3.2.*** $\langle j(\omega_H)\rangle$ is sensitive to motions at high frequencies [357] and hence, larger $\langle j(\omega_H)\rangle$ values indicate flexibility and fast motions in the ps timescale.[386]

Considering localized changes, the $\langle j(\omega_H)\rangle$ component is slightly elevated for residues 11-17 (first helix) and 146-160 (part of H5 and the N-terminal region of the C-tail) independent of sample concentration and for all three proteins, a trend which is in agreement with the previous studies of the WT protein. This distinction in dynamics is also apparent in $j(0)$ with a drop in H5, especially so for the reduced W_1M, but not for the first helix. The loop regions within the folded core of the W unit can also be identified based on the localized increases in $\langle j(\omega_H)\rangle$ values, and decreases in $j(0)$ values, a feature that is more prominent at high field strength for $\langle j(\omega_H)\rangle$ (***Fig. 5.10.)*** and at low field strength for $j(0)$ ***(Fig. 5.12.).***

A significant difference is observed for $j(\omega_N)$ values recorded at different field strengths. with largescale differences observed for the folded core and tails for the data acquired at 11.7 T relative to 18.8 T (***Fig. 5.11.).*** Namely, at 18.8 T relatively little variability in the density for $j(\omega_N)$ is observed across the entire length of the protein, although relative to the WT W_1 an increase is

observed over H5 in the stapled W_1M vs. a decrease over H5 for reduced W_1M. Relatively low variation between the independent datasets determined at 18.8 T also provides credence to support these differences between field strength (***Appendix F).*** At 11.7 T, the trend in $j(\omega_N)$ more closely mirrors that in $j(0)$, with a drop in the density in the tails relative to the globular core.

The standard deviation for the spectral densities estimated using data recorded at 11.7 T is notably larger than the corresponding values inferred using data recorded at 18.8 T, owing to the inherently lower sensitivity at a lower magnetic field resulting in a lower peak dispersion. Apart from that, the magnet at 11.7 T is equipped with a room temperature probe while the other two at 18.8 T were equipped with much more sensitive cryoprobes. Other than this, the spectral density analysis based upon data acquired on three different magnets at two different field strengths rules out discrepancies in behaviour arising from the sample concentrations and serves to confirm the reliability of the observed results with some clear, valuable field strength-dependent differences in behaviour.

Comparing the reduced spectral densities of WT to those of the reduced mutant, a minor dip in magnitude at loop locations is evident from $j(0)$. In contrast to these, the stapled mutant demonstrates an overall elevation in magnitude with significant deviations in H5 and the following C-terminus. In order to quantify these changes, average spectral density values were determined over each distinct region of the W unit (***Fig. 5.13.***). Changes in the behaviour of stapled mutant relative to WT W_1 and reduced mutant are apparent**.** Specifically, while $j(0)$ is elevated throughout the stapled mutant, higher values of $\langle j(\omega_H)\rangle$ are observed for the protein core and H5. Interestingly, $j(\omega_N)$ values are reduced for both forms of the mutant in comparison to WT.

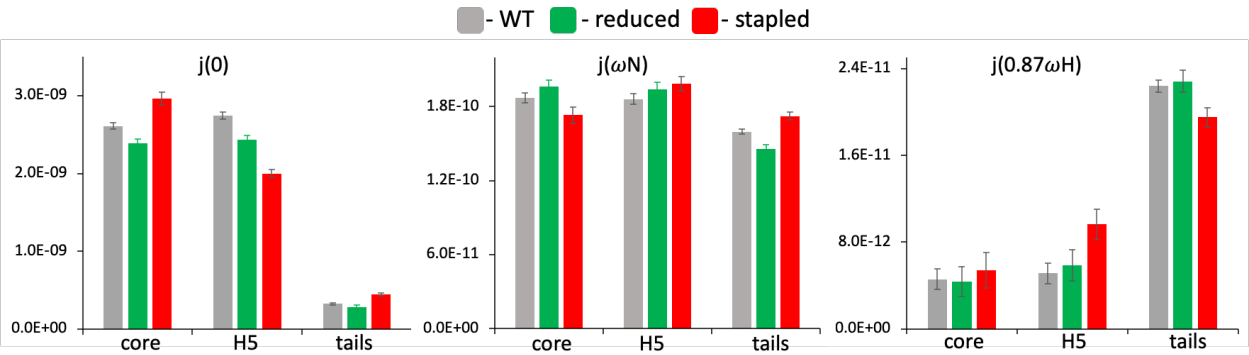


Fig. 5.13 *Bar graphs representing the average frequency difference in spectral density between the strategic regions in the W unit − core: 12–149, H5:135–149 and tails: 1–11 and 150–199 (from data collected at 18.8 T).*

5.3.5. No evidence for conformational exchange

With R_2^{eff} plotted as a function of residue number, except for one, no residues demonstrated significant values (***Fig. 5.14.).*** Hence the probability of conformational exchange, suspected to be the reason for the observed disparities between the two forms of mutant was ruled out in the µs-ms timescale.

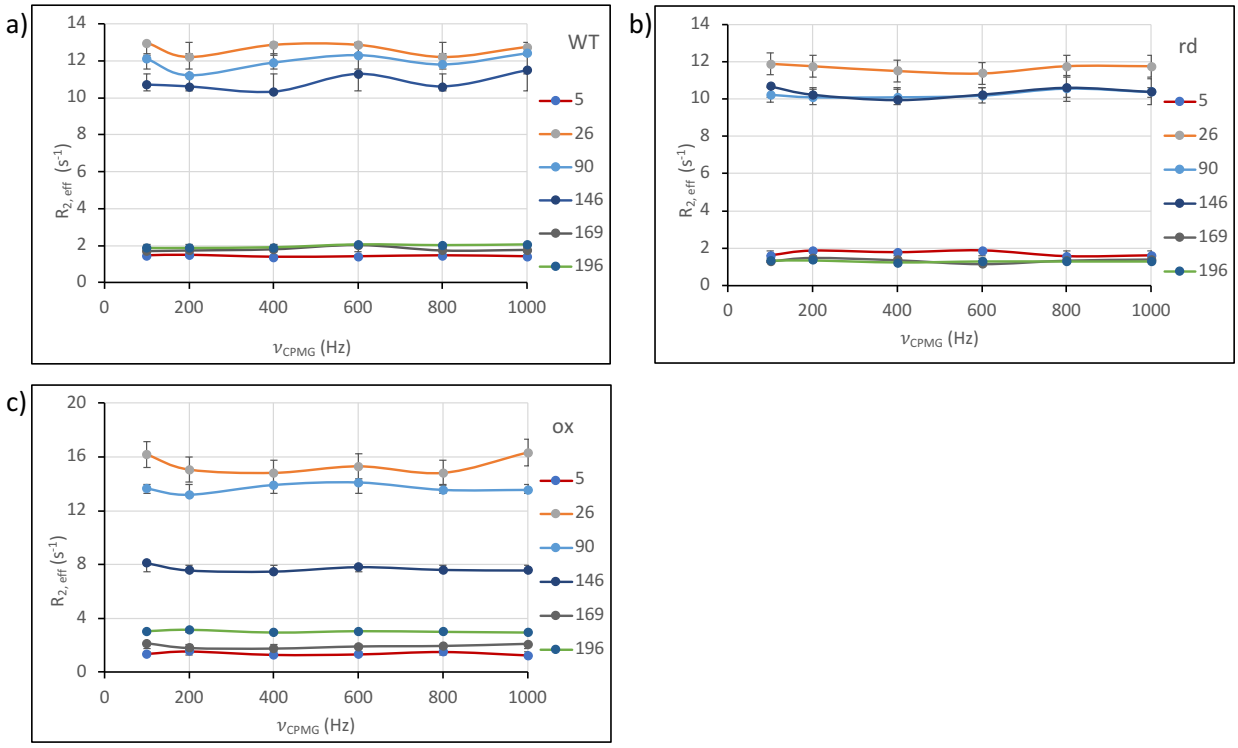


Fig. 5.14. *^{15}N R_2^{eff} relaxation decay constants plotted for various amino acids from strategic positions for a) WT W_1, b) reduced W_1M and c) stapled W_1M collected at 18.8 T. Residues were chosen from disordered N-terminal region (5), core region near the mutated Cys29 (26), hydrophobic core of protein (90), H5 near the mutated Cys143(146), initial region of the C-terminal tail flanking H5 (169) and C-terminal tail(196).*

5.4. SUMMARY

The ^{15}N spin relaxation parameters measured in this chapter mainly depend on N-H bond reorientations as a function of time, with contributions from both molecular tumbling and, when present, local motion within the molecule. This allows investigation of motions in the ps-ns (T1, T2, and het-NOE, as evaluated through reduced spectral density mapping) and µs-ms (CPMG-based relaxation dispersion measurements) timescales. Spin relaxation parameters estimated at two different magnetic fields herein allowed the quantification of both global and local dynamics within a molecular system. Since N-H bonds are mostly found in the protein backbone, results

obtained henceforth reflect motions of backbone peptide bonds occurring in the mentioned timescales.

Molecular motions of the backbone ^{15}N-^{1}H bonds were probed using het-NOE, R_1 and R_2 relaxation rates and CPMG-based relaxation dispersion experiments for W_1 and W_1M in both the reduced and stapled forms. R_1 and R_2 relaxation rate constants and heteronuclear NOE were determined as a function of residue. Further analysis using reduced spectral density mapping revealed that the repetitive W units display distinct trends in the stapled vs. WT/reduced mutant. Even though the reduced form of W_1M appears similar to WT, minor differences are observed in H5 and the C-tail dynamics. This might be the reason for the larger self-assembly and stronger fibres reported in ***Chapter 3***. Interestingly, the stapled version displayed a contrasting result where the dynamics of the C-terminal tail are reduced, with an increase in dynamics observed for H5. Disparities in the H5 and C-tail region of the stapled W_1M indicate attenuation of ps-ns dynamics in response to the covalent linkage introduced through intramolecular disulphide lock. Even though H5 adopts a less structured conformation, causing a resultant elongation of the C-tail altogether, motional restriction of one of the side chains in H5 onto the protein core is limiting the otherwise fast motions operating in the C-tail. Given that such drastic changes are observed in a single W unit, it would be worthwhile to extrapolate these backbone dynamic studies to W_2 mutant proteins to understand the behaviour of linker regions (C-terminal of one W unit connecting to the N-terminus of the other W unit) in response to the disulphide lock.

Relaxation dispersion experiments did not provide any evidence for conformational exchange occurring in any of these protein samples, which was suspected as the driving force behind all the disparities observed between the two mutant forms. With the relaxation experiments repeated at different field strengths (11.7 T and 18.8 T) and on different magnets (datasets collected on two

magnets as 18.8 T), discrepancies arising from sample quality or effect of magnetic field strength were ruled out, revealing the intriguing role of conformational freedom at the atomic level influencing the self-assembly within the dope state and quaternary fibre structure.

Chapter 6

CORRELATING EFFECTS OF DISULPHIDE-LOCKING UPON STRUCTURE, DYNAMICS & SELF-ASSEMBLY IN THE WRAPPING SILK REPEAT UNIT

6.1. Role of site-specific mutation in the soluble state of protein

6.1.1. Protein expression and oxidation

In recombinant W unit protein expression, the dame recombinant expression protocol was used for both WT and mutant plasmid DNAs. Interestingly, better expression yields were apparent for the mutant W unit in comparison to WT (both W_1M and W_2M), promising towards prospective large-scale production of this particular variant. Given that the second Cys mutation replaces one of the Ser residues from the poly-Ser region, greater variation is codon usage and subsequently lowered demand for Ser tRNA is suspected as a contributory towards such improvement in protein expression rates.

Owing to the reduced cytoplasmic environment within *E. coli*, the freshly recovered mutant proteins will be in the reduced form. During self-oxidation, even though the -SH groups of the mutated Cys residues are proximal to each other and are able to assume favourable side-chain dihedral orientations for disulphide bond formation,[387] it was difficult to accelerate the oxidation process. The inability to accelerate disulphide bond formation may be attributable to a buried positioning of the mutated residues, which would render them inaccessible/inert to oxidizing reagents.

6.1.2. Thermal stability and refolding ability

The observation of similar melting curves for both mutant forms of the W unit under thermal denaturation (detailed in ***section 3.3.5.***) suggest that stabilization through non-covalent interactions is prevalent over the introduction of a single disulphide bond. This behaviour of the mutant in either oxidation state could again be correlated to the position and side chain proximity of the mutated residues. Importantly, while the reduced form exhibits the capability to reversibly refold like the WT W unit, the temperature-induced irreversible secondary structural changes and increased aggregation propensity of the stapled form (***section 3.3.5.***) could be due to subtle changes in the free energy landscape – particularly for intermolecular interactions - arising from disulphide-induced motional constraints. It is interesting to note that there are comparable trends in a mutant MaSp1 construct that has two adjacent Ala→Cys mutations, where both reduced and stapled forms of this protein did not undergo disulphide-dependent oligomerization and exhibited a right-shifted thermal melting curve, all of which are attributed to the reduced tendency for conversion into β-sheet aggregates rather than disulphide-induced conformational stability.[388]

Given that protein turnover and proteolytic susceptibility are thermodynamically controlled, literature precedents suggest the contribution of *in-vitro* thermal stability towards higher *in-vivo* turnover rates.[389, 390] The unexpected improvement in protein expression rates and thermal stability obtained for both W_1 and W_2 with the noted Ser to Cys substitutions is not only beneficial from the perspective of these particular AcSp1 mutants but is also promising towards the genetic engineering of other prospective recombinant spider silk mutants with enhancements in yield, scalability and thermal stability.

6.1.3. Differences in protein structure, backbone dynamics and hydrodynamics

Previous studies in our group identified elevated dynamics in residues 146-160 of both the W_1 and W_2 WT proteins (with this numbering applied within a given W unit for W_2). This specifically fell within H5 (135-149) and the proximal disordered linker or C-tail regions of the W unit.[276, 277] Further to this, the H5 region was shown to be less stable than the remainder of the globular core whether subjected to urea, guanidinium chloride, or the detergent DPC.[5, 277] Based on these findings, this region of the W unit is postulated to undergo conformational changes that initiate protein-protein interactions leading to self-assembly in solution, which would, in turn, induce fibrillogenesis (***Fig. 3.2.***).

Secondary structural differences between the two mutant forms in CD spectroscopy (***section 3.3.5.***), when scrutinized at the atomic-level using solution-state NMR, revealed marked differences specifically at H5 (***Chapter 4)***. Interestingly, the CS-assignment based DANGLE predictions are suggestive of short segments with β-strand character around the mutation sites (***Fig. 4.9.***). ^{15}N nuclear spin relaxation experiments and reduced spectral density mapping correspondingly demonstrated that backbone-level dynamics are modulated not only in H5 but also extend across the disordered C-terminal tail of the stapled form. While the protein core of the reduced mutant very similar to WT, the C-tail region exhibits a significant elevation in dynamics (***section 5.3.4.)***. Given the distinct nature of H5 and the proximal C-tail region being hypothesized to initiate fibrillogenesis in WT, further verification is required to comprehend how these peculiar structural and dynamic differences are constructively contributing towards various stages of fibrillogenesis in either mutant form (vs. WT).

The translational diffusion coefficient, D_C, for W_1M is lower than that of WT not only in the reduced state but also in the stapled form, indicative of a larger hydrodynamic radius, r_H, for both

forms (***section 4.3.2)***. This could be a consequence of differences in the nature and extent of intramolecular forces atleast around the mutation sites, if not globally. Elongation of the C-tail through compromised helicity in H5 would render more flexibility and mobility to the stapled form. Apart from the elevated dynamics of C-tail, Cys side chains are suspected to bring changes to the reduced form – this was observed around the mutation sites as well as in the intrinsically disordered C-terminal region. While DTT-protected Cys residues would lead to an elongated side chain with greater surface area that would promote a greater degree of side chain-solvent interactions, -OH groups from DTT could also be strong contributors to H-bonding, thus altering the local interaction profile within the reduced mutant. This would in turn increase the hydration radius and thus hydrodynamic drag, reflected in a smaller diffusion coefficient. Hence, the disulphide chemistry, reduced helicity at H5 and reduced C-tail dynamics are inferred to be responsible for changes in the stapled form, while DTT protection of the Cys residues and elevated dynamics of C-tail might be the contributing to the differential trends in the reduced form.

6.2. Effect of disulphide lock on prefibrillar assembly

The dope state is considered to be the precursor for fibre formation where individual protein molecules undergo decompaction without substantial changes in secondary structure and protein solubility.[278] Prior work in the Rainey lab has demonstrated a key role for the choice of dope solvent in determining the nature of self-assembly before extrusion in the wet-spinning process. The degree of heterogeneity of the size distribution of such self-assemblies was directly tied to the resulting fibre quality, defined in terms of mechanical properties.[182, 278] I am proposing this distinction would result from differences in the preferential solvation of the protein by the dope solvent with the extent of solvation and nature of intermolecular interactions differing between solvent mixtures.

With the change of side-chain functionality from an -OH → -SH between the wildtype and the mutant protein, the ability to undergo homogeneous self-assembly was not affected in the reduced state. A morphological examination of the spinning dope state revealed that the reduced W_2M protein formed comparatively smaller particles relative to WT W_2 in the dope solution, although larger particles were formed over time owing to the dynamic nature of the dope state (as discussed in more depth below). This relative decrease in size could be attributed to the improved stability of the monomeric protein which would resist decompaction and retard the self-assembly process, resulting in smaller prefibrillar assemblies. However, introduction of the intramolecular -S-S- covalent linkage in the stapled mutant is observed to correlate with heterogeneous self-assembly. Namely, relatively larger particles (micron- vs. nanometer-scale) and even nanofibrillar assemblies are witnessed in the dope solution formed under conditions that are identical except for the oxidation state of the Cys residues. This contrast in dope-state self-assembly for the reduced vs. stapled forms of the protein is surprisingly similar to those previously reported for wildtype W_3 in two different dope solvents.[278]

According to my solution-state NMR studies (detailed in ***chapters 4 and 5***), disulphide locking 1) depletes the helical nature at H5, thus elongating the C-tail region and increasing the hydration sphere size and 2) reduces the stability of the stapled form, making the protein more susceptible to environmental changes relative to the WT/reduced mutant. Assuming dope solvent-mediated preferential solvation as the driving force for liquid-liquid phase separated self-assembly, I propose that conformational instability and elongation of the dynamic C-tail (H5 and C-tail in WT) in the protein chain promotes protein decompaction and accelerates intramolecular associations. This, in turn, leads to robust prefibrillar assemblies for the stapled form in the dope state. Hence, stapled is more prone to initial round of structural rearrangements, with disulphide altering the growth

kinetics in the succeeding stages. Whereas in reduced form, even though the initial processes are slow, latter stages are favoured for rapid fibre formation.

6.3. Effect of particle size distribution on fibre characteristics

Linear arrangement of the nanoparticles, as previously reported for WT, holds good in the case of reduced mutant as well, suggestive of linear fusion of these NPs towards prefibrillar assemblies and microfibrils leading to fibre formation. Relatively large prefibrillar assemblies and nanofibrils observed in the dope state of the stapled W_2 mutant suggest an increased affinity for primary aggregate formation. However, inability of these proteins to form regular fibres through wet spinning indicate that these microfibrils (observable by naked eye) fail to hierarchically associate, as in WT or reduced mutant protein. This could be attributed the loss of chain plasticity that limits exposure of (buried) hydrophobic residues and conformationally accessible transient structures, in turn compromising with extensive intermolecular interactions that drive fusion of prefibrillar assemblies and interfere with their ability to bundle microfibrils into stronger fibres.

In converse to this, the observation that the reduced mutant forms relatively smaller nanoparticles (vs. WT) in the dope state that undergo extensive fibrillogenesis suggests a distinctive yet favourable energy landscape for the reduced form vs. WT. This may be achieved through the promotion of decompaction and elongation of the protein chain, exposing buried hydrophobic residues for interfibrillar association and, hence, fibrillogenesis.

In native spidroins, Cys residues in the conserved CTD domain (absent in the W unit) are reported to guide unidirectional arrangement of protein chains through disulphide-mediated dimerization followed by controlled micellar arrangement, a significant step in fibre formation.[391] Free Cys/DTT protecting Cys is suspected to perform a similar role through interchain association.

Even though an -SH-containing side chain would not promote H-bonding, R-S-S- observed at low pH conditions is reported to be highly reactive. Suspecting a scenario similar to that of scleroproteins of silkworm silk, wool, keratin and vulcanized rubber,[392-394] the reduced mutant fibres should be probed for potential interchain disulphide bonding, herein proposed to mediate fibre strengthening through interfibrillar cross-linking and hence mechanical properties in the reduced mutant.

6.4. Fibre response to water: Supercontraction

Natural spider dragline silk fibres exhibit softening and shrinkage in response to water.[118, 395, 396] This phenomenon leads to a decrease in length accompanied by an increase in fibre diameter and is caused by the wetting of fibres by water molecules and subsequent breakage of intrafibrillar hydrogen bonds and reorientation of secondary structural motifs. Depending on the amount of water incorporated within a fibre, such responses can be qualitatively classified as either cyclic relaxation or supercontraction.

Backbone hydrogen bonds of disordered regions and α-helices are comparatively weaker than those of β-sheet motifs due to orientational specificity. Water molecules are initially incorporated through H-bond rearrangement and motif reorientation within the disordered and helical regions of an unstrained fibre. This is a reversible contraction response to wetting and drying that occurs in the amorphous regions. It is characterized as being unidirectional, with contraction being similar in magnitude over the affected region, with this phenomenon often being attributed as the key feature providing torsional shape memory to spider silks.[6] However, extreme humidity or exposure to liquid water induce stress even in the crystalline regions of restrained fibres. This disrupts β-sheet motifs, causing a permanent structural transition from a highly oriented glassy phase to a

disoriented rubbery phase.[180, 396] The result of this transition is irreversible radial swelling characterized by a decrease in fibre length and increase in cross-sectional diameter, termed supercontraction. This phenomenon is reported in native dragline silk,[396-398] and is attributed to specific sequence motifs.[399, 400]

Since response to water can differ according to hydration rate,[401] the possibility of reversible cyclic relaxation, its efficiency and the limit of its applicability before supercontraction need to be tested through humidity-controlled experiments. Supercontraction occurs at ambient temperatures in natural silk, while its induction requires elevated temperatures or harsh solvents in artificial fibres.[332] Given that supercontraction has not been reported in WT AcSp1 or any other recombinant aciniform silk, the potential that excess water exposure drives supercontraction in this mutant at room temperature is a fascinating scenario. With such response being absent in WT recombinant fibres, the specific role of the mutation in inducing this behaviour needs to be addressed.

6.5. Summary

Structural instability and elevated local dynamics in the soluble protein promote decompaction aiding the initial nucleation events required for robust liquid-liquid phase separation. This is quite evident from the stapled form of mutant being highly prone to prefibrillar assembly in the dope state (vs. WT/reduced form), owing to the elongated C-tail region. Contrasting trends in dope-state self-assembly for the identically prepared reduced vs. stapled form of the same protein demonstrates the critical role of conformational freedom in determining the nature of self-assembly, suspected to be the primary precursor for fibrillogenesis. Secondary nucleation events are guided through secondary structural changes and associated decompaction of the individual protein molecules that expose buried hydrophobic residues to promote extensive intermolecular

interactions required for microfibrillar arrangements. This step is retarded/ arrested in the stapled form due to disulphide locking of the protein chain and might be the cause for fibril fragmentation (brittle nature of microfibrils), similar to those reported in A-β proteins. However, presence of Cys residues promote these processes in the reduced form. Fibre mechanical properties echo the extend of intermolecular forces operational within the nano/microfibrillar assemblies – a greater number of forces strengthen such associations. Elongational forces during post-spin stretching procedures promote further orientation of secondary structural motifs and association of microfibrillar units (visible from the fusion of microfibrils in ***Fig. 3.15.b***), hence strengthening fibres.

Previous results from the native protein have served as the basis to understand the structure and function of AcSp1 proteins. Further NMR measurements of the sort described here, along with Numerous biophysical characterizations conducted on a closely related variant (double Cys mutant; detailed in ***chapters 3-5***) has broadened our understanding about this fascinating class of protein molecules providing crucial insights about various stages of soluble protein-to-solid fibre conversion. Improved fibre mechanical properties indicate the potential of fibre quality improvement by 1) fine-tuning the self-assembly formation occurring in the dope stage, and 2) through the usage of mutants to achieve the same, thus promising bioengineering of superior fibres with even few W units. This latter point is quite important since expression yields tend to become lower as protein molecular weight is increased.[7]

Chapter 7

CONCLUSIONS AND FUTURE DIRECTIONS

7.1. SORICIDIN

7.1.1. Conclusions

My success in developing the first reported methodology to produce full-length soricidin through recombinant technology using a bacterial system like *E. coli* provides a cost-effective means for industrial scaling the production of this protein. According to Anfinsen's thermodynamic hypothesis, the native structure of a protein is the thermodynamically stable structure.[402] Higher expression rates in the C41(DE3) *E. coli* strain are in agreement with the improved effectiveness of transformation and tolerance to overexpression-induced toxicity in this mutated strain of BL21(DE3), as previously reported in the case of many toxic and membrane proteins.[403, 404] Working under the conjecture that the bioactive conformer would be the state with the global minimum in the disulphide-bonding and folding energy profile and, hence, the most favoured conformer, the concept of enthalpic trapping was adopted here. Importantly, this was implemented through a thiol-disulphide exchange-mediated oxidative folding and was shown to dramatically augment the process involved in establishing the correct disulphide bridging pattern required for the functional bioactive conformer of soricidin, bringing down the time from ~14 days (as reported by *Soricimed Inc.*) to ~12-16 hours. Such a radical improvement in the rate of trapping of the correct folded state, complemented by a higher refolding efficiency (~95% vs. 60% previously reported by *Soricimed Inc.*), could be highly beneficial for future large-scale industrial production where a 14-day wait with ~60% efficiency in product recovery would be enormously detrimental. Conserved Cys motifs in the primary structure, a CSα/α helix-loop-helix hairpin fold, a potential surface dyad, and a pseudo cylic secondary structure were all identified on the basis of the tertiary

structure reported herin. Pronounced similarities and differences of this protein to previously reported (venom- and plant-derived) proteins (discussed in ***section 2.5.***)[88, 95, 242, 255-261] derived from very different sources, exhibiting variations in protein size and sequence, changes in number, conformation and positioning of disulphide bonds, the spatial orientation of helices and the interconnecting loop length,[93] suggest the probability of both convergent and divergent evolution mechanisms being operative.

7.1.2. Scope of the structure

In a protein like soricidin, which has been shown to have multiple classes of bioactivity, the identification and isolation of domains responsible for each of these functions are crucial for research and development. In particular, this compliments drug design by allowing for the potential avoidance of undesired secondary effects arising from distinct functional domains in a multi-domain protein under the clinical and pharmaceutical application. To date, soricidin has been shown to be bifunctional, having the ability to target and modify with more than one biochemical pathway. The comprehensive understanding of the 3D structure achieved through my work will directly support such domain separation. With the dynamic C-tail, constituting SOR-13 which is now being tested as an emerging anti-cancer drug, resolving the tertiary structure of soricidin in the soluble form could also support the prospective pharmaceutical development of the structured N-terminal region of soricidin as a paralytic or analgesic drug. Over and above this improvement in understanding of soricidin domain structuring, the identification of the localized surface positioning of the putative dyad provides greater scope for the development of shorter functional derivatives and could be used to direct protein engineering efforts to improve the specificity/potency of this bifunctional protein as a molecular probe or drug.[273]

Hence, the work carried out and detailed in ***Chapter 2*** introduces a novel, optimized protein production, purification, and folding procedures enabling me to provide the first structural information about a paralytic mammalian venom-derived protein, soricidin. These protocols and data are likely to be key in ongoing applied research and pharmaceutical advancement based upon soricidin.

7.1.3. Future directions

Higher levels of protein expression were achieved with the C41(DE3) *E. coli* strain, and the novel refolding protocol that I developed have immensely improved bioactive soricidin recovery levels. Slow expression rates, high cost and issues with scaling up restrict the usage of eukaryotic expression systems like Chinese Hamster Ovary (CHO), yeast, insect cells etc. for expressing complex multi-disulfide-bonded proteins. Given relative cost-effectiveness, high speed, user-friendliness, better optimized expression protocols and availability of large numbers of genetic tools for optimization purposes, prokaryotic systems are still the most popular expression systems.[405] With the increasing popularity of SHuffle strain, an engineered mutant of *E. coli* capable of promoting disulfide bond formation in the cytoplasm, effectiveness of this expression system for the production of soricidin should be tested as this could further reduce the downstream processes involving purification, refolding, etc. and hence improve protein recovery rates, and reduce production cost.[405]

Two primary causes for the second, lower intensity set of soricidin resonances observed by solution-state NMR spectroscopy were noted. First, an impurity remaining following recombinant protein expression, purification, and oxidative refolding may be the cause. This would be consistent with the unidentified peaks in mass spectrum, although is hard to rationalize with respect

to the purity implied by RP-HPLC chromatogram analysis.) Alternatively, conformational heterogeneity arising from slow exchange between two states may be the cause for the second set of resonances. Owing to the low number of observed NOE contacts for this second, lower intensity set of resonances, structural calculation could not be performed for the minor conformer. To distinguish between these hypotheses, a synthetic peptide with the elucidated disulphide bridging is being procured. NMR spectroscopy and/ or mass spectrometric analyses would be beneficial to test if the observed resonances are a cross contamination from recombinant protein expression or a result of conformational sampling in the bioactive form of soricidin. If these peaks still prevail, I suggest a thermal denaturation study to identify the melting temperature of soricidin followed by NMR experiments at temperatures higher [406] than those investigated by me.

Finally, even though the basic-aromatic amino acid pair identified on the protein surface is in agreement with the functional dyad in other venom proteins, the role of this putative dyad in soricidin function needs to be established. The efficacy of the full protein as a venom relative to the enclosed loop alone also should be tested through binding studies owing to the disparities in their binding efficiencies reported in proteins with similar structures.[86, 88] These studies will be instrumental in the prospective development of the analgesic drug and domain separation required for pharmaceutical applications.

7.2. DOUBLE CYS MUTANTS OF ACINIFORM SILK

7.2.1. Conclusions

Both the single-repeat unit (W_1) and double-repeat unit concatemer (W_2) of variations of the double-mutant aciniform silk protein were produced recombinantly and successfully prepared in pure samples under two oxidation states: the reduced state, with -SH groups of the introduced Cys

residues maintained in the reduced form in the presence of DTT (or TCEP, in the case of CD spectroscopy); and, the stapled state with H5 locked to the protein core through an intramolecular disulphide bond(s) formed specifically within each W unit. The fibre formation ability was tested under each redox state of the W_2 concatemer, while W_1 proteins were employed in solution-state NMR studies. These latter studies relied upon the fact that the modular nature of the W unit was established for the mutant W unit in both oxidation states through comparison of NMR spectral properties, demonstrating that the W_1 mutants serve as tractable models for the study of W_2 biophysics.

Attempts to modify protein sequences to improve control of the mechanical properties of synthetic and recombinant fibres have been reported in the literature.[293, 407] In the case of the W unit, a series mutants with Trp substitutions were produced in our lab at four disparate sites within both W_1 and W_2, where each site was mutated in both units of W_2 (***Fig. 7.1.)***. These site-specific Trp mutations, whether in a native Trp or a ^{19}F-labelled 5-fluorotryptophan state, neither affected the global soluble structure (in W_1) nor the fibre-forming ability of W_2 concatemers of this protein.[277] This prior study implied that the W unit is relatively amenable to the accommodation of local side chain mutation and concatenation, without significant structural perturbation(s). This also supports the modularity of the W unit and highlights a potential advantage in bioengineering this class of proteins.[3, 275, 277] Conversely to this, here I demonstrate that side-chain mutations evoked in two residues within the W unit, and the consecutive introduction of a disulphide lock as in the stapled form, compromises both the α-helicity and local dynamics of H5 (detailed in ***Chapters 4-5***.) alongside the native fibre forming ability of the protein (***Chapters 3***). However, it should be emphasized that in the reduced state, the Cys-substituted W unit maintains its structuring and forms W_2M fibres that exhibit augmented mechanics relative to WT W_2. As such, I have effectively

introduced and demonstrated a chemically switchable form of the aciniform silk protein, whereby reduction can be used to convert it from a state incapable of fibre formation to a state which produced improved fibres relative to the wildtype.

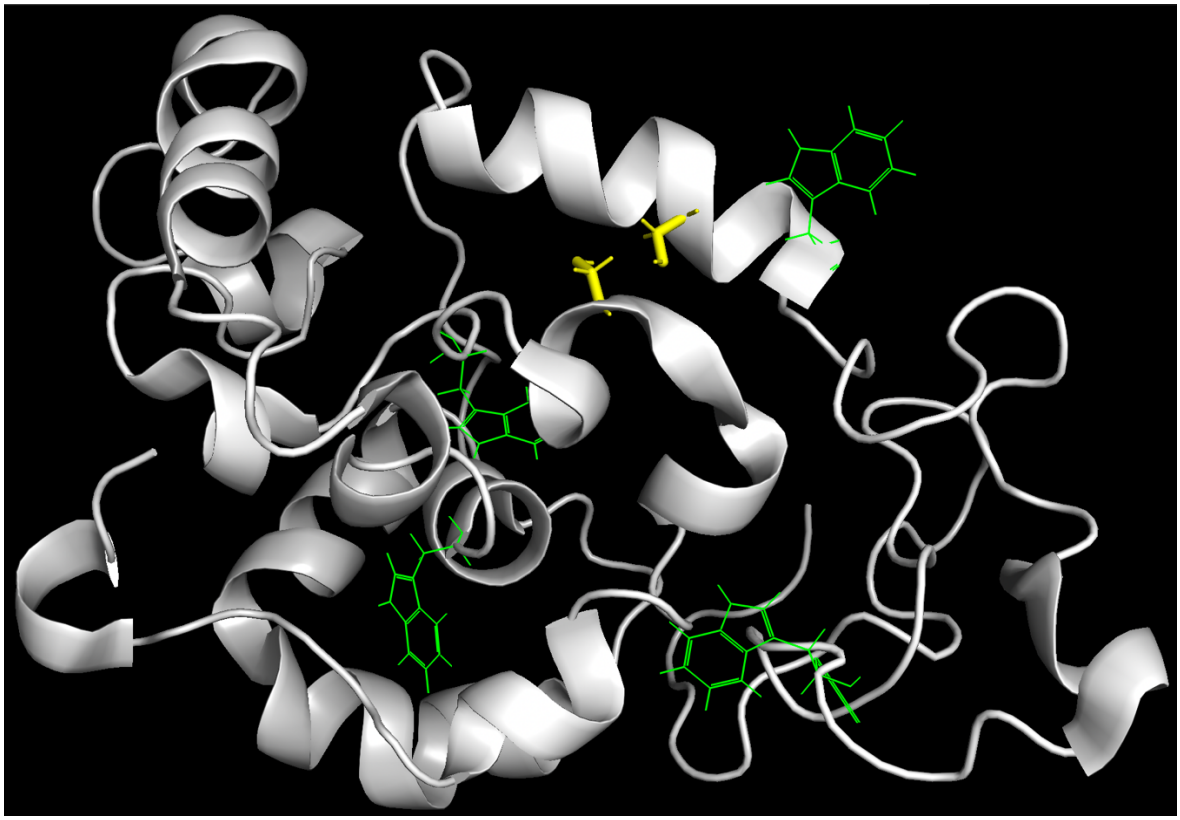


Fig. 7.1. *W_1 protein with tryptophan mutation at four sites. (cartoon representation in grey: protein backbone represented as ribbon, green: side chain of Trp residues at 36^{th}, 90^{th}, 146^{th} and 169^{th} positions represented as lines, yellow: Ser residues at 29^{th} and 143^{rd} positions as sticks)*

Correlative studies performed using my Cys mutantW_1M protein - produced in two oxidation states, the reduced and stapled - provide valuable insights about the critical role of motional freedom within the soluble silk protein state in initiating protein decompaction leading to self-assembly. This, in turn, affects the fibre formation ability of the aciniform silk repetitive unit and, ultimately, the mechanical properties of the fibres formed henceforth. Recognition of such a structural trigger for fibrillogenesis not only expands our understanding of the native processes

underlying the transition from the soluble protein state to solid fibres but also demonstrates the possibility of bioengineering conformationally switchable silk proteins with tunable properties.

7.2.2. Future directions

7.2.2.1. NMR studies on the native protein for unravelling fibrillogenesis

Although we know the solution-state structure of the W unit, it is also important to understand the protein secondary structural conformation in the solid fibre. Raman spectroscopy has provided preliminary understanding of the conformational state and orientation of both α-helical and β-sheet components in recombinant W unit-based fibres, but it is important to understand these details at the atomic level for rational protein engineering and material improvements or modifications. The optimization of appropriate isotope labelling scheme(s) followed by solid-state NMR studies are essential for this purpose. An atomic-level understanding of the protein in the fibre form could serve as a guide to identify the transient conformation(s) adopted by the protein during liquid-liquid and liquid-solid phase separations, i.e., an initial protein decompaction process with retention of secondary structural features followed by conformational rearrangements promoting prefibrillar assembly preceding fibre formation.

Dark-state exchange saturation transfer (DEST) is an NMR technique gaining popularity as a probe for characterizing the interaction between an NMR-'visible' free species and an NMR-'invisible' species, for example a species that is transiently bound to a high-molecular-weight complexe (>1 MDa).[408-410] Given that the NP-sized prefibrillar assemblies initially formed in the dope solution would fall in this size regime, the potential of this technique needs to be considered in light of the dynamic nature of the dope state. For example, even though W_1 proteins are incapable of forming strong fibres, the characterization of their prefibrillar assemblies/microfibrils could prove

informative towards understanding of their growth kinetics, as in the case of protofibrils of *Tau*, A-β or prion proteins.[408] W_1 may, thus, be an ideal model protein with which to apply DEST and understand AcSp1 pre-assembly.

To better apply DEST, and probe such otherwise "invisible" state, I propose that application of pressure-based perturbation may be beneficial. According to Hawley's theory, both temperature and pressure are crucial in the thermodynamics of protein unfolding. This may be described in terms of a elliptical pressure-temperature (p-T) phase diagram, which becomes more complex when intermolecular forces are also accounted for.[411-413] Chemical denaturation and structural perturbation studies were advantageous for preliminary identification of H5 dynamics.[5] Further investigations, however, could be restricted due to confinement of the protein in a small micelle/hydrodynamic shell volume, [414, 415] thus limiting the identification of heterogeneous populations due to conformational averaging.[413] This would limitat the ability to observe transient states, an issue that could be overcome by varying either/both temperature and pressure.[413, 416] Low pH conditions and chemical detaturants (GmHCl, urea etc.) in moderate amounts are reported to reduce the size of the p-T ellipse, thus favouring phase transitions. This could prove useful in driving phase transitions into experimentally attainable ranges during NMR studies. [195 417-422]

Based on our current understanding, silk fibre formation involves at least two types of phase transitions with intermolecular interactions driving subsequent hierarchical structural modifications in response to changes in pH, ionic concentration and shear force. Since these three factors are understood to play vital roles in the native fibre spinning processes, these parameters either independently or in combination could be modified to enable the identification of transient conformational states and, hence, identify the underlying mechanism of fibrillogenesis. Hence, a combination of the pressure, temperature, pH and/ or denaturants while performing VT and/or

high-pressure NMR studies alongside other biophysical characterization techniques could favourably expose transient states that might not be otheriwise visible. This would, hence, provide a comprehensive understanding of the thermodynamics and kinetics involved in fibre formation by spidroins.

Finally, given that H-bonding patterns and rearrangements of these play a crucial role in the structural transformation that occurs during fibrillogenesis, and given that the H^N CS temperature coefficient reflects the strength of hydrogen bond,[194,195] VT-NMR could be exploited to characterize local stability and quantify the destabilization as a function of position within the protein at various stages of fibrillogenesis. Apart from evaluating this in an aqueous state, I suggest examining the effect of dehydrating reagents like MeOH, EtOH, etc. to understand the consequences of destabilization of polar interactions and reduction of the influence of the hydrophobic effect.[413 422]

7.2.2.2. Applying reduced mutant for unravelling fibrillogenesis

The above detailed NMR studies could be extrapolated to the double Cys-mutant protein as well. One suggestion is to monitor the behavior of the linker region through the same series of structural and nuclear spin relaxation studies (detailed in ***chapters 4-5***) using W_2 proteins. Apart from this, I suggest the detailed series of investigations outlined below in order to understand the observations detailed in this project and aid in design of new protein constructs through rational protein engineering.

7.2.2.2.1. Probe free-energy landscape of mutant proteins

Given that protein function is dependent on the conformational freedom of the individual molecule, it is important to understand the free energy landscapes of the various AcSp1 proteins used in the current project as well as for bioengineering of novel mutants with desired functionality. Despite the observation of variation in both conformation and dynamics between the reduced and stapled W_1 and W_2 proteins, both forms of the mutant exhibit a similar melting point. However, discrepancies in the renaturation process suggest probable differences in both local and global responses to thermal energy. Namely, conformational restriction (in the stapled form) could be altering intramolecular non-covalent interactions that are responsible for the accessible conformational space of not only the mutated residues but also spatially proximal residues, concomitantly limiting or varying the nature, type and number of transient states that are otherwise thermally accessible to the reduced mutant protein. Such changes could potentially alter the kinetic and/or thermodynamic profiles corresponding to the free energy landscape of both mutant forms instigating alternate interaction pathways.[423, 424] VT-NMR and /or computational analyses should be performed to understand this aspect.

7.2.2.2.2. Characterization of mutant fibres

Due to the COVID-19 lockdown, mechanical testing was performed almost ~1.5 year after the initial wet spinning of the WT and mutant W_2 fibres. Testing of aged reduced W_2 mutant fibres has implied them to be mechanically superior to the WT state and is promising in terms of demonstrating a long-term shelf-life without compromising mechanical performance. However, the entire process of fibre spinning, and fibre characterization need to be repeated on a larger scale, without a significant delay between spinning and mechanical characterization, to confirm

reproducibility and ensure full control of conditions. Another gap in fibre characterization relates to the number of fibre samples that were studied. Given the relevance of reduced fibres in future material development, the analysis of a larger pool of samples is required for accurate quantitative evaluation of their mechanical properties. Further, more post-stretching conditions need to be tested on these fibres given the influence that this may have[7, 8] on fibre mechanics.

Various biophysical techniques like Raman spectroscopy,[425, 426] Fourier-transform infrared (FTIR),[427, 428] X-ray diffraction (XRD)[429, 430] and solid-state NMR[431, 119] has been extensively employed to understand the secondary structural features of spider silk fibres. Second harmonic generation (SHG) is a relatively new technique gaining popularity to understand fibril organization, orientational anisotropy, etc. in various fibrillogenic proteins.[432-434] Scope of these technique should be considered and tested on WT and reduced mutant fibres to rationalize the improved mechanics. Evaluation of glass transition temperature (Tg) could provide insights on H-bond density and hence the structural organization within these silk fibres. on their mechanical performance is also evaluated through their glass transition temperature.[435, 436]

The remarkable structural integrity of the stapled microfibrils in EtOH (>2 years) is indicative of a prolonged stability and shelf-life for these fibrillar structures. Since this is one of the intermediate states and macroscopic structures in the hierarchical association process, the scope of these structures 1) for understanding the fibrillogenesis mechanism and 2) for prospective silk-based material development should be considered.

7.2.2.2.3. Probe for interchain disulphides in reduced mutant fibres

In native spidroins, Cys residue in the conserved CTD domain (absent in the W unit) is reported to guide unidirectional arrangement of protein chains through disulphide-mediated dimerization

followed by controlled micellar arrangement, a significant step in fibre formation process.[391] Throughout the course of my thesis work, Cys residues retained in the reduced form (in the presence of DTT) yielded smaller self-assemblies and superior-quality fibres in comparison to the WT protein. One hypothesis for this finding is that the presence of DTT might have modified particle self-assembly and, later, the microfibrillar arrangement during fibrillogenesis, resulting in improved packing within the silk fibres formed Free Cys/DTT protecting Cys is suspected to perform a similar role through interchain association. Even though -SH side chain does not promote H-bonding, R-S-S- observed at low pH conditions is reported to be highly reactive. Suspecting a scenario similar to that of scleroproteins of silkworm silk, wool, keratin and vulcanized rubber,[392-394] the reduced mutant fibres should be probed for potential interchain disulphide bonding, hence fibre compactness and density to elucidate the structure-mechanical property relationship.

7.2.2.2.4. Studies inspired from similar fibrillogenic systems

Maleic acid imides (maleimides) are popular reagents employed in the covalent derivatization and fluorescent labelling of proteins.[437-439] Owing to their high site-specificity, these reagents could efficiently react with solvent-accessible Cys residues in the reduced state of the mutant protein at pH 6.5 - 7.5.[440] Given that disulphide bonds do not react with maleimides,[438] this technique could be employed to track local conformational changes and probe intra/intermolecular interactions around Cys residues using fluorescence assays and, hence, gain a better overview of the of mutant protein. systems, could be adopted to explore the unknown realms of spider silk fibrillogenesis.

Considering fibre formation, I feel that both alternative chemical methods and biophysical methodologies would be highly beneficial to explore. To enhance fibre formation itself, one

approach that I feel would be worth investigating would be to employ a reagent such as heparin, that is widely used to trigger *in-vitro* aggregation in proteins like *Tau, amyloid-β* etc. while monitoring the aggregation propensity, conformational stacking and kinetics of aggregation, with recent reports citing the use of nanopore sensors in revealing the mechanism of heparin-induced aggregation in the ms timescale.[441] Recent advancements in optical tweezers is promising towards their application for the detection and characterization of the kinetics, mechanics and dynamics of transient protein-protein interactions, protein (un)folding, particle self-assembly, higher-order protein assemblies etc. at a variety of size scales.[442, 443] This would require detailed thought about whether the Cys mutations at the present locations vs. towards the termini would be most beneficial. Given the amyloidogenic nature of spider silk and its prospects as a model system for testing β-sheet formation hypotheses in other fibrillogenic systems,[444] biophysical methods with their exciting demonstration of value for other protein.

7.2.2.3. Bioengineering new mutants

Owing to the sequence-structure-function correlation, it is important to understand the macroscopic properties and the microscopic structure for prospective bioengineering of novel materials.[445, 446] Hence, role of side chain mutation (change in hydrophobicity, H-bonding propensity etc.) and/or DTT protection in the speculated conformational changes leading to such quality improvement should also be considered. This can be tested using single Cys and/or non-Cys mutants in future. Pro is known to be a helix-breaker residue. Effect of such a mutation in H5 could be insightful towards understanding the role of H5 in fibre formation.

Due to the presence of two Cys residues in close vicinity, reagents like DTT (or TCEP) were required to impede disulphide bond formation while characterizing the W_1 and W_2 mutant proteins

in the reduced state. Another interesting project could evolve by replacing the Cys mutation in H5 with residues like Pro, short peptide sequences (3-4 residues), or molecular spacers (chemicals like polyols) that could promote intermolecular interactions without the risk of unwanted covalent chemistry promoting protein aggregation. Computational studies could be useful to understand the local steric effects with such side chain mutations and hence in the design of such mutants.

7.2.2.4. Rational architecture of silk-based hybrid materials

My results from the reduced W_2M protein provide a promising approach for the development of silk-based hybrid functional materials. Taking advantage of the hierarchical protein assembly involving intra- and intermolecular interactions,[447] [448] this could be realized through a variety of strategies. Some examples that I envision include the addition of 1) suitable side-chain modifiers for functionalizing or derivatizing side chains; 2) bio-conjugating reagents to promote polymerization and cross-linking and, hence, improve fibre mechanical performance; 3) molecular spacers to create cavities or vacant spaces within microfibrils/fibres; 4) proteins/oligomers exhibiting natural fibre-forming tendency (e.g., silkworm silk, keratin) to create hybrid fibres with novel properties; or, 5) other (bio)chemical reagents like nanoparticles, nanosheets, nanowires etc. at various stages of fibre formation (e.g. to dope solution, while fibre spinning, treatments on wet spun fibres (coating, spraying, etc.) would enable bioimaging, biomolecular detection, sensors, microelectronics etc. In any of these cases, hybrid functional materials would be obtained, allowing for the development of higher-order structures like fibres with tunable mechanical, physical, optical, surface plasmonic, thermal and/or chemical properties, broadening the span of applications for recombinant aciniform silk (***Fig. 6.2***.).

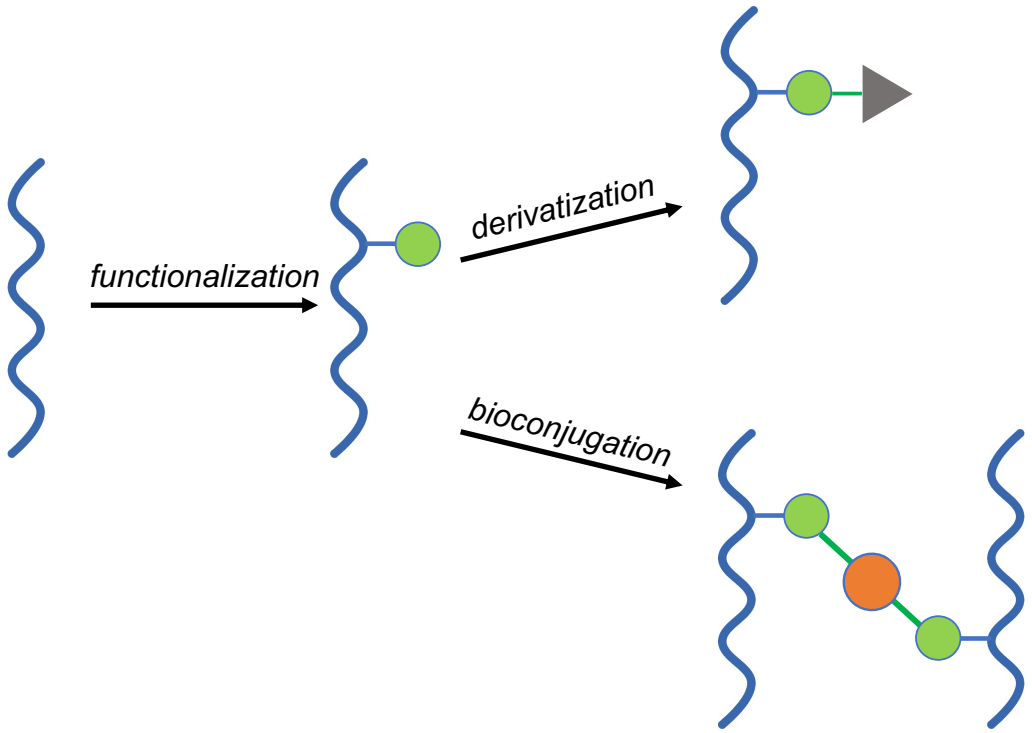


Fig. 6.2. *Proposed functionalization, derivatization and bioconjugation of reduced mutant protein. The blue wavy lines represent protein chain, green circles represent prospective side chain modifiers enabling derivatization and/or functionalization, and the orange circle represents a second order derivatization and/or functionalization achieved through bioconjugation through the sidechain modifier.*

6.3. SUMMARY OF THESIS

During the course of the research detailed in this thesis, I worked with 2 different classes of protein: a) the mammalian venom-derived paralytic protein soricidin (***Chapter 2***) and b) Ser to Cys mutants of the aciniform spider silk repetitive unit. Both classes of protein were recombinantly produced using *E. coli* as an expression system, with subsequent oxidative folding achieved for each. While the former protein inherently contains six Cys residues, two Cys residues were introduced within each W unit through Ser→ Cys mutations in the latter. The institution of intramolecular disulphide bonds within these disparate proteins, both of which interestingly adopt a helix-loop-helix fold,

led to clearly contrasting effects in their native functionality. Namely, while soricidin gained its paralytic activity from the formation of a specific topology of three disulphide bonds (***Chapter 2***), the mutant aciniform silk concatemer demonstrated a loss-of-function in response to the introduction of a single positionally-engineered disulphide bond within each W unit. (***Chapters 3-5***). This latter situation allowed me to introduce a chemical trigger for silk fibrillogenesis, whereupon reduction of the disulphide bonds restored functionality. These contrasting functional outcomes in terms of disulphide bond formation exemplify the fundamentals of protein biochemistry – the fact that biomolecules like proteins are rendered functional through their unique conformational states, with minute changes at the atomic level having the potential to amend both native structure and dynamics with either adverse or favourable functional consequences.

BIBLIOGRAPHY

1. Blamires, S. J.; Blackledge, T. A.; Tso, I.-M., Physicochemical property variation in spider silk: ecology, evolution, and synthetic production. *Annual review of entomology* **2017,** *62*, 443-460.

2. Mukhopadhyay, D.; Dasso, M., Modification in reverse: the SUMO proteases. *Trends in Biochemical Sciences* **2007,** *32* (6), 286-295.

3. Xu, L.; Rainey, J. K.; Meng, Q.; Liu, X.-Q., Recombinant minimalist spider wrapping silk proteins capable of native-like fiber formation. *PLoS One* **2012,** *7* (11), e50227.

4. Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M., Using Xplor–NIH for NMR molecular structure determination. *Progress in nuclear magnetic resonance spectroscopy* **2006,** *48* (1), 47-62.

5. Tremblay, M.-L.; Xu, L.; Lefèvre, T.; Sarker, M.; Orrell, K. E.; Leclerc, J.; Meng, Q.; Pézolet, M.; Auger, M.; Liu, X.-Q., Spider wrapping silk fibre architecture arising from its modular soluble protein precursor. *Scientific reports* **2015,** *5* (1), 1-15.

6. Heim, M.; Keerl, D.; Scheibel, T., Spider silk: from soluble protein to extraordinary fiber. *Angewandte Chemie International Edition* **2009,** *48* (20), 3584-3596.

7. Weatherbee-Martin, N.; Xu, L.; Hupe, A.; Kreplak, L.; Fudge, D. S.; Liu, X.-Q.; Rainey, J. K., Identification of wet-spinning and post-spin stretching methods amenable to recombinant spider aciniform silk. *Biomacromolecules* **2016,** *17* (8), 2737-2746.

8. Simmons, J. R.; Xu, L.; Rainey, J. K., Recombinant pyriform silk fiber mechanics are modulated by wet-spinning conditions. *ACS Biomaterials Science & Engineering* **2019,** *5* (10), 4985-4993.

9. Xu, L.; Tremblay, M.-L.; Meng, Q.; Liu, X.-Q.; Rainey, J. K., 1H, 13C and 15N NMR assignments of the aciniform spidroin (AcSp1) repetitive domain of *Argiope trifasciata* wrapping silk. *Biomolecular NMR Assignments* **2012,** *6* (2), 147-151.

10. Palmer III, A. G., NMR characterization of the dynamics of biomacromolecules. *Chemical reviews* **2004,** *104* (8), 3623-3640.

11. Augusti, K.; Benaim, M., Effect of essential oil of onion (allyl propyl disulphide) on blood glucose, free fatty acid and insulin levels of normal subjects. *Clinica chimica acta* **1975,** *60* (1), 121-123.

12. Alagon, A. C.; King, T. P., Activation of polysaccharides with 2-iminothiolane and its uses. *Biochemistry* **1980,** *19* (18), 4341-4345.

13. Adams, G. G.; Kök, S. M.; Imran, S.; Harding, S. E.; Ilyas, M.; Tatham, A. S., The interaction of dietary fibres with disulphide bonds (SS) and a potential strategy to reduce the

toxicity of the gluten proteins in coeliac disease. *Biotechnology and Genetic Engineering Reviews* **2012,** *28* (1), 115-130.

14. Sarkar, B.; Jayaraman, N., Glycoconjugations of biomolecules by chemical methods. *Front Chem* **2020,** *8*, 570185-570185.

15. Ribeiro Morais, G.; Falconer, R. A., Glycosyl disulfides: importance, synthesis and application to chemical and biological systems. *Organic & Biomolecular Chemistry* **2021,** *19* (1), 82-100.

16. Miseta, A.; Csutora, P., Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Molecular biology and evolution* **2000,** *17* (8), 1232-1239.

17. Jordan, I. K.; Kondrashov, F. A.; Adzhubei, I. A.; Wolf, Y. I.; Koonin, E. V.; Kondrashov, A. S.; Sunyaev, S., A universal trend of amino acid gain and loss in protein evolution. *Nature* **2005,** *433* (7026), 633-638.

18. Brooks, D. J.; Fresco, J. R., Increased frequency of cysteine, tyrosine, and phenylalanine residues since the last universal ancestor. *Molecular & Cellular Proteomics* **2002,** *1* (2), 125-131.

19. Broekaert, W. F.; Cammue, B. P.; De Bolle, M. F.; Thevissen, K.; De Samblanx, G. W.; Osborn, R. W.; Nielson, K., Antimicrobial peptides from plants. *Critical reviews in plant sciences* **1997,** *16* (3), 297-323.

20. Srivastava, S.; Dashora, K.; Ameta, K. L.; Singh, N. P.; El‐Enshasy, H. A.; Pagano, M. C.; Hesham, A. E. L.; Sharma, G. D.; Sharma, M.; Bhargava, A., Cysteine‐rich antimicrobial peptides from plants: The future of antimicrobial therapy. *Phytotherapy Research* **2021,** *35* (1), 256-277.

21. Hammami, R.; Ben Hamida, J.; Vergoten, G.; Fliss, I., PhytAMP: a database dedicated to antimicrobial plant peptides. *Nucleic acids research* **2009,** *37* (suppl_1), D963-D968.

22. Colgrave, M. L.; Craik, D. J., Thermal, chemical, and enzymatic stability of the cyclotide kalata B1: the importance of the cyclic cystine knot. *Biochemistry* **2004,** *43* (20), 5965-5975.

23. Østergaard, H.; Tachibana, C.; Winther, J. R., Monitoring disulfide bond formation in the eukaryotic cytosol. *The Journal of cell biology* **2004,** *166* (3), 337-345.

24. Chakravarthi, S.; Jessop, C. E.; Bulleid, N. J., The role of glutathione in disulphide bond formation and endoplasmic‐reticulum‐generated oxidative stress. *EMBO reports* **2006,** *7* (3), 271-275.

25. Vivancos, P. D.; Dong, Y.; Ziegler, K.; Markovic, J.; Pallardó, F. V.; Pellny, T. K.; Verrier, P. J.; Foyer, C. H., Recruitment of glutathione into the nucleus during cell proliferation adjusts whole‐cell redox homeostasis in *Arabidopsis thaliana* and lowers the oxidative defence shield. *The Plant Journal* **2010,** *64* (5), 825-838.

26. Cumming, R. C.; Andon, N. L.; Haynes, P. A.; Park, M.; Fischer, W. H.; Schubert, D., Protein disulfide bond formation in the cytoplasm during oxidative stress. *Journal of biological chemistry* **2004,** *279* (21), 21749-21758.

27. Tu, B. P.; Weissman, J. S., Oxidative protein folding in eukaryotes mechanisms and consequences. *Journal of Cell Biology* **2004,** *164* (3), 341-346.

28. Nishinaka, Y.; Masutani, H.; Nakamura, H.; Yodoi, J., Regulatory roles of thioredoxin in oxidative stress-induced cellular responses. *Redox Report* **2001,** *6* (5), 289-295.

29. Bains, J. S.; Shaw, C. A., Neurodegenerative disorders in humans: the role of glutathione in oxidative stress-mediated neuronal death. *Brain research reviews* **1997,** *25* (3), 335-358.

30. Wedemeyer, W. J.; Welker, E.; Narayan, M.; Scheraga, H. A., Disulfide bonds and protein folding. *Biochemistry* **2000,** *39* (15), 4207-4216.

31. Narayan, M.; Welker, E.; Wedemeyer, W. J.; Scheraga, H. A., Oxidative folding of proteins. *Accounts of chemical research* **2000,** *33* (11), 805-812.

32. Arolas, J. L.; Aviles, F. X.; Chang, J.-Y.; Ventura, S., Folding of small disulfide-rich proteins: clarifying the puzzle. *Trends in Biochemical Sciences* **2006,** *31* (5), 292-301.

33. Wei, C.; Tang, B.; Zhang, Y.; Yang, K., Oxidative refolding of recombinant prochymosin. *Biochem J* **1999,** *340 (Pt 1)*, 345-51.

34. Arai, K.; Kumakura, F.; Iwaoka, M., Characterization of kinetic and thermodynamic phases in the prefolding process of bovine pancreatic ribonuclease A coupled with fast SS formation and SS reshuffling. *Biochemistry* **2010,** *49* (49), 10535-10542.

35. Creighton, T. E.; Zapun, A.; Darby, N. J., Mechanisms and catalysts of disulphide bond formation in proteins. *Trends in biotechnology* **1995,** *13* (1), 18-23.

36. Zhang, Y. Y.; Yang, K. Y., [Oxidative refolding of proteins]. *Sheng Wu Gong Cheng Xue Bao* **2003,** *19* (1), 1-8.

37. Madar, D. J.; Patel, A. S.; Lees, W. J., Comparison of the oxidative folding of lysozyme at a high protein concentration using aromatic thiols versus glutathione. *Journal of biotechnology* **2009,** *142* (3-4), 214-219.

38. Beld, J.; Woycechowsky, K. J.; Hilvert, D., Diselenides as universal oxidative folding catalysts of diverse proteins. *Journal of biotechnology* **2010,** *150* (4), 481-489.

39. Aboye, T. L.; Clark, R. J.; Burman, R.; Roig, M. B.; Craik, D. J.; Göransson, U., Interlocking disulfides in circular proteins: toward efficient oxidative folding of cyclotides. *Antioxidants & redox signaling* **2011,** *14* (1), 77-86.

40. Tienson, H. L.; Dabir, D. V.; Neal, S. E.; Loo, R.; Hasson, S. A.; Boontheung, P.; Kim, S.-K.; Loo, J. A.; Koehler, C. M., Reconstitution of the mia40-erv1 oxidative folding pathway for the small tim proteins. *Molecular biology of the cell* **2009,** *20* (15), 3481-3490.

41. Lees, W. J., Oxidative protein folding with small molecules. In *Folding of Disulfide Proteins*, Springer: 2011; pp 109-132.

42. Tanford, C.; Kawahara, K.; Lapanje, S., Proteins in 6-M guanidine hydrochloride. Demonstration of random coil behavior. *J Biol Chem* **1966,** *241* (8), 1921-3.

43. Anfinsen, C.; Scheraga, H., Experimental and theoretical aspects of protein folding. *Advances in protein chemistry* **1975,** *29*, 205-300.

44. Jackson, S. E., How do small single-domain proteins fold? *Fold Des* **1998,** *3* (4), R81-91.

45. Feige, M. J.; Braakman, I.; Hendershot, L. M., CHAPTER 1.1 Disulfide bonds in protein folding and stability. in *oxidative folding of proteins: basic principles, cellular regulation and engineering*, The Royal Society of Chemistry: 2018; pp 1-33.

46. Thornton, J., Disulphide bridges in globular proteins. *Journal of molecular biology* **1981,** *151* (2), 261-287.

47. Richardson, J. S., The anatomy and taxonomy of protein structure. *Advances in protein chemistry* **1981,** *34*, 167-339.

48. Ben-Naim, A., Pitfalls in Anfinsen's thermodynamic hypothesis. *Chemical Physics Letters* **2011,** *511* (1-3), 126-128.

49. Hirata, F.; Sugita, M.; Yoshida, M.; Akasaka, K., Perspective: Structural fluctuation of protein and Anfinsen's thermodynamic hypothesis. *The Journal of chemical physics* **2018,** *148* (2), 020901.

50. Vaz, D. C., Enthalpic and entropic contributions mediate the role of disulfide bonds on the conformational stability of Interleukin-4. *Protein Science* **2006,** *15* (1), 33-44.

51. Bryliński, M.; Prymula, K.; Jurkowski, W.; Kochańczyk, M.; Stawowczyk, E.; Konieczny, L.; Roterman, I., Prediction of functional sites based on the fuzzy oil drop model. *PLoS computational biology* **2007,** *3* (5), e94.

52. Banach, M.; Konieczny, L.; Roterman, I., The fuzzy oil drop model, based on hydrophobicity density distribution, generalizes the influence of water environment on protein structure and function. *Journal of theoretical biology* **2014,** *359*, 6-17.

53. Banach, M.; Konieczny, L.; Roterman, I., Fuzzy oil drop model application—from globular proteins to amyloids. In *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes*, Springer: 2019; pp 639-658.

54. Baker, D.; Agard, D. A., Kinetics versus thermodynamics in protein folding. *Biochemistry* **1994,** *33* (24), 7505-7509.

55. Baldwin, R. L., Energetics of protein folding. *J Mol Biol* **2007,** *371* (2), 283-301.

56. Fersht, A. R., From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat Rev Mol Cell Biol* **2008,** *9* (8), 650-4.

57. Wouters, M.; George, R.; Haworth, N., "Forbidden" disulfides: their role as redox switches. *Current protein and peptide science* **2007,** *8* (5), 484-495.

58. Disulfides as Redox Switches: From molecular mechanisms to functional significance. *Antioxidants & Redox Signaling* **2010,** *12* (1), 53-91.

59. Dishman, A. F.; Volkman, B. F., Unfolding the mysteries of protein metamorphosis. *ACS Chemical Biology* **2018,** *13* (6), 1438-1446.

60. Murzin, A. G., Metamorphic proteins. *Science* **2008**.

61. Creighton, T. E., Disulphide bonds and protein stability. *Bioessays* **1988,** *8* (2), 57-63.

62. Betz, S. F., Disulfide bonds and the stability of globular proteins. *Protein Science* **1993,** *2* (10), 1551-1558.

63. Schwartz, J.-L.; Juteau, M.; Grochulski, P.; Cygler, M.; Préfontaine, G.; Brousseau, R.; Masson, L., Restriction of intramolecular movements within the Cry1Aa toxin molecule of Bacillus thuringiensis through disulfide bond engineering. *FEBS Letters* **1997,** *410* (2), 397-402.

64. Tokuriki, N.; Tawfik, D. S., Stability effects of mutations and protein evolvability. *Current opinion in structural biology* **2009,** *19* (5), 596-604.

65. Dombkowski, A. A.; Sultana, K. Z.; Craig, D. B., Protein disulfide engineering. *FEBS letters* **2014,** *588* (2), 206-212.

66. Wu, S. C.; Ng, K. K. S.; Wong, S. L., Engineering monomeric streptavidin and its ligands with infinite affinity in binding but reversibility in interaction. *Proteins: Structure, Function, and Bioinformatics* **2009,** *77* (2), 404-412.

67. Shen, Y.; Joachimiak, A.; Rich Rosner, M.; Tang, W.-J., Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. *Nature* **2006,** *443* (7113), 870-874.

68. Johnson, K. A.; Bhushan, S.; Ståhl, A.; Hallberg, B. M.; Frohn, A.; Glaser, E.; Eneqvist, T., The closed structure of presequence protease PreP forms a unique 10 000 Å3 chamber for proteolysis. *The EMBO Journal* **2006,** *25* (9), 1977-1986.

69. Seeger, M. A.; Von Ballmoos, C.; Eicher, T.; Brandstätter, L.; Verrey, F.; Diederichs, K.; Pos, K. M., Engineered disulfide bonds support the functional rotation mechanism of multidrug efflux pump AcrB. *Nature Structural & Molecular Biology* **2008,** *15* (2), 199-205.

70. Glynn, S. E.; Nager, A. R.; Baker, T. A.; Sauer, R. T., Dynamic and static components power unfolding in topologically closed rings of a AAA^+ proteolytic machine. *Nature Structural & Molecular Biology* **2012,** *19* (6), 616-622.

71. Smallwood, T. B.; Clark, R. J., Advances in venom peptide drug discovery: where are we at and where are we heading? *Expert Opinion on Drug Discovery* **2021**, 1-11.

72. Härd, T., Protein engineering to stabilize soluble amyloid β-protein aggregates for structural and functional studies. *FEBS Journal* **2011,** *278* (20), 3884-3892.

73. Sandberg, A.; Luheshi, L. M.; Söllvander, S.; Pereira de Barros, T.; Macao, B.; Knowles, T. P. J.; Biverstål, H.; Lendel, C.; Ekholm-Petterson, F.; Dubnovitsky, A.; Lannfelt, L.; Dobson, C. M.; Härd, T., Stabilization of neurotoxic Alzheimer amyloid-β oligomers by protein engineering. *Proceedings of the National Academy of Sciences* **2010,** *107* (35), 15595-15600.

74. Müller-Schiffmann, A.; Andreyeva, A.; Horn, A. H. C.; Gottmann, K.; Korth, C.; Sticht, H., Molecular engineering of a secreted, highly homogeneous, and neurotoxic Aβ dimer. *ACS Chemical Neuroscience* **2011,** *2* (5), 242-248.

75. Yamaguchi, T.; Yagi, H.; Goto, Y.; Matsuzaki, K.; Hoshino, M., A Disulfide-linked amyloid-β peptide dimer forms a protofibril-like oligomer through a distinct pathway from amyloid fibril formation. *Biochemistry* **2010,** *49* (33), 7100-7107.

76. Nazari, M.; Hosseinkhani, S., Design of disulfide bridge as an alternative mechanism for color shift in firefly luciferase and development of secreted luciferase. *Photochemical & Photobiological Sciences* **2011,** *10* (7), 1203.

77. Imani, M.; Hosseinkhani, S.; Ahmadian, S.; Nazari, M., Design and introduction of a disulfide bridge in firefly luciferase: increase of thermostability and decrease of pH sensitivity. *Photochemical & Photobiological Sciences* **2010,** *9* (8), 1167.

78. Nazari, M.; Hosseinkhani, S.; Hassani, L., Step-wise addition of disulfide bridge in firefly luciferase controls color shift through a flexible loop: a thermodynamic perspective. *Photochemical & Photobiological Sciences* **2013,** *12* (2), 298-308.

79. Ménez, A., Functional architectures of animal toxins: a clue to drug design? *Toxicon* **1998,** *36* (11), 1557-72.

80. Ueberheide, B. M.; Fenyö, D.; Alewood, P. F.; Chait, B. T., Rapid sensitive analysis of cysteine rich peptide venom components. *Proceedings of the National Academy of Sciences* **2009,** *106* (17), 6910-6915.

81. Shu, Q.; Lu, S.-Y.; Gu, X.-C.; Liang, S.-P., The structure of spider toxin huwentoxin-II with unique disulfide linkage: evidence for structural evolution. *Protein Sci* **2002,** *11* (2), 245-252.

82. Gao, B.; Harvey, Peta J.; Craik, David J.; Ronjat, M.; De Waard, M.; Zhu, S., Functional evolution of scorpion venom peptides with an inhibitor cystine knot fold. *Bioscience Reports* **2013,** *33* (3).

83. Daly, N. L.; Craik, D. J., Bioactive cystine knot proteins. *Current Opinion in Chemical Biology* **2011,** *15* (3), 362-368.

84. Walkinshaw, M. D.; Saenger, W.; Maelicke, A., Three-dimensional structure of the" long" neurotoxin from cobra venom. *Proceedings of the National Academy of Sciences* **1980,** *77* (5), 2400-2404.

85. Fraenkel-Conrat, H., Snake venom neurotoxins related to phospholipase A2. *Journal of Toxicology: Toxin Reviews* **1982,** *1* (2), 205-221.

86. Eiselé, J.-L.; Bertrand, S.; Galzi, J.-L.; Devillers-Thiéry, A.; Changeux, J.-P.; Bertrand, D., Chimaeric nicotinic–serotonergic receptor combines distinct ligand binding and channel specificities. *Nature* **1993,** *366* (6454), 479-483.

87. Ullah, A.; Masood, R., The sequence and three-dimensional structure characterization of snake venom phospholipases b. *Frontiers in molecular biosciences* **2020,** *7*, 175.

88. Servent, D.; Winckler-Dietrich, V.; Hu, H.-Y.; Kessler, P.; Drevet, P.; Bertrand, D.; Ménez, A., Only snake curaremimetic toxins with a fifth disulfide bond have high affinity for the neuronal α7 nicotinic receptor. *Journal of Biological Chemistry* **1997,** *272* (39), 24279-24286.

89. Wells, M. A., Evidence that the phospholipases A2 of *Crotalus adamanteus* venom are dimers. *Biochemistry* **1971,** *10* (22), 4074-4078.

90. Doley, R.; Kini, R., Protein complexes in snake venom. *Cellular and molecular life sciences* **2009,** *66* (17), 2851-2871.

91. Bénard-Valle, M.; Neri-Castro, E.; Yañez-Mendoza, M. F.; Lomonte, B.; Olvera, A.; Zamudio, F.; Restano-Cassulini, R.; Possani, L. D.; Jiménez-Ferrer, E.; Alagón, A., Functional, proteomic and transcriptomic characterization of the venom from *Micrurus browni browni*: Identification of the first lethal multimeric neurotoxin in coral snake venom. *Journal of Proteomics* **2020,** *225*, 103863.

92. Calvete, J. J.; Marcinkiewicz, C.; Sanz, L., Snake venomics of Bitis gabonica gabonica. Protein family composition, subunit organization of venom toxins, and characterization of dimeric disintegrins bitisgabonin-1 and bitisgabonin-2. *Journal of proteome research* **2007,** *6* (1), 326-336.

93. Balaji, R. A.; Ohtake, A.; Sato, K.; Gopalakrishnakone, P.; Kini, R. M.; Seow, K. T.; Bay, B.-H., λ-Conotoxins, a new family of conotoxins with unique disulfide pattern and protein folding. *Journal of Biological Chemistry* **2000,** *275* (50), 39516-39522.

94. Hill, J. M.; Oomen, C. J.; Miranda, L. P.; Bingham, J.-P.; Alewood, P. F.; Craik, D. J., Three-dimensional solution structure of α-conotoxin MII by NMR spectroscopy: effects of solution environment on helicity. *Biochemistry* **1998,** *37* (45), 15621-15630.

95. Möller, C.; Rahmankhah, S.; Lauer-Fields, J.; Bubis, J.; Fields, G. B.; Marí, F., A novel conotoxin framework with a helix− loop− helix (Cs α/α) fold. *Biochemistry* **2005,** *44* (49), 15986-15996.

96. Fry, B. G.; Roelants, K.; Champagne, D. E.; Scheib, H.; Tyndall, J. D.; King, G. F.; Nevalainen, T. J.; Norman, J. A.; Lewis, R. J.; Norton, R. S., The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annual review of genomics and human genetics* **2009,** *10*, 483-511.

97. Wang, X.; Umetsu, Y.; Gao, B.; Ohki, S.; Zhu, S., Mesomartoxin, a new Kv1. 2-selective scorpion toxin interacting with the channel selectivity filter. *Biochemical pharmacology* **2015,** *93* (2), 232-239.

98. Wang, F.; Li, H.; Liu, M.-n.; Song, H.; Han, H.-m.; Wang, Q.-l.; Yin, C.-c.; Zhou, Y.-c.; Qi, Z.; Shu, Y.-y., Structural and functional analysis of natrin, a venom protein that targets various ion channels. *Biochemical and biophysical research communications* **2006,** *351* (2), 443-448.

99. Zhu, L.; Gao, B.; Luo, L.; Zhu, S., Two dyad-free Shaker-type K^+ channel blockers from scorpion venom. *Toxicon* **2012,** *59* (3), 402-407.

100. Chen, T.; Walker, B.; Zhou, M.; Shaw, C., Molecular cloning of a novel putative potassium channel-blocking neurotoxin from the venom of the North African scorpion, *Androctonus amoreuxi*. *Peptides* **2005,** *26* (5), 731-736.

101. Peigneur, S.; Orts, D. J.; da Silva, A. R. P.; Oguiura, N.; Boni-Mitake, M.; de Oliveira, E. B.; Zaharenko, A. J.; de Freitas, J. C.; Tytgat, J., Crotamine pharmacology revisited: novel insights based on the inhibition of KV channels. *Molecular pharmacology* **2012,** *82* (1), 90-96.

102. Banijamali, S. E.; Amininasab, M.; Elmi, M. M., Characterization of a new member of kunitz-type protein family from the venom of Persian false-horned viper, *Pseudocerastes persicus*. *Archives of biochemistry and biophysics* **2019,** *662*, 1-6.

103. Mouhat, S.; Jouirou, B.; Mosbah, A.; De Waard, M.; Sabatier, J.-M., Diversity of folds in animal toxins acting on ion channels. *Biochemical Journal* **2004,** *378* (3), 717-726.

104. Nirthanan, S.; Pil, J.; Abdel-Mottaleb, Y.; Sugahara, Y.; Gopalakrishnakone, P.; Joseph, J. S.; Sato, K.; Tytgat, J., Assignment of voltage-gated potassium channel blocking activity to κ-KTx1. 3, a non-toxic homologue of κ-hefutoxin-1, from Heterometrus spinifer venom. *Biochemical pharmacology* **2005,** *69* (4), 669-678.

105. Dauplais, M.; Lecoq, A.; Song, J.; Cotton, J.; Jamin, N.; Gilquin, B.; Roumestand, C.; Vita, C.; de Medeiros, C. C.; Rowan, E. G., On the convergent evolution of animal toxins: conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures. *Journal of Biological Chemistry* **1997,** *272* (7), 4302-4309.

106. Dash, T. S.; Shafee, T.; Harvey, P. J.; Zhang, C.; Peigneur, S.; Deuis, J. R.; Vetter, I.; Tytgat, J.; Anderson, M.; Craik, D. J., A centipede toxin family defines a new ancient class of CSαβ defensins. *Available at SSRN 3188389* **2018**.

107. Klint, J. K.; Senff, S.; Saez, N. J.; Seshadri, R.; Lau, H. Y.; Bende, N. S.; Undheim, E. A. B.; Rash, L. D.; Mobli, M.; King, G. F., Production of recombinant disulfide-rich venom peptides for structural and functional analysis via expression in the periplasm of *E. coli*. *PLoS ONE* **2013,** *8* (5), e63865.

108. Meyers, M. A.; Chen, P.-Y.; Lin, A. Y.-M.; Seki, Y., Biological materials: Structure and mechanical properties. *Progress in Materials Science* **2008,** *53* (1), 1-206.

109. Eisoldt, L.; Smith, A.; Scheibel, T., Decoding the secrets of spider silk. *Materials Today* **2011,** *14* (3), 80-86.

110. Hayashi, C. Y., Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family. *Molecular Biology and Evolution* **2004,** *21* (10), 1950-1959.

111. Tian, M.; Lewis, R. V., Molecular characterization and evolutionary study of spider tubuliform (eggcase) silk protein. *Biochemistry* **2005,** *44* (22), 8006-8012.

112. Zhao, A.-C.; Zhao, T.-F.; Nakagaki, K.; Zhang, Y.-S.; SiMa, Y.-H.; Miao, Y.-G.; Shiomi, K.; Kajiura, Z.; Nagata, Y.; Takadera, M., Novel molecular and mechanical properties of egg case silk from wasp spider, Argiope bruennichi. *Biochemistry* **2006,** *45* (10), 3348-3356.

113. Huang, X.; Liu, G.; Wang, X., New secrets of spider silk: Exceptionally high thermal conductivity and its abnormal change under stretching. *Advanced Materials* **2012,** *24* (11), 1482-1486.

114. Parry, D. A. D.; Squire, J. M., Fibrous proteins: new structural and functional aspects revealed. In *Advances in Protein Chemistry*, Academic Press: 2005; Vol. 70, pp 1-10.

115. Heslot, H., Artificial fibrous proteins: a review. *Biochimie* **1998,** *80* (1), 19-31.

116. Savage, K. N. The role of elastomeric network chains in the mechanics of spider silks. University of British Columbia, 2006.

117. Zhu, B.; Li, W.; Lewis, R. V.; Segre, C. U.; Wang, R., E-spun composite fibers of collagen and dragline silk protein: fiber mechanics, biocompatibility, and application in stem cell differentiation. *Biomacromolecules* **2015,** *16* (1), 202-213.

118. Blackledge, T. A.; Boutry, C.; Wong, S.-C.; Baji, A.; Dhinojwala, A.; Sahni, V.; Agnarsson, I., How super is supercontraction? Persistent versus cyclic responses to humidity in spider dragline silk. *Journal of Experimental Biology* **2009,** *212* (13), 1981-1989.

119. Creager, M. S.; Jenkins, J. E.; Thagard-Yeaman, L. A.; Brooks, A. E.; Jones, J. A.; Lewis, R. V.; Holland, G. P.; Yarger, J. L., Solid-state NMR comparison of various spiders' dragline silk fiber. *Biomacromolecules* **2010,** *11* (8), 2039-2043.

120. Wang, S.-P.; Guo, T.-Q.; Guo, X.-Y.; Huang, J.-T.; Lu, C.-D., Structural analysis of fibroin heavy chain signal peptide of silkworm Bombyx mori. *Acta biochimica et biophysica Sinica* **2006,** *38* (7), 507-513.

121. Nova, A.; Keten, S.; Pugno, N. M.; Redaelli, A.; Buehler, M. J., Molecular and nanostructural mechanisms of deformation, strength and toughness of spider silk fibrils. *Nano letters* **2010,** *10* (7), 2626-2634.

122. Lefèvre, T.; Boudreault, S.; Cloutier, C.; Pézolet, M., Diversity of molecular transformations involved in the formation of spider silks. *Journal of molecular biology* **2011,** *405* (1), 238-253.

123. Simmons, A. H.; Michal, C. A.; Jelinski, L. W., Molecular orientation and two-component nature of the crystalline fraction of spider dragline silk. *Science* **1996,** *271* (5245), 84-87.

124. Gotte, L.; Giro, M.; Volpin, D.; Horne, R., The ultrastructural organization of elastin. *Journal of ultrastructure research* **1974,** *46* (1), 23-33.

125. Becker, N.; Oroudjev, E.; Mutz, S.; Cleveland, J. P.; Hansma, P. K.; Hayashi, C. Y.; Makarov, D. E.; Hansma, H. G., Molecular nanosprings in spider capture-silk threads. *Nature materials* **2003,** *2* (4), 278-283.

126. Kreplak, L.; Doucet, J.; Dumas, P.; Briki, F., New aspects of the α-helix to β-sheet transition in stretched hard α-keratin fibers. *Biophysical journal* **2004,** *87* (1), 640-647.

127. Fraser, R.; MacRae, T. P.; Sparrow, L. G.; Parry, D., Disulphide bonding in α-keratin. *International Journal of Biological Macromolecules* **1988,** *10* (2), 106-112.

128. Cruz, C. F.; Martins, M.; Egipto, J.; Osório, H.; Ribeiro, A.; Cavaco-Paulo, A., Changing the shape of hair with keratin peptides. *RSC advances* **2017,** *7* (81), 51581-51592.

129. Sweetman, B.; Maclaren, J., The reduction of wool keratin by tertiary phosphines. *Australian Journal of Chemistry* **1966,** *19* (12), 2347-2354.

130. Hagn, F.; Eisoldt, L.; Hardy, J. G.; Vendrely, C.; Coles, M.; Scheibel, T.; Kessler, H., A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature* **2010,** *465* (7295), 239-242.

131. Wang, S.; Huang, W.; Yang, D., Structure and function of C-terminal domain of aciniform spidroin. *Biomacromolecules* **2014,** *15* (2), 468-477.

132. Rousseau, M.-E.; Lefèvre, T.; Pézolet, M., Conformation and orientation of proteins in various types of silk fibers produced by *Nephila clavipes* spiders. *Biomacromolecules* **2009,** *10* (10), 2945-2953.

133. Blackledge, T. A.; Hayashi, C. Y., Silken toolkits: biomechanics of silk fibers spun by the orb web spider *Argiope argentata* (Fabricius 1775). *Journal of experimental biology* **2006,** *209* (13), 2452-2461.

134. Hinman, M. B.; Jones, J. A.; Lewis, R. V., Synthetic spider silk: a modular fiber. *Trends in biotechnology* **2000,** *18* (9), 374-379.

135. Xue, Y.; Lofland, S.; Hu, X., Thermal conductivity of protein-based materials: A review. *Polymers* **2019,** *11* (3), 456.

136. Xu, S.; Xu, Z.; Starrett, J.; Hayashi, C.; Wang, X., Cross-plane thermal transport in micrometer-thick spider silk films. *Polymer* **2014,** *55* (7), 1845-1853.

137. Belbéoch, C.; Lejeune, J.; Vroman, P.; Salaün, F., Silkworm and spider silk electrospinning: a review. *Environmental Chemistry Letters* **2021**, 1-27.

138. Zhang, L.; Chen, T.; Ban, H.; Liu, L., Hydrogen bonding-assisted thermal conduction in β-sheet crystals of spider silk protein. *Nanoscale* **2014,** *6* (14), 7786-7791.

139. Nguyen, T. P.; Nguyen, Q. V.; Nguyen, V.-H.; Le, T.-H.; Huynh, V. Q. N.; Vo, D.-V. N.; Trinh, Q. T.; Kim, S. Y.; Le, Q. V., Silk fibroin-based biomaterials for biomedical applications: A review. *Polymers* **2019,** *11* (12), 1933.

140. Kim, J. H.; Park, C. H.; Lee, O. J.; Lee, J. M.; Kim, J. W.; Park, Y. H.; Ki, C. S., Preparation and in vivo degradation of controlled biodegradability of electrospun silk fibroin nanofiber mats. *Journal of biomedical materials research Part A* **2012,** *100* (12), 3287-3295.

141. Meinel, L.; Hofmann, S.; Karageorgiou, V.; Kirker-Head, C.; McCool, J.; Gronowicz, G.; Zichner, L.; Langer, R.; Vunjak-Novakovic, G.; Kaplan, D. L., The inflammatory responses to silk films *in vitro* and *in vivo*. *Biomaterials* **2005,** *26* (2), 147-155.

142. Maitz, M. F.; Sperling, C.; Wongpinyochit, T.; Herklotz, M.; Werner, C.; Seib, F. P., Biocompatibility assessment of silk nanoparticles: hemocompatibility and internalization by human blood cells. *Nanomedicine: Nanotechnology, Biology and Medicine* **2017,** *13* (8), 2633-2642.

143. Koh, L.-D.; Cheng, Y.; Teng, C.-P.; Khin, Y.-W.; Loh, X.-J.; Tee, S.-Y.; Low, M.; Ye, E.; Yu, H.-D.; Zhang, Y.-W., Structures, mechanical properties and applications of silk fibroin materials. *Progress in Polymer Science* **2015,** *46*, 86-110.

144. Altman, G. H.; Diaz, F.; Jakuba, C.; Calabro, T.; Horan, R. L.; Chen, J.; Lu, H.; Richmond, J.; Kaplan, D. L., Silk-based biomaterials. *Biomaterials* **2003,** *24* (3), 401-416.

145. Vendrely, C.; Scheibel, T., Biotechnological production of spider‐silk proteins enables new applications. *Macromolecular bioscience* **2007,** *7* (4), 401-409.

146. Huby, N.; Vié, V.; Renault, A.; Beaufils, S.; Lefevre, T.; Paquet-Mercier, F.; Pézolet, M.; Bêche, B., Native spider silk as a biological optical fiber. *Applied Physics Letters* **2013,** *102* (12), 123702.

147. Monks, J. N.; Yan, B.; Hawkins, N.; Vollrath, F.; Wang, Z., Spider silk: mother nature's bio-superlens. *Nano letters* **2016,** *16* (9), 5842-5845.

148. Rising, A.; Widhe, M.; Johansson, J.; Hedhammar, M., Spider silk proteins: recent advances in recombinant production, structure–function relationships and biomedical applications. *Cellular and Molecular Life Sciences* **2011,** *68* (2), 169-184.

149. Gellynck, K.; Verdonk, P. C.; Van Nimmen, E.; Almqvist, K. F.; Gheysens, T.; Schoukens, G.; Van Langenhove, L.; Kiekens, P.; Mertens, J.; Verbruggen, G., Silkworm and spider silk scaffolds for chondrocyte support. *Journal of Materials Science: Materials in Medicine* **2008,** *19* (11), 3399-3409.

150. Horan, R. L.; Antle, K.; Collette, A. L.; Wang, Y.; Huang, J.; Moreau, J. E.; Volloch, V.; Kaplan, D. L.; Altman, G. H., *In vitro* degradation of silk fibroin. *Biomaterials* **2005,** *26* (17), 3385-3393.

151. Borkner, C. B.; Elsner, M. B.; Scheibel, T., Coatings and films made of silk proteins. *ACS Applied Materials & Interfaces* **2014,** *6* (18), 15611-15625.

152. Fu, C.; Shao, Z.; Fritz, V., Animal silks: their structures, properties and artificial production. *Chemical Communications* **2009,** (43), 6515-6529.

153. Vollrath, F., Biology of spider silk. *Int J Biol Macromol* **1999,** *24* (2-3), 81-8.

154. Madsen, B.; Shao, Z. Z.; Vollrath, F., Variability in the mechanical properties of spider silks on three levels: interspecific, intraspecific and intraindividual. *Int J Biol Macromol* **1999,** *24* (2-3), 301-6.

155. Lazaris, A., Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* **2002,** *295* (5554), 472-476.

156. Xia, X. X.; Qian, Z. G.; Ki, C. S.; Park, Y. H.; Kaplan, D. L.; Lee, S. Y., Native-sized recombinant spider silk protein produced in metabolically engineered *Escherichia coli* results in a strong fiber. *Proceedings of the National Academy of Sciences* **2010,** *107* (32), 14059-14063.

157. Adrianos, S. L.; Teulé, F.; Hinman, M. B.; Jones, J. A.; Weber, W. S.; Yarger, J. L.; Lewis, R. V., *Nephila clavipes* flagelliform silk-like GGX motifs contribute to extensibility and spacer motifs contribute to strength in synthetic spider silk fibers. *Biomacromolecules* **2013,** *14* (6), 1751-1760.

158. Heidebrecht, A.; Eisoldt, L.; Diehl, J.; Schmidt, A.; Geffers, M.; Lang, G.; Scheibel, T., biomimetic fibers made of recombinant spidroins with the same toughness as natural spider silk. *Advanced Materials* **2015,** *27* (13), 2189-2194.

159. Panavas, T.; Sanders, C.; Butt, T. R., SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems. Humana Press: 2009; pp 303-317.

160. Churchfield, S., *The natural history of shrews*. Cornell University Press: 1990.

161. Dufton, M. J., Venomous mammals. *Pharmacology & therapeutics* **1992,** *53* (2), 199-215.

162. Bowen, C. V.; DeBay, D.; Ewart, H. S.; Gallant, P.; Gormley, S.; Ilenchuk, T. T.; Iqbal, U.; Lutes, T.; Martina, M.; Mealing, G., *In vivo* detection of human TRPV6-rich tumors with anti-cancer peptides derived from soricidin. *PloS one* **2013,** *8* (3), e58866.

163. Tomasi, T. E., Function of venom in the short-tailed shrew, *Blarina brevicauda*. *Journal of mammalogy* **1978,** *59* (4), 852-854.

164. Kita, M.; Nakamura, Y.; Okumura, Y.; Ohdachi, S. D.; Oba, Y.; Yoshikuni, M.; Kido, H.; Uemura, D., Blarina toxin, a mammalian lethal venom from the short-tailed shrew *Blarina brevicauda*: Isolation and characterization. *Proceedings of the National Academy of Sciences* **2004,** *101* (20), 7542-7547.

165. Pearson, O. P., A toxic substance from the salivary glands of a mammal (short-tailed shrew). *Venoms* **1956,** *44*, 55-8.

166. Hamilton, W., The food of the Soricidae. *Journal of Mammalogy* **1930,** *11* (1), 26-39.

167. Martin, I. G., Venom of the short-tailed shrew (*Blarina brevicauda*) as an insect immobilizing agent. *Journal of Mammalogy* **1981,** *62* (1), 189-192.

168. POURNELLE, G. H., Classification, biology, and description of the venom apparatus of insectivores of the genera *Solenodon, Neomys, and Blarina*. In *Venomous animals and their venoms*, Elsevier: 1968; pp 31-42.

169. Davey, M.; Dugourd, D.; Lutes, T.; Rice, C.; Gormley, S.; Pierre, S. S.; Stewart, J. M., Abstract LB-051: Cancer cell binding, internalization and intracellular localization of a novel therapeutic peptide SOR-C13 targeting the TRPV6 oncochannel. AACR: 2016.

170. Lutes, T.; Dugourd, D.; Davey, M.; Rice, C.; St-Pierre, S.; Stewart, J. M., Abstract LB-128: High prevalence of elevated TRPV6 mRNA in pancreatic ductal adenocarcinoma. AACR: 2016.

171. Rice, C.; Lutes, T.; Davey, M.; Lloyd, V.; MacCormack, T. J.; Stewart, J. M.; Dugourd, D., The central role of NFAT signalling in the mechanism of action of the TRPV6 oncochannel inhibitor and clinical candidate SOR-C13. AACR: 2018.

172. Novakovic, S. D.; Eglen, R. M.; Hunter, J. C., Regulation of Na^+ channel distribution in the nervous system. *Trends in neurosciences* **2001,** *24* (8), 473-478.

173. Atanassoff, P. G.; Hartmannsgruber, M. W.; Thrasher, J.; Wermeling, D.; Longton, W.; Gaeta, R.; Singh, T.; Mayo, M.; McGuire, D.; Luther, R. R., Ziconotide, a new N-type calcium

channel blocker, administered intrathecally for acute postoperative pain. *Regional Anesthesia & Pain Medicine* **2000,** *25* (3), 274-278.

174. Franklin, J. B.; Rajesh, R. P., A sleep-inducing peptide from the venom of the Indian cone snail *Conus araneosus*. *Toxicon* **2015,** *103*, 39-47.

175. Livett, B. G.; Gayler, K. R.; Khalil, Z., Drugs from the sea: conopeptides as potential therapeutics. *Current medicinal chemistry* **2004,** *11* (13), 1715-1723.

176. Aguilar, M. B.; Lezama-Monfil, L.; Maillo, M.; Pedraza-Lara, H.; López-Vera, E.; de la Cotera, E. P. H., A biologically active hydrophobic T-1-conotoxin from the venom of *Conus spurius*. *Peptides* **2006,** *27* (3), 500-505.

177. Gowd, K. H.; Sabareesh, V.; Sudarslal, S.; Iengar, P.; Franklin, B.; Fernando, A.; Dewan, K.; Ramaswami, M.; Sarma, S. P.; Sikdar, S., Novel peptides of therapeutic promise from Indian Conidae. *Annals of the New York Academy of Sciences* **2005,** *1056*, 462-473.

178. Rigby, A. C.; Lucas-Meunier, E.; Kalume, D. E.; Czerwiec, E.; Hambe, B.; Dahlqvist, I.; Fossier, P.; Baux, G.; Roepstorff, P.; Baleja, J. D., A conotoxin from Conus textile with unusual posttranslational modifications reduces presynaptic Ca^{2+} influx. *Proceedings of the National Academy of Sciences* **1999,** *96* (10), 5758-5763.

179. Alonso, D.; Khalil, Z.; Satkunanthan, N.; Livett, B., Drugs from the sea: conotoxins as drug leads for neuropathic pain and other neurological conditions. *Mini Reviews in Medicinal Chemistry* **2003,** *3* (7), 785-787.

180. Lewis, R. V., Spider silk: ancient ideas for new biomaterials. *Chemical Reviews* **2006,** *106* (9), 3762-3774.

181. Hayashi, C. Y.; Lewis, R. V., Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *Journal of molecular biology* **1998,** *275* (5), 773-784.

182. Xu, L.; Tremblay, M.-L.; Orrell, K. E.; Leclerc, J.; Meng, Q.; Liu, X.-Q.; Rainey, J. K., Nanoparticle self-assembly by a highly stable recombinant spider wrapping silk protein subunit. *FEBS letters* **2013,** *587* (19), 3273-3280.

183. Vasanthavada, K.; Hu, X.; Falick, A. M.; La Mattina, C.; Moore, A. M. F.; Jones, P. R.; Yee, R.; Reza, R.; Tuton, T.; Vierra, C., Aciniform spidroin, a constituent of egg case sacs and wrapping silk fibers from the black widow spider *Latrodectus hesperus*. *Journal of Biological Chemistry* **2007,** *282* (48), 35088-35097.

184. Bass, R. B.; Butler, S. L.; Chervitz, S. A.; Gloor, S. L.; Falke, J. J., Use of site‐directed cysteine and disulfide chemistry to probe protein structure and dynamics: applications to soluble and transmembrane receptors of bacterial chemotaxis. Elsevier: 2007; pp 25-51.

185. McIntosh, L. P.; Dahlquist, F. W., Biosynthetic Incorporation of ^{15}N and ^{13}C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins. *Quarterly Reviews of Biophysics* **1990,** *23* (1), 1-38.

186. Marley, J.; Lu, M.; Bracken, C., A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR* **2001,** *20* (1), 71-75.

187. Gill, S. C.; von Hippel, P. H., Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **1989,** *182* (2), 319-26.

188. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989,** *246* (4926), 64-71.

189. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **2006,** *1* (6), 2876.

190. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism. *Biochim Biophys Acta* **2005,** *1751* (2), 119-39.

191. Kinney, D. R.; Chuang, I. S.; Maciel, G. E., Water and the silica surface as studied by variable-temperature high-resolution proton NMR. *Journal of the American Chemical Society* **1993,** *115* (15), 6786-6794.

192. Ardenkjær-Larsen, J. H.; Fridlund, B.; Gram, A.; Hansson, G.; Hansson, L.; Lerche, M. H.; Servin, R.; Thaning, M.; Golman, K., Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR. *Proceedings of the National Academy of Sciences* **2003,** *100* (18), 10158-10163.

193. Ohnishi, M.; Urry, D. W., Temperature dependence of amide proton chemical shifts: the secondary structures of gramicidin S and valinomycin. *Biochem Biophys Res Commun* **1969,** *36* (2), 194-202.

194. Cierpicki, T.; Otlewski, J., Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J Biomol NMR* **2001,** *21* (3), 249-61.

195. Tomlinson, J. H.; Williamson, M. P., Amide temperature coefficients in the protein G B1 domain. *J Biomol NMR* **2012,** *52* (1), 57-64.

196. Cordier, F.; Grzesiek, S., Temperature-dependence of protein hydrogen bond properties as studied by high-resolution NMR. *J Mol Biol* **2002,** *317* (5), 739-52.

197. Hong, J.; Jing, Q.; Yao, L., The protein amide 1HN chemical shift temperature coefficient reflects thermal expansion of the N–H···O=C hydrogen bond. *Journal of Biomolecular NMR* **2013,** *55* (1), 71-78.

198. Trainor, K.; Palumbo, J. A.; Mackenzie, D. W. S.; Meiering, E. M., Temperature dependence of NMR chemical shifts: Tracking and statistical analysis. *Protein Science* **2020,** *29* (1), 306-314.

199. Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A., Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969)* **1990,** *89* (3), 496-514.

200. Bax, A.; Ikura, M., 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. *Journal of biomolecular NMR* **1991,** *1* (1), 99-104.

201. Clubb, R. T.; Thanabal, V.; Wagner, G., A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue $^{1}H^{N}$, ^{15}N, and $^{13}C'$ chemical shifts in ^{15}N--^{13}C-labelled proteins. *Journal of Magnetic Resonance (1969)* **1992,** *97* (1), 213-217.

202. Vuister, G. W.; Bax, A., Quantitative J correlation: a new approach for measuring homonuclear three-bond J (HNH. alpha.) coupling constants in ^{15}N-enriched proteins. *Journal of the American Chemical Society* **1993,** *115* (17), 7772-7777.

203. Archer, S. J.; Ikura, M.; Torchia, D. A.; Bax, A., An alternative 3D NMR technique for correlating backbone ^{15}N with side chain Hβ resonances in larger proteins. *Journal of Magnetic Resonance (1969)* **1991,** *95* (3), 636-641.

204. Richard, J.-A.; Kelly, I.; Marion, D.; Pézolet, M.; Auger, M., Interaction between β-purothionin and dimyristoylphosphatidylglycerol: a ^{31}P-NMR and infrared spectroscopic study. *Biophysical journal* **2002,** *83* (4), 2074-2083.

205. Marion, D.; Driscoll, P. C.; Kay, L. E.; Wingfield, P. T.; Bax, A.; Gronenborn, A. M.; Clore, G. M., 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* **1989,** *28* (15), 6150-6156.

206. Olejniczak, E. T.; Xu, R. X.; Fesik, S. W., A 4D HCCH-TOCSY experiment for assigning the side chain 1 H and 13 C resonances of proteins. *Journal of biomolecular NMR* **1992,** *2* (6), 655-659.

207. Diercks, T.; Coles, M.; Kessler, H., An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. *Journal of biomolecular NMR* **1999,** *15* (2), 177-180.

208. Gandier, J.-A.; Langelaan, D. N.; Won, A.; O'Donnell, K.; Grondin, J. L.; Spencer, H. L.; Wong, P.; Tillier, E.; Yip, C.; Smith, S. P., Characterization of a Basidiomycota hydrophobin reveals the structural basis for a high-similarity Class I subdivision. *Scientific reports* **2017,** *7* (1), 1-9.

209. Keeler, J., *Understanding NMR spectroscopy*. John Wiley & Sons: 2011.

210. Kharchenko, V.; Nowakowski, M.; Jaremko, M.; Ejchart, A.; Jaremko, Ł., Dynamic ^{15}N $\{^{1}H\}$ NOE measurements: a tool for studying protein dynamics. *Journal of Biomolecular NMR* **2020,** *74* (12), 707-716.

211. Arakawa, S.; Shirayama, Y.; Fujita, Y.; Ishima, T.; Horio, M.; Muneoka, K.; Iyo, M.; Hashimoto, K., Minocycline produced antidepressant-like effects on the learned helplessness rats with alterations in levels of monoamine in the amygdala and no changes in BDNF levels in the hippocampus at baseline. *Pharmacology Biochemistry and Behavior* **2012,** *100* (3), 601-606.

212. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* **1995,** *6* (3), 277-293.

213. Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D., The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins: structure, function, and bioinformatics* **2005,** *59* (4), 687-696.

214. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D., ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *Journal of biomolecular NMR* **1995,** *6* (2), 135-140.

215. Rule, G. S.; Hitchens, T. K., *Fundamentals of protein NMR spectroscopy*. Springer Science & Business Media: 2006; Vol. 5.

216. Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E., Backbone dynamics of a free and a phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **1994,** *33* (19), 5984-6003.

217. Wang, B.; Wang, Y.; Wishart, D. S., A probabilistic approach for validating protein NMR chemical shift assignments. *J Biomol NMR* **2010,** *47* (2), 85-99.

218. Wishart, D. S.; Sykes, B. D.; Richards, F. M., The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992,** *31* (6), 1647-1651.

219. Wishart, D. S.; Sykes, B. D., The ^{13}C Chemical-Shift Index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *Journal of Biomolecular NMR* **1994,** *4* (2), 171-180.

220. Mielke, S. P.; Krishnan, V. V., Characterization of protein secondary structure from NMR chemical shifts. *Prog Nucl Magn Reson Spectrosc* **2009,** *54* (3-4), 141-165.

221. Cheung, M.-S.; Maguire, M. L.; Stevens, T. J.; Broadhurst, R. W., DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. *Journal of Magnetic Resonance* **2010,** *202* (2), 223-233.

222. Skinner, S. P.; Goult, B. T.; Fogh, R. H.; Boucher, W.; Stevens, T. J.; Laue, E. D.; Vuister, G. W., Structure calculation, refinement and validation using CcpNmr Analysis. *Acta Crystallographica Section D: Biological Crystallography* **2015,** *71* (1), 154-161.

223. Nilges, M.; Macias, M. J.; O'Donoghue, S. I.; Oschkinat, H., Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *Journal of molecular biology* **1997,** *269* (3), 408-422.

224. Linge, J. P.; Williams, M. A.; Spronk, C. A.; Bonvin, A. M.; Nilges, M., Refinement of protein structures in explicit solvent. *Proteins: Structure, Function, and Bioinformatics* **2003,** *50* (3), 496-506.

225. Matsumura, M.; Signor, G.; Matthews, B. W., Substantial increase of protein stability by multiple disulphide bonds. *Nature* **1989,** *342* (6247), 291-3.

226. Fry, B. G.; Wüster, W.; Kini, R. M.; Brusic, V.; Khan, A.; Venkataraman, D.; Rooney, A., Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *Journal of molecular evolution* **2003,** *57* (1), 110-129.

227. Olamendi‐Portugal, T.; Batista, C. V.; Restano‐Cassulini, R.; Pando, V.; Villa‐Hernandez, O.; Zavaleta‐Martínez‐Vargas, A.; Salas‐Arruz, M. C.; de la Vega, R. C. R.; Becerril, B.; Possani, L. D., Proteomic analysis of the venom from the fish eating coral snake Micrurus surinamensis: novel toxins, their function and phylogeny. *Proteomics* **2008,** *8* (9), 1919-1932.

228. Undheim, E. A.; Jenner, R. A.; King, G. F., Centipede venoms as a source of drug leads. *Expert opinion on drug discovery* **2016,** *11* (12), 1139-1149.

229. Uzair, B.; Bint-e-Irshad, S.; Khan, B. A.; Azad, B.; Mahmood, T.; Rehman, M. U.; Braga, V. A., Scorpion venom peptides as a potential source for human drug candidates. *Protein and peptide letters* **2018,** *25* (7), 702-708.

230. Abyzov, A.; Salvi, N.; Schneider, R.; Maurin, D.; Ruigrok, R. W. H.; Jensen, M. R.; Blackledge, M., Identification of dynamic modes in an intrinsically disordered protein using temperature-dependent nmr relaxation. *Journal of the American Chemical Society* **2016,** *138* (19), 6240-6251.

231. Sattler, M.; Schleucher, J.; Griesinger, C., Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution. *Progress in nuclear magnetic resonance spectroscopy* **1999,** *34*, 93-158.

232. Wang, Y.; Zhao, S.; Somerville, R. L.; Jardetzky, O., Solution structure of the DNA‐binding domain of the TyrR protein of Haemophilus influenzae. *Protein Science* **2001,** *10* (3), 592-598.

233. Sharma, D.; Rajarathnam, K., ^{13}C NMR chemical shifts can predict disulfide bond formation. *Journal of Biomolecular NMR* **2000,** *18* (2), 165-171.

234. Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M., ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **2007,** *23* (3), 381-382.

235. Schwieters, C. D.; Bermejo, G. A.; Clore, G. M., Xplor‐NIH for molecular structure determination from NMR and other data sources. *Protein Science* **2018,** *27* (1), 26-40.

236. Langelaan, D. N.; Reddy, T.; Banks, A. W.; Dellaire, G.; Dupré, D. J.; Rainey, J. K., Structural features of the apelin receptor N-terminal tail and first transmembrane segment implicated in ligand binding and receptor trafficking. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **2013,** *1828* (6), 1471-1483.

237. DeLano, W. L., Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography* **2002,** *40* (1), 82-92.

238. Ramachandran, G.; Chandrasekaran, R.; Kopple, K. D., Variation of the NH–CαH coupling constant with dihedral angle in the NMR spectra of peptides. *Biopolymers: Original Research on Biomolecules* **1971,** *10* (11), 2113-2131.

239. LeMaster, D. M., NMR relaxation order parameter analysis of the dynamics of protein side chains. *Journal of the American Chemical Society* **1999,** *121* (8), 1726-1742.

240. Ding, J.; Rainey, J. K.; Xu, C.; Sykes, B. D.; Fliegel, L., Structural and functional characterization of transmembrane segment VII of the Na^+/H^+ exchanger isoform 1. *J Biol Chem* **2006,** *281* (40), 29817-29.

241. Reddy, T.; Ding, J.; Li, X.; Sykes, B. D.; Rainey, J. K.; Fliegel, L., Structural and functional characterization of transmembrane segment IX of the NHE1 isoform of the Na^+/H^+ exchanger. *J Biol Chem* **2008,** *283* (32), 22018-30.

242. Chagot, B.; Pimentel, C.; Dai, L.; Pil, J.; Tytgat, J.; Nakajima, T.; Corzo, G.; Darbon, H.; Ferrat, G., An unusual fold for potassium channel blockers: NMR structure of three toxins from the scorpion *Opisthacanthus madagascariensis*. *Biochemical Journal* **2005,** *388* (1), 263-271.

243. Dutertre, S.; Lewis, R. J., Use of venom peptides to probe ion channel structure and function. *J Biol Chem* **2010,** *285* (18), 13315-20.

244. Owsianik, G.; Talavera, K.; Voets, T.; Nilius, B., PERMEATION AND SELECTIVITY OF TRP CHANNELS. *Annual Review of Physiology* **2006,** *68* (1), 685-717.

245. Possani, L. D.; Merino, E.; Corona, M.; Bolivar, F.; Becerril, B., Peptides and genes coding for scorpion toxins that affect ion-channels. *Biochimie* **2000,** *82* (9), 861-868.

246. Almeida, D. D.; Torres, T. M.; Barbosa, E. G.; Lima, J. P. M. S.; de Freitas Fernandes-Pedrosa, M., Molecular approaches for structural characterization of a new potassium channel blocker from *Tityus stigmurus* venom: cDNA cloning, homology modeling, dynamic simulations and docking. *Biochemical and Biophysical Research Communications* **2013,** *430* (1), 113-118.

247. Saito, A.; Ueda, K.; Imamura, M.; Miura, N.; Atsumi, S.; Tabunoki, H.; Sato, R., Purification and cDNA cloning of a novel antibacterial peptide with a cysteine-stabilized αβ motif

from the longicorn beetle, *Acalolepta luxuriosa*. *Developmental & Comparative Immunology* **2004,** *28* (1), 1-7.

248. Seo, J.-Y.; Yaneva, R.; Cresswell, P., Viperin: A multifunctional, interferon-inducible protein that regulates virus replication. *Cell Host & Microbe* **2011,** *10* (6), 534-539.

249. Dhawan, R.; Joseph, S.; Sethi, A.; Lala, A. K., Purification and characterization of a short insect toxin from the venom of the scorpion *Buthus tamulus*. *FEBS Letters* **2002,** *528* (1), 261-266.

250. Booker, S. J.; Grove, T. L., Mechanistic and functional versatility of radical SAM enzymes. *F1000 Biol Rep* **2010,** *2*, 52-52.

251. Wang, S. C.; Frey, P. A., S-adenosylmethionine as an oxidant: the radical SAM superfamily. *Trends in Biochemical Sciences* **2007,** *32* (3), 101-110.

252. Simpson, L. L.; Maksymowych, A. B.; Park, J. B.; Bora, R. S., The role of the interchain disulfide bond in governing the pharmacological actions of botulinum toxin. *J Pharmacol Exp Ther* **2004,** *308* (3), 857-64.

253. Nayak, S. K.; Rathore, D.; Batra, J. K., Role of individual cysteine residues and disulfide bonds in the structure and function of Aspergillus ribonucleolytic toxin restrictocin. *Biochemistry* **1999,** *38* (31), 10052-8.

254. Payne, C. D.; Vadlamani, G.; Fisher, M. F.; Zhang, J.; Clark, R. J.; Mylne, J. S.; Rosengren, K. J., Defining the familial fold of the vicilin-buried peptide family. *Journal of Natural Products* **2020,** *83* (10), 3030-3040.

255. Zhang, J.; Payne, C. D.; Pouvreau, B.; Schaefer, H.; Fisher, M. F.; Taylor, N. L.; Berkowitz, O.; Whelan, J.; Rosengren, K. J.; Mylne, J. S., An ancient peptide family buried within vicilin precursors. *ACS chemical biology* **2019,** *14* (5), 979-993.

256. Barnham, K. J.; Dyke, T. R.; Kem, W. R.; Norton, R. S., Structure of neurotoxin B-IV from the marine Worm *Cerebratulus lacteus*: a helical hairpin cross-linked by disulphide bonding 11 Edited by P.E. Wright. *Journal of Molecular Biology* **1997,** *268* (5), 886-902.

257. Oparin, P. B.; Mineev, K. S.; Dunaevsky, Y. E.; Arseniev, A. S.; Belozersky, M. A.; Grishin, E. V.; Egorov, T. A.; Vassilevski, A. A., Buckwheat trypsin inhibitor with helical hairpin structure belongs to a new family of plant defence peptides. *Biochemical Journal* **2012,** *446* (1), 69-77.

258. Nolde, S. B.; Vassilevski, A. A.; Rogozhin, E. A.; Barinov, N. A.; Balashova, T. A.; Samsonova, O. V.; Baranov, Y. V.; Feofanov, A. V.; Egorov, T. A.; Arseniev, A. S., Disulfide-stabilized helical hairpin structure and activity of a novel antifungal peptide EcAMP1 from seeds of barnyard grass (*Echinochloa crus-galli*). *Journal of Biological Chemistry* **2011,** *286* (28), 25145-25153.

259. Srinivasan, K. N.; Sivaraja, V.; Huys, I.; Sasaki, T.; Cheng, B.; Kumar, T. K. S.; Sato, K.; Tytgat, J.; Yu, C.; San, B. C. C., κ-Hefutoxin1, a novel toxin from the scorpion *Heterometrus*

fulvipes with unique structure and function: IMPORTANCE OF THE FUNCTIONAL DIAD IN POTASSIUM CHANNEL SELECTIVITY. *Journal of Biological Chemistry* **2002,** *277* (33), 30040-30047.

260. Shai, Y.; Fox, J.; Caratsch, C.; Shih, Y.-L.; Edwards, C.; Lazarovici, P., Sequencing and synthesis of pardaxin, a polypeptide from the Red Sea Moses sole with ionophore activity. *FEBS letters* **1988,** *242* (1), 161-166.

261. Conners, R.; Konarev, A. V.; Forsyth, J.; Lovegrove, A.; Marsh, J.; Joseph-Horne, T.; Shewry, P.; Brady, R. L., An unusual helix-turn-helix protease inhibitory motif in a novel trypsin inhibitor from seeds of *Veronica (Veronica hederifolia L.*). *Journal of Biological Chemistry* **2007,** *282* (38), 27760-27768.

262. Sharpe, I. A.; Gehrmann, J.; Loughnan, M. L.; Thomas, L.; Adams, D. A.; Atkins, A.; Palant, E.; Craik, D. J.; Adams, D. J.; Alewood, P. F., Two new classes of conopeptides inhibit the α1-adrenoceptor and noradrenaline transporter. *Nature neuroscience* **2001,** *4* (9), 902-907.

263. Moller, C.; Rahmankhah, S.; Lauer-Fields, J.; Bubis, J.; Fields, G. B.; Mari, F., A novel conotoxin framework with a helix-loop-helix (Cs alpha/alpha) fold. *Biochemistry* **2005,** *44* (49), 15986-96.

264. McIntosh, J. M.; Corpuz, G. O.; Layer, R. T.; Garrett, J. E.; Wagstaff, J. D.; Bulaj, G.; Vyazovkina, A.; Yoshikami, D.; Cruz, L. J.; Olivera, B. M., Isolation and characterization of a novel conus peptide with apparent antinociceptive activity. *Journal of Biological Chemistry* **2000,** *275* (42), 32391-32397.

265. Chairatana, P.; Niramitranon, J.; Pongprayoon, P., Dynamics of human defensin 5 (HD5) self-assembly in solution: Molecular simulations/insights. *Computational biology and chemistry* **2019,** *83*, 107091.

266. Tang, Y.-Q.; Yuan, J.; Österlund, G.; Österlund, K.; Tran, D.; Miller, C. J.; Ouellette, A. J.; Selsted, M. E., A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated α-defensins. *Science* **1999,** *286* (5439), 498-502.

267. Selsted, M. E., θ-Defensins: cyclic antimicrobial peptides produced by binary ligation of truncated α-defensins. *Current Protein and Peptide Science* **2004,** *5* (5), 365-371.

268. Conibear, A. C.; Rosengren, K. J.; Daly, N. L.; Henriques, S. T.; Craik, D. J., The cyclic cystine ladder in θ-defensins is important for structure and stability, but not antibacterial activity. *Journal of Biological Chemistry* **2013,** *288* (15), 10830-10840.

269. Joseph, R.; Pahari, S.; Hodgson, W. C.; Kini, R. M., Hypotensive agents from snake venoms. *Current Drug Targets-Cardiovascular & Hematological Disorders* **2004,** *4* (4), 437-459.

270. Guo, M.; Teng, M.; Niu, L.; Liu, Q.; Huang, Q.; Hao, Q., Crystal structure of the cysteine-rich secretory protein stecrisp reveals that the cysteine-rich domain has a K^+ channel inhibitor-like fold. *Journal of Biological Chemistry* **2005,** *280* (13), 12405-12412.

271. Mourão, C. B.; Schwartz, E. F., Protease inhibitors from marine venomous animals and their counterparts in terrestrial venomous animals. *Marine drugs* **2013,** *11* (6), 2069-2112.

272. Hasegawa, Y.; Honma, T.; Nagai, H.; Ishida, M.; Nagashima, Y.; Shiomi, K., Isolation and cDNA cloning of a potassium channel peptide toxin from the sea anemone *Anemonia erythraea*. *Toxicon* **2006,** *48* (5), 536-542.

273. Sun, P.; Wu, F.; Wen, M.; Yang, X.; Wang, C.; Li, Y.; He, S.; Zhang, L.; Zhang, Y.; Tian, C., A distinct three-helix centipede toxin SSD609 inhibits Iks channels by interacting with the KCNE1 auxiliary subunit. *Scientific Reports* **2015,** *5* (1), 13399.

274. Rabdano, S. O.; Izmailov, S. A.; Luzik, D. A.; Groves, A.; Podkorytov, I. S.; Skrynnikov, N. R., Onset of disorder and protein aggregation due to oxidation-induced intermolecular disulfide bonds: case study of RRM2 domain from TDP-43. *Scientific Reports* **2017,** *7* (1), 11161.

275. Tremblay, M.-L.; Xu, L.; Liu, P. X.-Q.; Rainey, J. K., Characterizing recombinant spider wrapping silk monomers and fibers by NMR and AFM. *Biophysical journal* **2011,** *100* (3), 602a-603a.

276. Tremblay, M.-L.; Xu, L.; Sarker, M.; Liu, X.-Q.; Rainey, J. K., Characterizing aciniform silk repetitive domain backbone dynamics and hydrodynamic modularity. *International journal of molecular sciences* **2016,** *17* (8), 1305.

277. Sarker, M.; Orrell, K. E.; Xu, L.; Tremblay, M.-L.; Bak, J. J.; Liu, X.-Q.; Rainey, J. K., Tracking transitions in spider wrapping silk conformation and dynamics by ^{19}F nuclear magnetic resonance spectroscopy. *Biochemistry* **2016,** *55* (21), 3048-3059.

278. Xu, L.; Weatherbee‐Martin, N.; Liu, X. Q.; Rainey, J. K., Biomaterials: recombinant silk fiber properties correlate to prefibrillar self‐assembly (Small 12/2019). *Small* **2019,** *15* (12), 1970065.

279. Cordero, B.; Gómez, V.; Platero-Prats, A. E.; Revés, M.; Echeverría, J.; Cremades, E.; Barragán, F.; Alvarez, S., Covalent radii revisited. *Dalton Transactions* **2008,** (21), 2832-2838.

280. Ayoub, N. A.; Garb, J. E.; Kuelbs, A.; Hayashi, C. Y., Ancient properties of spider silks revealed by the complete gene sequence of the prey-wrapping silk protein (AcSp1). *Molecular biology and evolution* **2013,** *30* (3), 589-601.

281. Vollrath, F.; Knight, D. P., Liquid crystalline spinning of spider silk. *Nature* **2001,** *410* (6828), 541-548.

282. Jin, H.-J.; Kaplan, D. L., Mechanism of silk processing in insects and spiders. *Nature* **2003,** *424* (6952), 1057-1061.

283. Römer, L.; Scheibel, T., The elaborate structure of spider silk: structure and function of a natural high performance fiber. *Prion* **2008,** *2* (4), 154-161.

284. Ko, F. K.; Jovicic, J., Modeling of mechanical properties and structural design of spider web. *Biomacromolecules* **2004,** *5* (3), 780-785.

285. Hijirida, D. H.; Do, K. G.; Michal, C.; Wong, S.; Zax, D.; Jelinski, L. W., ^{13}C NMR of *Nephila clavipes* major ampullate silk gland. *Biophysical journal* **1996,** *71* (6), 3442-3447.

286. Riekel, C.; Bränden, C.; Craig, C.; Ferrero, C.; Heidelbach, F.; Müller, M., Aspects of X-ray diffraction on single spider fibers. *International Journal of Biological Macromolecules* **1999,** *24* (2-3), 179-186.

287. Sparkes, J.; Holland, C., Analysis of the pressure requirements for silk spinning reveals a pultrusion dominated process. *Nature communications* **2017,** *8* (1), 1-10.

288. Koeppel, A.; Laity, P. R.; Holland, C., Extensional flow behaviour and spinnability of native silk. *Soft Matter* **2018,** *14* (43), 8838-8845.

289. Arndt, T.; Laity, P. R.; Johansson, J.; Holland, C.; Rising, A., Native-like flow properties of an artificial spider silk dope. *ACS Biomaterials Science & Engineering* **2021,** *7* (2), 462-471.

290. Ayutsede, J.; Gandhi, M.; Sukigara, S.; Micklus, M.; Chen, H.-E.; Ko, F., Regeneration of *Bombyx mori* silk by electrospinning. Part 3: characterization of electrospun nonwoven mat. *Polymer* **2005,** *46* (5), 1625-1634.

291. Arcidiacono, S.; Mello, C. M.; Butler, M.; Welsh, E.; Soares, J. W.; Allen, A.; Ziegler, D.; Laue, T.; Chase, S., Aqueous processing and fiber spinning of recombinant spider silks. *Macromolecules* **2002,** *35* (4), 1262-1266.

292. Yan, J.; Zhou, G.; Knight, D. P.; Shao, Z.; Chen, X., Wet-spinning of regenerated silk fiber from aqueous silk fibroin solution: discussion of spinning parameters. *Biomacromolecules* **2010,** *11* (1), 1-5.

293. Teulé, F.; Furin, W. A.; Cooper, A. R.; Duncan, J. R.; Lewis, R. V., Modifications of spider silk sequences in an attempt to control the mechanical properties of the synthetic fibers. *Journal of Materials Science* **2007,** *42* (21), 8974-8985.

294. Teulé, F.; Addison, B.; Cooper, A. R.; Ayon, J.; Henning, R. W.; Benmore, C. J.; Holland, G. P.; Yarger, J. L.; Lewis, R. V., Combining flagelliform and dragline spider silk motifs to produce tunable synthetic biopolymer fibers. *Biopolymers* **2012,** *97* (6), 418-431.

295. Ki, C. S.; Lee, K. H.; Baek, D. H.; Hattori, M.; Um, I. C.; Ihm, D. W.; Park, Y. H., Dissolution and wet spinning of silk fibroin using phosphoric acid/formic acid mixture solvent system. *Journal of applied polymer science* **2007,** *105* (3), 1605-1610.

296. Kluge, J. A.; Rabotyagova, O.; Leisk, G. G.; Kaplan, D. L., Spider silks and their applications. *Trends in biotechnology* **2008,** *26* (5), 244-251.

297. An, B.; Hinman, M. B.; Holland, G. P.; Yarger, J. L.; Lewis, R. V., Inducing β-sheets formation in synthetic spider silk fibers by aqueous post-spin stretching. *Biomacromolecules* **2011,** *12* (6), 2375-2381.

298. Gnesa, E.; Hsia, Y.; Yarger, J. L.; Weber, W.; Lin-Cereghino, J.; Lin-Cereghino, G.; Tang, S.; Agari, K.; Vierra, C., Conserved C-terminal domain of spider tubuliform spidroin 1 contributes to extensibility in synthetic fibers. *Biomacromolecules* **2012,** *13* (2), 304-312.

299. Albertson, A. E.; Teulé, F.; Weber, W.; Yarger, J. L.; Lewis, R. V., Effects of different post-spin stretching conditions on the mechanical properties of synthetic spider silk fibers. *Journal of the mechanical behavior of biomedical materials* **2014,** *29*, 225-234.

300. Tucker, C. L.; Jones, J. A.; Bringhurst, H. N.; Copeland, C. G.; Addison, J. B.; Weber, W. S.; Mou, Q.; Yarger, J. L.; Lewis, R. V., Mechanical and physical properties of recombinant spider silk films using organic and aqueous solvents. *Biomacromolecules* **2014,** *15* (8), 3158-3170.

301. ISHIZAKA, H.; WATANABE, Y.; ISHIDA, K.; FUKUMOTO, O., Regenerated silk prepeared from ortho phosphoric acid solution of fibroin. *The Journal of Sericultural Science of Japan* **1989,** *58* (2), 87-95.

302. Yao, J.; Masuda, H.; Zhao, C.; Asakura, T., Artificial spinning and characterization of silk fiber from *Bombyx mori* silk fibroin in hexafluoroacetone hydrate. *Macromolecules* **2002,** *35* (1), 6-9.

303. Inglis, A.; Nicholls, P.; Roxburgh, C., Hydrolysis of the peptide bond and amino acid modification with hydriodic acid. *Australian journal of biological sciences* **1971,** *24* (4), 1235-1240.

304. dos Santos-Pinto, J. R. A.; Lamprecht, G.; Chen, W.-Q.; Heo, S.; Hardy, J. G.; Priewalder, H.; Scheibel, T. R.; Palma, M. S.; Lubec, G., Structure and post-translational modifications of the web silk protein spidroin-1 from *Nephila* spiders. *Journal of proteomics* **2014,** *105*, 174-185.

305. Kozlovskaya, V.; Baggett, J.; Godin, B.; Liu, X.; Kharlampieva, E., Hydrogen-bonded multilayers of silk fibroin: from coatings to cell-mimicking shaped microcontainers. *ACS Macro Letters* **2012,** *1* (3), 384-387.

306. Wei, G.; Shea, J.-E., Effects of solvent on the structure of the alzheimer amyloid-β (25–35) Peptide. *Biophysical Journal* **2006,** *91* (5), 1638-1647.

307. Jones, J. A.; Harris, T. I.; Tucker, C. L.; Berg, K. R.; Christy, S. Y.; Day, B. A.; Gaztambide, D. A.; Needham, N. J.; Ruben, A. L.; Oliveira, P. F., More than just fibers: an aqueous method for the production of innovative recombinant spider silk protein materials. *Biomacromolecules* **2015,** *16* (4), 1418-1425.

308. Elices, M.; Guinea, G. V.; Pérez-Rigueiro, J.; Plaza, G. R., Finding inspiration in *Argiope trifasciata* spider silk fibers. *Jom* **2005,** *57* (2), 60-66.

309. Zhang, K.-H.; Yu, Q.-Z.; Mo, X.-M., Fabrication and intermolecular interactions of silk fibroin/hydroxybutyl chitosan blended nanofibers. *International journal of molecular sciences* **2011,** *12* (4), 2187-2199.

310. Perea, G. B.; Solanas, C.; Marí-Buyé, N.; Madurga, R.; Agulló-Rueda, F.; Muinelo, A.; Riekel, C.; Burghammer, M.; Jorge, I.; Vázquez, J., The apparent variability of silkworm (*Bombyx mori*) silk and its relationship with degumming. *European Polymer Journal* **2016,** *78*, 129-140.

311. Shao, Z.; Young, R. J.; Vollrath, F., The effect of solvents on spider silk studied by mechanical testing and single-fibre Raman spectroscopy. *International journal of biological macromolecules* **1999,** *24* (2-3), 295-300.

312. Zhang, F.; Lu, Q.; Yue, X.; Zuo, B.; Qin, M.; Li, F.; Kaplan, D. L.; Zhang, X., Regeneration of high-quality silk fibroin fiber by wet spinning from $CaCl_2$–formic acid solvent. *Acta Biomaterialia* **2015,** *12*, 139-145.

313. Um, I. C.; Kweon, H.; Lee, K. G.; Ihm, D. W.; Lee, J.-H.; Park, Y. H., Wet spinning of silk polymer: I. Effect of coagulation conditions on the morphological feature of filament. *International journal of biological macromolecules* **2004,** *34* (1-2), 89-105.

314. Lee, K. H.; Baek, D. H.; Ki, C. S.; Park, Y. H., Preparation and characterization of wet spun silk fibroin/poly(vinyl alcohol) blend filaments. *International Journal of Biological Macromolecules* **2007,** *41* (2), 168-172.

315. Koeppel, A.; Holland, C., Progress and trends in artificial silk spinning: a systematic review. *ACS Biomaterials Science & Engineering* **2017,** *3* (3), 226-237.

316. Tremblay, M.-L., The structural characterization of *Argiope trifasciata* spider wrapping silk by solution-state NMR. **2016**.

317. Maulik, V.; Jennifer, S.; Teruna, J., The role of thiols and disulfides in protein chemical and physical stability. *Curr Protein Pept Sci* **2009,** *10*, 614-25.

318. Wang, L.; Li, S.-j.; Sidhu, A.; Zhu, L.; Liang, Y.; Freedman, R. B.; Wang, C.-c., Reconstitution of human Ero1-Lα/protein-disulfide isomerase oxidative folding pathway *in vitro*: position-dependent differences in role between the a and a' domains of protein-disulfide isomerase. *Journal of Biological Chemistry* **2009,** *284* (1), 199-206.

319. Tu, B. P.; Weissman, J. S., The FAD-and O_2-dependent reaction cycle of Ero1-mediated oxidative protein folding in the endoplasmic reticulum. *Molecular cell* **2002,** *10* (5), 983-994.

320. Callister, W. D.; Rethwisch, D. G., *Materials science and engineering*. John wiley & sons New York: 2011; Vol. 5.

321. Bauer, E., Surface electron microscopy: the first thirty years. *Surface science* **1994,** *299*, 102-115.

322. Einstein, A., On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. *Annalen der physik* **1905,** *17* (549-560), 208.

323. Uhlenbeck, G. E.; Ornstein, L. S., On the theory of the Brownian motion. *Physical review* **1930,** *36* (5), 823.

324. Strutt, J. W., LVIII. On the scattering of light by small particles. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1871,** *41* (275), 447-454.

325. Chang, T.; Han, C.; Wheeler, L.; Lodge, T. P., Comparison of diffusion coefficients in ternary polymer solutions measured by dynamic light scattering and forced Rayleigh scattering. *Macromolecules* **1988,** *21* (6), 1870-1872.

326. Lechner, M., Influence of Mie scattering on nanoparticles with different particle sizes and shapes: photometry and analytical ultracentrifugation with absorption optics. *Journal of the Serbian Chemical Society* **2005,** *70* (3), 361-369.

327. Stetefeld, J.; McKenna, S. A.; Patel, T. R., Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical reviews* **2016,** *8* (4), 409-427.

328. Phelps, C., Dynamic light scattering, with application to chemistry, biology and physics: BJ Berne & R. Pecora. Pp. 376. John Wiley 1976.£ 14.00 or $27.00. Wiley Online Library: 1977.

329. Morrison, I. D.; Grabowski, E.; Herb, C., Improved techniques for particle size determination by quasi-elastic light scattering. *Langmuir* **1985,** *1* (4), 496-501.

330. Provencher, S. W., CONTIN: a general purpose constrained regularization program for inverting noisy linear algebraic and integral equations. *Computer Physics Communications* **1982,** *27* (3), 229-242.

331. Scotti, A. e.; Liu, W.; Hyatt, J.; Herman, E.; Choi, H.; Kim, J.; Lyon, L.; Gasser, U.; Fernandez-Nieves, A., The CONTIN algorithm and its application to determine the size distribution of microgel suspensions. *The Journal of chemical physics* **2015,** *142* (23), 234905.

332. Liu, Y.; Shao, Z.; Vollrath, F., Relationships between supercontraction and mechanical properties of spider silk. *Nature materials* **2005,** *4* (12), 901-905.

333. Morris, K. F.; Johnson Jr, C. S., Diffusion-ordered two-dimensional nuclear magnetic resonance spectroscopy. *Journal of the American chemical society* **1992,** *114* (8), 3139-3141.

334. Wu, D.; Chen, A.; Johnson, C. S., An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. *Journal of magnetic resonance, Series A* **1995,** *115* (2), 260-264.

335. Stejskal, E. O.; Tanner, J. E., Spin diffusion measurements: spin echoes in the presence of a time‐dependent field gradient. *The journal of chemical physics* **1965,** *42* (1), 288-292.

336. Schumann, F. H.; Riepl, H.; Maurer, T.; Gronwald, W.; Neidig, K.-P.; Kalbitzer, H. R., Combined chemical shift changes and amino acid specific chemical shift mapping of protein–protein interactions. *Journal of biomolecular NMR* **2007,** *39* (4), 275-289.

337. Markley, J. L.; Putter, I.; Jardetzky, O., High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease. *Science* **1968,** *161* (3847), 1249-1251.

338. Markley, J.; Jardetzky, O., Nuclear magnetic resonance studies of the structure and binding sites of enzymes: XIV. Inhibitor binding to staphylococcal nuclease. *Journal of molecular biology* **1970,** *50* (2), 223-233.

339. Spera, S.; Bax, A., Empirical correlation between protein backbone conformation and C. alpha. and C. beta. ^{13}C nuclear magnetic resonance chemical shifts. *Journal of the American Chemical Society* **1991,** *113* (14), 5490-5492.

340. De Dios, A. C.; Pearson, J. G.; Oldfield, E., Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. *Science* **1993,** *260* (5113), 1491-1496.

341. Le, H.-b.; Pearson, J. G.; De Dios, A. C.; Oldfield, E., Protein structure refinement and prediction via NMR chemical shifts and quantum chemistry. *Journal of the American Chemical Society* **1995,** *117* (13), 3800-3807.

342. Fiebig, K. M.; Rice, L. M.; Pollock, E.; Brunger, A. T., Folding intermediates of SNARE complex assembly. *Nature structural biology* **1999,** *6* (2), 117-123.

343. Peterson, F. C.; Gordon, N. C.; Gettins, P. G., Formation of a noncovalent serpin-proteinase complex involves no conformational change in the serpin. Use of ^{1}H-^{15}N HSQC NMR as a sensitive nonperturbing monitor of conformation. *Biochemistry* **2000,** *39* (39), 11884-92.

344. Avbelj, F.; Kocjan, D.; Baldwin, R. L., Protein chemical shifts arising from α-helices and β-sheets depend on solvent exposure. *Proceedings of the National Academy of Sciences* **2004,** *101* (50), 17394-17397.

345. Eghbalnia, H. R.; Wang, L.; Bahrami, A.; Assadi, A.; Markley, J. L., Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements. *Journal of biomolecular NMR* **2005,** *32* (1), 71-81.

346. McNulty, B. C.; Tripathy, A.; Young, G. B.; Charlton, L. M.; Orans, J.; Pielak, G. J., Temperature-induced reversible conformational change in the first 100 residues of α-synuclein. *Protein Science* **2006,** *15* (3), 602-608.

347. Joseph, P. R. B.; Mosier, P. D.; Desai, U. R.; Rajarathnam, K., Solution NMR characterization of chemokine CXCL8/IL-8 monomer and dimer binding to glycosaminoglycans: structural plasticity mediates differential binding interactions. *Biochemical Journal* **2015,** *472* (1), 121-133.

348. Chakravarty, D.; Janin, J.; Robert, C. H.; Chakrabarti, P., Changes in protein structure at the interface accompanying complex formation. *IUCrJ* **2015,** *2* (6), 643-652.

349. Jameson, C. J., UNDERSTANDING NMR CHEMICAL SHIFTS. *Annual Review of Physical Chemistry* **1996,** *47* (1), 135-169.

350. Fushman, D.; Tjandra, N.; Cowburn, D., An approach to direct determination of protein dynamics from^{15}N NMR relaxation at multiple fields, independent of variable ^{15}N chemical shift anisotropy and chemical exchange contributions. *Journal of the American Chemical Society* **1999,** *121* (37), 8577-8582.

351. Göbl, C.; Tjandra, N., Application of solution NMR spectroscopy to study protein dynamics. *Entropy* **2012,** *14* (3), 581-598.

352. Kovermann, M.; Rogne, P.; Wolf-Watz, M., Protein dynamics and function from solution state NMR spectroscopy. *Quarterly reviews of biophysics* **2016,** *49*.

353. Gust, D.; Moon, R. B.; Roberts, J. D., Applications of natural-abundance nitrogen-15 nuclear magnetic resonance to large biochemically important molecules. *Proceedings of the National Academy of Sciences* **1975,** *72* (12), 4696-4700.

354. d'Auvergne, E. J.; Gooley, P. R., The use of model selection in the model-free analysis of protein dynamics. *Journal of biomolecular NMR* **2003,** *25* (1), 25-39.

355. Jarymowycz, V. A.; Stone, M. J., Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chemical reviews* **2006,** *106* (5), 1624-1671.

356. Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A. A.; Davidson, A. R.; Dobson, C. M.; Kay, L. E., Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **2004,** *430* (6999), 586-590.

357. Kadeřávek, P.; Zapletal, V.; Rabatinová, A.; Krásný, L.; Sklenář, V.; Žídek, L., Spectral density mapping protocols for analysis of molecular motions in disordered proteins. *Journal of biomolecular NMR* **2014,** *58* (3), 193-207.

358. Reddy, T.; Rainey, J. K., Interpretation of biomolecular NMR spin relaxation parameters. *Biochemistry and Cell Biology* **2010,** *88* (2), 131-142.

359. Farrow, N. A.; Zhang, O.; Szabo, A.; Torchia, D. A.; Kay, L. E., Spectral density function mapping using ^{15}N relaxation data exclusively. *Journal of Biomolecular NMR* **1995,** *6* (2), 153-162.

360. Ropars, V.; Bouguet-Bonnet, S.; Auguin, D.; Barthe, P.; Canet, D.; Roumestand, C., Unraveling protein dynamics through fast spectral density mapping. *Journal of biomolecular NMR* **2007,** *37* (3), 159-177.

361. Lefèvre, J.-F.; Dayie, K. T.; Peng, J. W.; Wagner, G., Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N− H spectral density functions. *Biochemistry* **1996,** *35* (8), 2674-2686.

362. Wang, Y.; Li, C.; Pielak, G. J., Effects of proteins on protein diffusion. *Journal of the American Chemical Society* **2010,** *132* (27), 9392-9397.

363. Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Skelton, N. J., *Protein NMR spectroscopy: principles and practice*. Elsevier: 1995.

364. Farrow, N. A.; Zhang, O.; Forman-Kay, J. D.; Kay, L. E., Characterization of the backbone dynamics of folded and denatured states of an SH3 domain. *Biochemistry* **1997,** *36* (9), 2390-2402.

365. Wen, Y.; Li, J.; Xiong, M.; Peng, Y.; Yao, W.; Hong, J.; Lin, D., Solution structure and dynamics of the I214V mutant of the rabbit prion protein. *PLoS One* **2010,** *5* (10), e13273.

366. Twomey, E. C.; Cordasco, D. F.; Wei, Y., Profound conformational changes of PED/PEA-15 in ERK2 complex revealed by NMR backbone dynamics. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* **2012,** *1824* (12), 1382-1393.

367. Ugwu, S. O.; Apte, S. P., The effect of buffers on protein conformational stability. *Pharmaceutical Technology* **2004,** *28* (3), 86-109.

368. Ishikawa, H.; Kwak, K.; Chung, J. K.; Kim, S.; Fayer, M. D., Direct observation of fast protein conformational switching. *Proceedings of the National Academy of Sciences* **2008,** *105* (25), 8619-8624.

369. Fuxreiter, M.; Tompa, P., Fuzzy complexes: a more stochastic view of protein function. *Fuzziness* **2012**, 1-14.

370. Ishima, R., CPMG Relaxation Dispersion. Humana Press: 2014; pp 29-49.

371. Baldwin, A. J., An exact solution for R2,eff in CPMG experiments in the case of two site chemical exchange. *Journal of Magnetic Resonance* **2014,** *244*, 114-124.

372. Levitt, M. H.; Freeman, R., NMR population inversion using a composite pulse. *Journal of Magnetic Resonance (1969)* **1979,** *33* (2), 473-476.

373. Majumdar, S.; Orphanoudakis, S.; Gmitro, A.; O'donnell, M.; Gore, J., Errors in the measurements of T2 using multiple ‐ echo MRI techniques. I. Effects of radiofrequency pulse imperfections. *Magnetic resonance in medicine* **1986,** *3* (3), 397-417.

374. Korzhnev, D. M.; Orekhov, V. Y.; Kay, L. E., Off-resonance R1ρ NMR studies of exchange dynamics in proteins with low spin-lock fields: an application to a Fyn SH3 domain. *Journal of the American Chemical Society* **2005,** *127* (2), 713-721.

375. Tollinger, M.; Skrynnikov, N. R.; Mulder, F. A.; Forman-Kay, J. D.; Kay, L. E., Slow dynamics in folded and unfolded states of an SH3 domain. *Journal of the American Chemical Society* **2001,** *123* (46), 11341-11352.

376. Bieri, M.; d'Auvergne, E. J.; Gooley, P. R., relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and μs motion of proteins. *Journal of biomolecular NMR* **2011,** *50* (2), 147-155.

377. Bhunia, A.; Domadia, P. N.; Torres, J.; Hallock, K. J.; Ramamoorthy, A.; Bhattacharjya, S., NMR Structure of Pardaxin, a Pore-forming Antimicrobial Peptide, in Lipopolysaccharide Micelles: MECHANISM OF OUTER MEMBRANE PERMEABILIZATION 2. *Journal of biological chemistry* **2010,** *285* (6), 3883-3895.

378. Kay, L. E.; Torchia, D. A.; Bax, A., Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **1989,** *28* (23), 8972-9.

379. Lipari, G.; Szabo, A., Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* **1982,** *104* (17), 4546-4559.

380. Lipari, G.; Szabo, A., Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *Journal of the American Chemical Society* **1982,** *104* (17), 4559-4570.

381. Choy, W.-Y.; Shortle, D.; Kay, L. E., Side chain dynamics in unfolded protein states: an NMR based 2H spin relaxation study of Δ131Δ. *Journal of the American Chemical Society* **2003,** *125* (7), 1748-1758.

382. Nederveen, A. J.; Bonvin, A. M., NMR relaxation and internal dynamics of ubiquitin from a 0.2 μs MD simulation. *Journal of Chemical Theory and Computation* **2005,** *1* (3), 363-374.

383. Showalter, S. A.; Brüschweiler, R., Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: application to the AMBER99SB force field. *Journal of chemical theory and computation* **2007,** *3* (3), 961-975.

384. Xue, Y.; Skrynnikov, N. R., Motion of a disordered polypeptide chain as studied by paramagnetic relaxation enhancements, ^{15}N relaxation, and molecular dynamics simulations: how fast is segmental diffusion in denatured ubiquitin? *Journal of the American Chemical Society* **2011,** *133* (37), 14614-14628.

385. Jensen, M. R.; Markwick, P. R. L.; Meier, S.; Griesinger, C.; Zweckstetter, M.; Grzesiek, S.; Bernadó, P.; Blackledge, M., Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure* **2009,** *17* (9), 1169-1185.

386. Viles, J. H.; Donne, D.; Kroon, G.; Prusiner, S. B.; Cohen, F. E.; Dyson, H. J.; Wright, P. E., Local structural plasticity of the prion protein. Analysis of NMR relaxation dynamics. *Biochemistry* **2001,** *40* (9), 2743-2753.

387. Nisthal, A.; Wang, C. Y.; Ary, M. L.; Mayo, S. L., Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proceedings of the National Academy of Sciences* **2019,** *116* (33), 16367-16377.

388. Grip, S.; Johansson, J.; Hedhammar, M., Engineered disulfides improve mechanical properties of recombinant spider silk. *Protein Science* **2009,** *18* (5), 1012-1022.

389. McLendon, G.; Radany, E., Is protein turnover thermodynamically controlled? *Journal of Biological Chemistry* **1978,** *253* (18), 6335-6337.

390. Parsell, D. A.; Sauer, R. T., The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*. *Journal of Biological Chemistry* **1989,** *264* (13), 7590-7595.

391. Heim, M.; Ackerschott, C. B.; Scheibel, T., Characterization of recombinantly produced spider flagelliform silk domains. *Journal of Structural Biology* **2010,** *170* (2), 420-425.

392. Kawasaki, H.; Sato, H.; Suzuki, M., Structural proteins in the silkworm egg-shells. *Insect Biochemistry* **1971,** *1* (2), 130-148.

393. Joseph, A. M.; George, B.; Madhusoodanan, K.; Alex, R., Current status of sulphur vulcanization and devulcanization chemistry: Process of vulcanization. *Rubber Science* **2015,** *28* (1), 82-121.

394. Plowman, J. E.; Miller, R. E.; Thomas, A.; Grosvenor, A. J.; Harland, D. P.; Deb‐Choudhury, S., A detailed mapping of the readily accessible disulphide bonds in the cortex of wool fibres. *Proteins: Structure, Function, and Bioinformatics* **2021,** *89* (6), 708-720.

395. Elices, M.; Plaza, G. R.; Pérez-Rigueiro, J.; Guinea, G. V., The hidden link between supercontraction and mechanical behavior of spider silks. *Journal of the Mechanical Behavior of Biomedical Materials* **2011,** *4* (5), 658-669.

396. Cohen, N.; Levin, M.; Eisenbach, C. D., On the Origin of supercontraction in spider silk. *Biomacromolecules* **2021,** *22* (2), 993-1000.

397. Bell, F. I.; McEwen, I. J.; Viney, C., Supercontraction stress in wet spider dragline. *Nature* **2002,** *416* (6876), 37-37.

398. Schäfer, A.; Vehoff, T.; Glišović, A.; Salditt, T., Spider silk softening by water uptake: an AFM study. *European Biophysics Journal* **2008,** *37* (2), 197-204.

399. Jelinski, L. W.; Blye, A.; Liivak, O.; Michal, C.; LaVerde, G.; Seidel, A.; Shah, N.; Yang, Z., Orientation, structure, wet-spinning, and molecular basis for supercontraction of spider dragline silk. *International journal of biological macromolecules* **1999,** *24* (2-3), 197-201.

400. Yang, Z.; Liivak, O.; Seidel, A.; LaVerde, G.; Zax, D. B.; Jelinski, L. W., Supercontraction and backbone dynamics in spider silk: ^{13}C and ^{2}H NMR studies. *Journal of the American Chemical Society* **2000,** *122* (37), 9019-9025.

401. Agnarsson, I.; Boutry, C.; Wong, S.-C.; Baji, A.; Dhinojwala, A.; Sensenig, A. T.; Blackledge, T. A., Supercontraction forces in spider dragline silk depend on hydration rate. *Zoology* **2009,** *112* (5), 325-331.

402. Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R., The protein folding problem. *Annu. Rev. Biophys.* **2008,** *37*, 289-316.

403. Miroux, B.; Walker, J. E., Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **1996,** *260* (3), 289-98.

404. Dumon-Seignovert, L.; Cariot, G.; Vuillard, L., The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif* **2004,** *37* (1), 203-6.

405. Lobstein, J.; Emrich, C. A.; Jeans, C.; Faulkner, M.; Riggs, P.; Berkmen, M., SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories* **2012,** *11* (1), 753.

406. Jung, A.; Bamann, C.; Kremer, W.; Kalbitzer, H. R.; Brunner, E., High-temperature solution NMR structure of TmCsp. *Protein Sci* **2004,** *13* (2), 342-350.

407. Chung, H.; Kim, T. Y.; Lee, S. Y., Recent advances in production of recombinant spider silk proteins. *Current Opinion in Biotechnology* **2012,** *23* (6), 957-964.

408. Fawzi, N. L.; Ying, J.; Ghirlando, R.; Torchia, D. A.; Clore, G. M., Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. *Nature* **2011,** *480* (7376), 268-272.

409. Fawzi, N. L.; Ying, J.; Torchia, D. A.; Clore, G. M., Probing exchange kinetics and atomic resolution dynamics in high-molecular-weight complexes using dark-state exchange saturation transfer NMR spectroscopy. *Nature protocols* **2012,** *7* (8), 1523-1533.

410. Anthis, N. J.; Clore, G. M., Visualizing transient dark states by NMR spectroscopy. *Quarterly Reviews of Biophysics* **2015,** *48* (1), 35-116.

411. Brandts, J. F.; Oliveira, R. J.; Westort, C., Thermodynamics of protein denaturation. Effect of pressure on the denaturation on ribonuclease A. *Biochemistry* **1970,** *9* (4), 1038-1047.

412. Hawley, S. A., Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry* **1971,** *10* (13), 2436-2442.

413. Vajpai, N.; Nisius, L.; Wiktor, M.; Grzesiek, S., High-pressure NMR reveals close similarity between cold and alcohol protein denaturation in ubiquitin. *Proceedings of the National Academy of Sciences* **2013,** *110* (5), E368-E376.

414. Baldwin, R. L., Temperature dependence of the hydrophobic interaction in protein folding. *Proceedings of the National Academy of Sciences* **1986,** *83* (21), 8069-8072.

415. Privalov, P. L., Stability of proteins small globular proteins. *Advances in protein chemistry* **1979,** *33*, 167-241.

416. Heremans, K.; Smeller, L., Protein structure and dynamics at high pressure. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology* **1998,** *1386* (2), 353-370.

417. Acampora, G.; Hermans, J., Reversible denaturation of sperm whale myoglobin. I. Dependence on temperature, pH, and composition. *Journal of the American Chemical Society* **1967,** *89* (7), 1543-1547.

418. Puett, D., The equilibrium unfolding parameters of horse and sperm whale myoglobin: effects of guanidine hydrochloride, urea, and acid. *Journal of Biological Chemistry* **1973,** *248* (13), 4623-4634.

419. Gayán, E.; Condón, S.; Álvarez, I.; Nabakabaya, M.; Mackey, B., Effect of pressure-induced changes in the ionization equilibria of buffers on inactivation of *Escherichia coli* and Staphylococcus aureus by high hydrostatic pressure. *Applied and environmental microbiology* **2013,** *79* (13), 4041-4047.

420. Zipp, A.; Kauzmann, W., Pressure denaturation of metmyoglobin. *Biochemistry* **1973,** *12* (21), 4217-4228.

421. Smeller, L., Pressure–temperature phase diagrams of biomolecules. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **2002,** *1595* (1), 11-29.

422. Bhate, S. H.; Udgaonkar, J. B.; Das, R., Destabilization of polar interactions in the prion protein triggers misfolding and oligomerization. *Protein Sci* **2021,** *30* (11), 2258-2271.

423. Schweickhardt, R. L.; Jiang, X.; Garone, L. M.; Brondyk, W. H., Structure-expression relationship of tumor necrosis factor receptor mutants that increase expression. *Journal of Biological Chemistry* **2003,** *278* (31), 28961-28967.

424. Sekhar, A.; Rumfeldt, J. A. O.; Broom, H. R.; Doyle, C. M.; Sobering, R. E.; Meiering, E. M.; Kay, L. E., Probing the free energy landscapes of ALS disease mutants of SOD1 by NMR spectroscopy. *Proceedings of the National Academy of Sciences* **2016,** *113* (45), E6939-E6945.

425. Sirichaisit, J.; Brookes, V. L.; Young, R. J.; Vollrath, F., Analysis of structure/property relationships in silkworm (Bombyx mori) and spider dragline (Nephila edulis) silks using Raman spectroscopy. *Biomacromolecules* **2003,** *4* (2), 387-394.

426. Shao, Z.; Vollrath, F.; Sirichaisit, J.; Young, R., Analysis of spider silk in native and supercontracted states using Raman spectroscopy. *Polymer* **1999,** *40* (10), 2493-2500.

427. Rabotyagova, O. S.; Cebe, P.; Kaplan, D. L., Self-assembly of genetically engineered spider silk block copolymers. *Biomacromolecules* **2009,** *10* (2), 229-236.

428. Dong, Z.; Lewis, R. V.; Middaugh, C. R., Molecular mechanism of spider silk elasticity. *Archives of biochemistry and biophysics* **1991,** *284* (1), 53-57.

429. Winkler, S.; Szela, S.; Avtges, P.; Valluzzi, R.; Kirschner, D. A.; Kaplan, D., Designing recombinant spider silk proteins to control assembly. *International journal of biological macromolecules* **1999,** *24* (2-3), 265-270.

430. Du, N.; Yang, Z.; Liu, X. Y.; Li, Y.; Xu, H. Y., Structural origin of the strain‐hardening of spider silk. *Advanced Functional Materials* **2011,** *21* (4), 772-778.

431. Holland, G. P.; Jenkins, J. E.; Creager, M. S.; Lewis, R. V.; Yarger, J. L., Solid-state NMR investigation of major and minor ampullate spider silk in the native and hydrated states. *Biomacromolecules* **2008,** *9* (2), 651-657.

432. Stoller, P.; Reiser, K. M.; Celliers, P. M.; Rubenchik, A. M., Polarization-modulated second harmonic generation in collagen. *Biophysical journal* **2002,** *82* (6), 3330-3342.

433. Chen, X.; Nadiarynkh, O.; Plotnikov, S.; Campagnola, P. J., Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. *Nature protocols* **2012,** *7* (4), 654-669.

434. Zhao, Y.; Hien, K. T. T.; Mizutani, G.; Rutt, H. N., Second-order nonlinear optical microscopy of spider silk. *Applied Physics B* **2017,** *123* (6), 188.

435. Guan, J.; Wang, Y.; Mortimer, B.; Holland, C.; Shao, Z.; Porter, D.; Vollrath, F., Glass transitions in native silk fibres studied by dynamic mechanical thermal analysis. *Soft Matter* **2016,** *12* (27), 5926-5936.

436. Malay, A. D.; Sato, R.; Yazawa, K.; Watanabe, H.; Ifuku, N.; Masunaga, H.; Hikima, T.; Guan, J.; Mandal, B. B.; Damrongsakkul, S., Relationships between physical properties and sequence in silkworm silks. *Scientific reports* **2016,** *6* (1), 1-11.

437. Tyagarajan, K.; Pretzer, E.; Wiktorowicz, J. E., Thiol‐reactive dyes for fluorescence labeling of proteomic samples. *Electrophoresis* **2003,** *24* (14), 2348-2358.

438. Shimada, K.; Mitamura, K., Derivatization of thiol-containing compounds. *Journal of Chromatography B: Biomedical Sciences and Applications* **1994,** *659* (1-2), 227-241.

439. Kalia, J.; Raines, R. T., Advances in bioconjugation. *Current organic chemistry* **2010,** *14* (2), 138-147.

440. Agarwal, P.; Bertozzi, C. R., Site-specific antibody–drug conjugates: the nexus of bioorthogonal chemistry, protein engineering, and drug development. *Bioconjugate chemistry* **2015,** *26* (2), 176-192.

441. Giamblanco, N.; Fichou, Y.; Janot, J.-M.; Balanzat, E.; Han, S.; Balme, S., Mechanisms of Heparin-Induced Tau Aggregation Revealed by a Single Nanopore. *ACS Sensors* **2020,** *5* (4), 1158-1167.

442. Bustamante, C. J.; Chemla, Y. R.; Liu, S.; Wang, M. D., Optical tweezers in single-molecule biophysics. *Nature Reviews Methods Primers* **2021,** *1* (1).

443. Lehmann, K.; Shayegan, M.; Blab, G. A.; Forde, N. R., Optical tweezers approaches for probing multiscale protein mechanics and assembly. *Frontiers in Molecular Biosciences* **2020,** *7*.

444. Kenney, J. M.; Knight, D.; Wise, M. J.; Vollrath, F., Amyloidogenic nature of spider silk. *European Journal of Biochemistry* **2002,** *269* (16), 4159-4163.

445. Grubb, D. T.; Jelinski, L. W., Fiber morphology of spider silk: the effects of tensile deformation. *Macromolecules* **1997,** *30* (10), 2860-2867.

446. Glišović, A.; Vehoff, T.; Davies, R. J.; Salditt, T., Strain dependent structural changes of spider dragline silk. *Macromolecules* **2008,** *41* (2), 390-398.

447. Reches, M.; Gazit, E., Controlled patterning of aligned self-assembled peptide nanotubes. *Nature nanotechnology* **2006,** *1* (3), 195-200.

448. Knowles, T. P.; Oppenheim, T. W.; Buell, A. K.; Chirgadze, D. Y.; Welland, M. E., Nanostructured films from hierarchical self-assembly of amyloidogenic proteins. *Nature nanotechnology* **2010,** *5* (3), 204-207.

APPENDICES

***APPENDIX A**: Chemical shift assignments (ppm) for soricidin.*

	H	N	N δ	N ε	Hα	Hβ	Hγ	Hδ	Hε	C	Cα	Cβ	Cγ	Cδ	Cε	Cζ
1 Asp					4.79	3.24					55.05	43.50				
2 Cys	8.43	122.70			4.53	3.25				176.32	58.66	39.43				
						2.83										
3 Ser	8.45	116.88				3.95				177.09	62.47	62.22				
3 Ser					3.96											
4 Gln	8.00	122.75		111.98	4.11	2.15	2.51		6.75	179.13	58.87	28.70	34.21			
							2.39		7.49							
5 Asp	8.39	122.38			4.39	2.78				179.05	57.75	40.07				
						2.62										
6 Cys	8.52	117.46			4.32	3.06				176.57	59.95	40.74				
						2.72										
7 Ala	7.95	124.63			4.15	1.54				180.59	55.39	17.71				
8 Ala	7.93	121.37			4.16	1.54				180.69	55.14	17.72				
9 Cys	8.25	116.54			4.06	3.16				175.88	57.55	38.73				
10 Ser	8.15	114.60			4.02	4.02				176.54	61.98	62.77				
11 Ile	7.33	120.73			3.78	1.96	0.91	0.86		178.48	64.31	38.08	28.64	12.57		
							1.66						17.33			
							1.21									
12 Leu								0.88						23.21		
								0.87						25.26		
12 Leu	7.53	120.16			4.09	1.72	1.74			177.73	56.91	42.22	27.03			
						1.50										
13 Ala	7.76	118.41			4.04	1.24				177.53	52.49	19.37				
14 Arg	7.07	117.90			4.33	1.81	2.10	3.23		175.01	55.99	29.24	27.03	43.71		
							1.78									
15 Pro					4.40	2.27	1.96	3.52		176.57	63.72	34.20	24.44	50.30		
						2.31	1.77	3.59								
16 Ala	8.54	129.37			4.31	1.40				176.56	52.91	18.95				
17 Glu	8.33	120.41			4.28	1.87	2.22			175.39	55.57	30.56	36.17			
						1.95	2.12									
18 Leu	8.24	126.17			4.31	1.54	1.37	0.73		175.86	54.52	43.89	27.11	25.39		
						1.28		0.71						25.03		
19 Asn	8.73	124.32	112.56		4.77	3.03		7.58		175.70	53.04	38.19				
						2.73		6.99								
20 Thr	8.53	123.03			3.63	4.11	1.18			175.37	66.73	68.49	22.46			
21 Glu	8.30	121.21			3.97	2.07	2.25			179.07	60.33	29.06	36.59			
						2.14	2.30									
22 Thr	7.83	114.60			3.90	4.13	1.22			175.73	66.20	68.38	22.40			
23 Cys	7.76	121.34			4.11	3.35				176.18	60.60	38.45				
						2.80										
24 Ile	8.50	122.98			3.39	1.89	1.86	0.80		177.57	66.33	37.77	17.33	13.35		
							0.71						30.17			
							0.82									
25 Leu	7.76	119.28			4.03	1.91	1.82	0.86		180.34	58.33	42.23	26.89	25.22		
						1.44		0.89						23.06		
26 Glu	8.28	119.59			4.17	2.11	2.21			179.40	59.16	29.45	36.11			
						2.17	2.44									
27 Cys	8.70	121.57			4.17	3.23				176.75	59.70	39.14				
						3.37										
28 Glu	8.77	119.44			4.21	2.06	2.33			177.96	57.92	29.97	36.84			
						2.15	2.58									

27 Cys	8.70	121.57			4.17	3.23				176.75	59.70	39.14				
						3.37										
28 Glu	8.77	119.44			4.21	2.06	2.33			177.96	57.92	29.97	36.84			
						2.15	2.58									
29 Gly	7.91	107.27			4.03					175.12	45.99					
30 Lys	8.00	119.24			4.30	1.78	1.45	1.67	2.98	176.45	56.75	33.19	25.31	29.18	42.48	
						1.99										
31 Leu	7.73	120.55			4.43	1.60	1.58	0.90		177.13	55.03	43.50	27.02	25.19		
						1.56		0.95						23.66		
32 Ser	8.31	116.36			4.49	3.88				174.55	58.32	63.91				
33 Ser	8.23	117.23			4.46	3.90				174.35	58.39	63.98				
						3.83										
34 Asn	8.39	120.11	112.66		4.71	2.85		6.84		174.95	53.58	38.91				
						2.74		7.54								
								6.99								
35 Asp	8.24	120.09			4.64	2.64				176.61	54.55	41.31				
						2.72										
36 Thr	8.07	113.55			4.29	4.25	1.18			175.11	62.36	69.69	21.71			
37 Glu	8.40	122.67			4.26	1.97	2.24			177.08	57.20	29.94	36.24			
						2.07										
38 Gly	8.36	109.29			3.93					174.91	45.67					
39 Gly	8.25	108.97			3.93					174.81	45.60					
					3.99											
40 Leu	8.28	122.13			4.15	1.66	1.63	0.90		178.05	57.12	42.62	26.74	24.78		
						1.57		0.85						24.07		
41 Cys	8.48	114.79			4.65	3.06				175.94	55.93	38.50				
						3.20										
42 Lys	7.82	120.19			3.93	1.79	1.41	1.66	2.97	177.67	59.52	32.28	24.63	29.06	42.15	
						1.84	1.40									
43 Glu	8.18	117.27			3.98	1.72	2.06			176.94	57.95	29.08	35.86			
44 Phe	7.58	116.28			4.50	3.21		7.25	7.21	175.97	58.31	39.95		131.48	131.88	129.76
						2.89										
45 Leu	7.54	118.39			4.19	1.28	1.64	0.77		176.60	55.27	42.57	27.03	25.34		
						1.55		0.84						23.06		
46 His	8.15	118.03		164.63	4.91	3.13		7.22	7.90	171.96	53.37	28.78		120.87	136.67	
									8.42							
47 Pro					4.48	2.31	1.97	3.53		177.41	63.56	32.17	27.32	50.50		
						1.92										
48 Ser	8.62	116.75			4.45	3.88				174.57	58.62	63.87				
49 Lys	8.24	123.18			4.38	1.74	1.42	1.66	2.97	176.20	56.32	33.11	24.67	28.74	42.30	
						1.85	1.36									
50 Val	8.02	120.63			4.13	2.04	0.88			175.46	62.06	33.01	21.15			
							0.89						20.25			
51 Asp	8.31	124.23			4.63	2.65				175.46	54.14	41.34				
						2.52										
52 Leu	7.98	123.74			4.60	1.56	1.64	0.84		175.06	52.92	41.99		24.30		
53 Pro					4.40	1.93	2.04	3.81		175.98	63.39	31.85	27.35	50.61		
						2.27	2.00	3.62								
54 Arg	7.87	126.53			4.16	1.84	1.61	3.18			57.27	31.68	27.18	43.59		
						1.70										

APPENDIX B: *PANAV verification of chemical shift assignments for soricidin*

Seq	C	CA	CB	N	H	HA	Beta sheet probability	Coil probability	Helix probability	SecStr
D1		55.0526	43.4985			4.7930	0.42	0.58	0.00	C
C2	176.3175	58.6556	39.4302	122.6950	8.4309	4.5332	0.18	0.48	0.34	C
S3	177.0919	62.4653	62.2204	116.8816	8.4499	3.9555	0.00	0.01	0.99	H
Q4	179.1255	58.8730	28.6955	122.7475	8.0018	4.1146	0.00	0.00	1.00	H
D5	179.0459	57.7496	40.0666	122.3774	8.3937	4.3947	0.00	0.00	1.00	H
C6	176.5716	59.9502	40.7366	117.4630	8.5224	4.3176	0.02	0.24	0.73	H
A7	180.5865	55.3853	17.7143	124.6286	7.9482	4.1472	0.00	0.00	1.00	H
A8	180.6913	55.1408	17.7194	121.3680	7.9310	4.1558	0.00	0.00	1.00	H
C9	175.8823	57.5476	38.7312	116.5387	8.2492	4.0562	0.07	0.33	0.61	H
S10	176.5403	61.9787	62.7731	114.5983	8.1544	4.0175	0.00	0.01	0.99	H
I11	178.4765	64.3094	38.0844	120.7346	7.3269	3.7789	0.00	0.01	0.99	H
L12	177.7269	56.9096	42.2187	120.1636	7.5257	4.0878	0.00	0.03	0.97	H
A13	177.5340	52.4867	19.3687	118.4083	7.7601	4.0387	0.01	0.96	0.03	C
R14	175.0081	55.9905	29.2364	117.9046	7.0675	4.3349	0.03	0.96	0.01	C
P15	176.5728	63.7164	34.2043			4.4022	0.17	0.83	0.00	C
A16	176.5554	52.9060	18.9548	129.3673	8.5449	4.3075	0.10	0.90	0.00	C
E17	175.3939	55.5724	30.5639	120.4135	8.3290	4.2757	0.43	0.57	0.00	C
L18	175.8604	54.5163	43.8856	126.1694	8.2420	4.3085	0.78	0.22	0.00	B
N19	175.6978	53.0443	38.1896	124.3165	8.7320	4.7726	0.33	0.67	0.00	C
T20	175.3698	66.7255	68.4877	123.0320	8.5264	3.6305	0.00	0.00	1.00	H
E21	179.0668	60.3277	29.0649	121.2140	8.2985	3.9672	0.00	0.00	1.00	H
T22	175.7306	66.1966	68.3848	114.6007	7.8276	3.9045	0.00	0.00	1.00	H
C23	176.1810	60.5977	38.4537	121.3368	7.7552	4.1111	0.00	0.09	0.91	H
I24	177.5733	66.3324	37.7738	122.9756	8.5023	3.3907	0.00	0.00	1.00	H
L25	180.3396	58.3323	42.2349	119.2805	7.7629	4.0350	0.00	0.00	1.00	H
E26	179.3994	59.1581	29.4518	119.5949	8.2798	4.1708	0.00	0.00	1.00	H
C27	176.7493	59.6999	39.1449	121.5699	8.6975	4.1730	0.05	0.33	0.62	H
E28	177.9610	57.9187	29.9650	119.4442	8.7681	4.2054	0.00	0.19	0.81	H
G29	175.1208	45.9852		107.2714	7.9139	4.0323	0.05	0.78	0.18	C
K30	176.4473	56.7498	33.1890	119.2426	7.9952	4.3046	0.05	0.93	0.01	C
L31	177.1284	55.0303	43.5015	120.5489	7.7276	4.4324	0.06	0.93	0.01	C
S32	174.5511	58.3246	63.9139	116.3601	8.3086	4.4900	0.12	0.87	0.01	C
S33	174.3502	58.3870	63.9793	117.2318	8.2286	4.4602	0.11	0.89	0.01	C
N34	174.9461	53.5806	38.9100	120.1129	8.3924	4.7053	0.10	0.85	0.05	C
D35	176.6094	54.5536	41.3060	120.0853	8.2446	4.6448	0.10	0.89	0.01	C
T36	175.1118	62.3640	69.6937	113.5478	8.0707	4.2892	0.23	0.70	0.06	C
E37	177.0848	57.2009	29.9372	122.6735	8.4022	4.2580	0.01	0.89	0.10	C

G38	174.9129	45.6744		109.2883	8.3573	3.9327	0.06	0.87	0.06	C
G39	174.8092	45.6013		108.9684	8.2453	3.9879	0.03	0.91	0.06	C
L40	178.0531	57.1178	42.6240	122.1322	8.2831	4.1456	0.00	0.04	0.96	H
C41	175.9426	55.9285	38.5017	114.7904	8.4807	4.6475	0.11	0.72	0.16	C
K42	177.6718	59.5239	32.2805	120.1857	7.8151	3.9306	0.00	0.01	0.99	H
E43	176.9388	57.9506	29.0831	117.2666	8.1846	3.9847	0.00	0.30	0.70	H
F44	175.9694	58.3142	39.9458	116.2814	7.5795	4.4996	0.02	0.90	0.08	C
L45	176.5984	55.2741	42.5651	118.3854	7.5431	4.1920	0.01	0.91	0.08	C
H46	171.9568	53.3651	28.7831	118.0268	8.1463	4.9138	0.35	0.65	0.00	C
P47	177.4054	63.5571	32.1670			4.4789	0.08	0.90	0.02	C
S48	174.5720	58.6151	63.8672	116.7452	8.6207	4.4507	0.21	0.78	0.01	C
K49	176.1971	56.3193	33.1050	123.1766	8.2439	4.3809	0.11	0.89	0.00	C
V50	175.4583	62.0598	33.0077	120.6298	8.0180	4.1349	0.36	0.64	0.00	C
D51	175.4551	54.1412	41.3392	124.2345	8.3077	4.6265	0.16	0.84	0.00	C
L52	175.0630	52.9191	41.9868	123.7435	7.9773	4.5984	0.70	0.30	0.00	B
P53	175.9789	63.3875	31.8531			4.3959	0.15	0.85	0.00	C
R54		57.2746	31.6829	126.5345	7.8729	4.1612	0.03	0.78	0.19	C

CSI and Sec Structure (No_Deviant_Ref_Calibrated)

Seq	C	CA	CB	N	H	HA	Beta sheet probability	Coil probability	Helix probability	SecStr
D1		54.79	43.68			4.7930	0.47	0.53	0.00	C
C2	176.15	58.40	39.61	122.64	8.4309	4.5332	0.23	0.48	0.29	C
S3	176.92	62.21	62.40	116.83	8.4499	3.9555	0.00	0.01	0.99	H
Q4	178.96	58.61	28.88	122.70	8.0018	4.1146	0.00	0.00	1.00	H
D5	178.88	57.49	40.25	122.33	8.3937	4.3947	0.00	0.01	0.99	H
C6	176.40	59.69	40.92	117.41	8.5224	4.3176	0.04	0.29	0.67	H
A7	180.42	55.13	17.89	124.58	7.9482	4.1472	0.00	0.00	1.00	H
A8	180.52	54.88	17.90	121.32	7.9310	4.1558	0.00	0.00	1.00	H
C9	175.71	57.29	38.91	116.49	8.2492	4.0562	0.09	0.37	0.54	H
S10	176.37	61.72	62.95	114.55	8.1544	4.0175	0.00	0.02	0.98	H
I11	178.31	64.05	38.26	120.68	7.3269	3.7789	0.00	0.01	0.99	H
L12	177.56	56.65	42.40	120.11	7.5257	4.0878	0.00	0.06	0.94	H
A13	177.36	52.23	19.55	118.36	7.7601	4.0387	0.02	0.97	0.01	C
R14	174.84	55.73	29.42	117.85	7.0675	4.3349	0.05	0.95	0.01	C
P15	176.40	63.46	34.38			4.4022	0.22	0.78	0.00	C
A16	176.39	52.65	19.13	129.32	8.5449	4.3075	0.16	0.84	0.00	C
E17	175.22	55.31	30.74	120.36	8.3290	4.2757	0.53	0.47	0.00	B
L18	175.69	54.26	44.07	126.12	8.2420	4.3085	0.84	0.16	0.00	B
N19	175.53	52.78	38.37	124.27	8.7320	4.7726	0.38	0.62	0.00	C
T20	175.20	66.47	68.67	122.98	8.5264	3.6305	0.00	0.00	1.00	H

E21	178.90	60.07	29.24	121.16	8.2985	3.9672	0.00	0.00	1.00	H
T22	175.56	65.94	68.56	114.55	7.8276	3.9045	0.00	0.00	1.00	H
C23	176.01	60.34	38.63	121.29	7.7552	4.1111	0.00	0.11	0.89	H
I24	177.40	66.07	37.95	122.93	8.5023	3.3907	0.00	0.00	1.00	H
L25	180.17	58.07	42.41	119.23	7.7629	4.0350	0.00	0.00	1.00	H
E26	179.23	58.90	29.63	119.54	8.2798	4.1708	0.00	0.01	0.99	H
C27	176.58	59.44	39.32	121.52	8.6975	4.1730	0.07	0.36	0.57	H
E28	177.79	57.66	30.14	119.39	8.7681	4.2054	0.00	0.38	0.62	H
G29	174.95	45.73		107.22	7.9139	4.0323	0.06	0.84	0.11	C
K30	176.28	56.49	33.37	119.19	7.9952	4.3046	0.09	0.91	0.01	C
L31	176.96	54.77	43.68	120.50	7.7276	4.4324	0.10	0.90	0.00	C
S32	174.38	58.06	64.09	116.31	8.3086	4.4900	0.16	0.84	0.00	C
S33	174.18	58.13	64.16	117.18	8.2286	4.4602	0.15	0.85	0.00	C
N34	174.78	53.32	39.09	120.06	8.3924	4.7053	0.12	0.85	0.03	C
D35	176.44	54.29	41.49	120.04	8.2446	4.6448	0.12	0.88	0.00	C
T36	174.94	62.10	69.87	113.50	8.0707	4.2892	0.29	0.68	0.03	C
E37	176.91	56.94	30.12	122.62	8.4022	4.2580	0.02	0.95	0.03	C
G38	174.74	45.41		109.24	8.3573	3.9327	0.07	0.89	0.04	C
G39	174.64	45.34		108.92	8.2453	3.9879	0.04	0.93	0.03	C
L40	177.88	56.86	42.80	122.08	8.2831	4.1456	0.00	0.08	0.92	H
C41	175.77	55.67	38.68	114.74	8.4807	4.6475	0.13	0.74	0.13	C
K42	177.50	59.26	32.46	120.14	7.8151	3.9306	0.00	0.02	0.98	H
E43	176.77	57.69	29.26	117.22	8.1846	3.9847	0.00	0.52	0.48	C
F44	175.80	58.05	40.13	116.23	7.5795	4.4996	0.03	0.92	0.05	C
L45	176.43	55.01	42.75	118.34	7.5431	4.1920	0.02	0.96	0.03	C
H46	171.79	53.11	28.96	117.98	8.1463	4.9138	0.38	0.62	0.00	C
P47	177.24	63.30	32.35			4.4789	0.11	0.89	0.01	C
S48	174.40	58.36	64.05	116.70	8.6207	4.4507	0.28	0.72	0.01	C
K49	176.03	56.06	33.28	123.13	8.2439	4.3809	0.16	0.84	0.00	C
V50	175.29	61.80	33.19	120.58	8.0180	4.1349	0.44	0.56	0.00	C
D51	175.29	53.88	41.52	124.18	8.3077	4.6265	0.20	0.80	0.00	C
L52	174.89	52.66	42.17	123.69	7.9773	4.5984	0.73	0.27	0.00	B
P53	175.81	63.13	32.03			4.3959	0.19	0.81	0.00	C
R54		57.01	31.86	126.48	7.8729	4.1612	0.05	0.83	0.12	C

Validation

3-Resdue Scan:

DCSQDCAACS ILARPAELNT ETCILECEGK LSSNDTEGGL CKEFLHPSKV DLPR

Original Assigment and Probability vs Suggested Assignment and Probability

LEC(25-27) 0.24% LCK(40-42) 100%
ECE(26-28) 0.03% QDC(4-6) 100%
CKE(41-43) 0.02% LEC(25-27) 100%

# of selected frags: 52; # of confirmed frags: 49; CONA Score: 94.23%

4-Resdue Scan:

DCSQDCAACS ILARPAELNT ETCILECEGK LSSNDTEGGL CKEFLHPSKV DLPR

Original Assigment and Probability vs Suggested Assignment and Probability

LECE(25-28) 0.24% LCKE(40-43) 100%
LCKE(40-43) 0.54% ILEC(24-27) 100%

# of selected frags: 51; # of confirmed frags: 49; CONA Score: 96.08%

5-Resdue Scan:

DCSQDCAACS ILARPAELNT ETCILECEGK LSSNDTEGGL CKEFLHPSKV DLPR

Original Assigment and Probability vs Suggested Assignment and Probability

# of selected frags: 50; # of confirmed frags: 50; CONA Score: 100.00%

6-Resdue Scan:

DCSQDCAACS ILARPAELNT ETCILECEGK LSSNDTEGGL CKEFLHPSKV DLPR

Original Assigment and Probability vs Suggested Assignment and Probability

# of selected frags: 49; # of confirmed frags: 49; CONA Score: 100.00%

Reference Offsets and Deviant Shifts

Detected reference offsets
CO: -0.17ppm CA: -0.26ppm CB: 0.18ppm N: -0.05ppm

Number of assignments: 311
Number of deviant assignments: 0

Number of suspicious assignments after ref-calibration when necessary: 0

APPENDIX C: *DOSY data analysis using Dynamic center**

Amide resonances were picked from 1D NMR for data analysis. Default data calculation Dynamic center was off by a factor of 2. Hence, Dc values were calculated by dividing the average value by 2.*

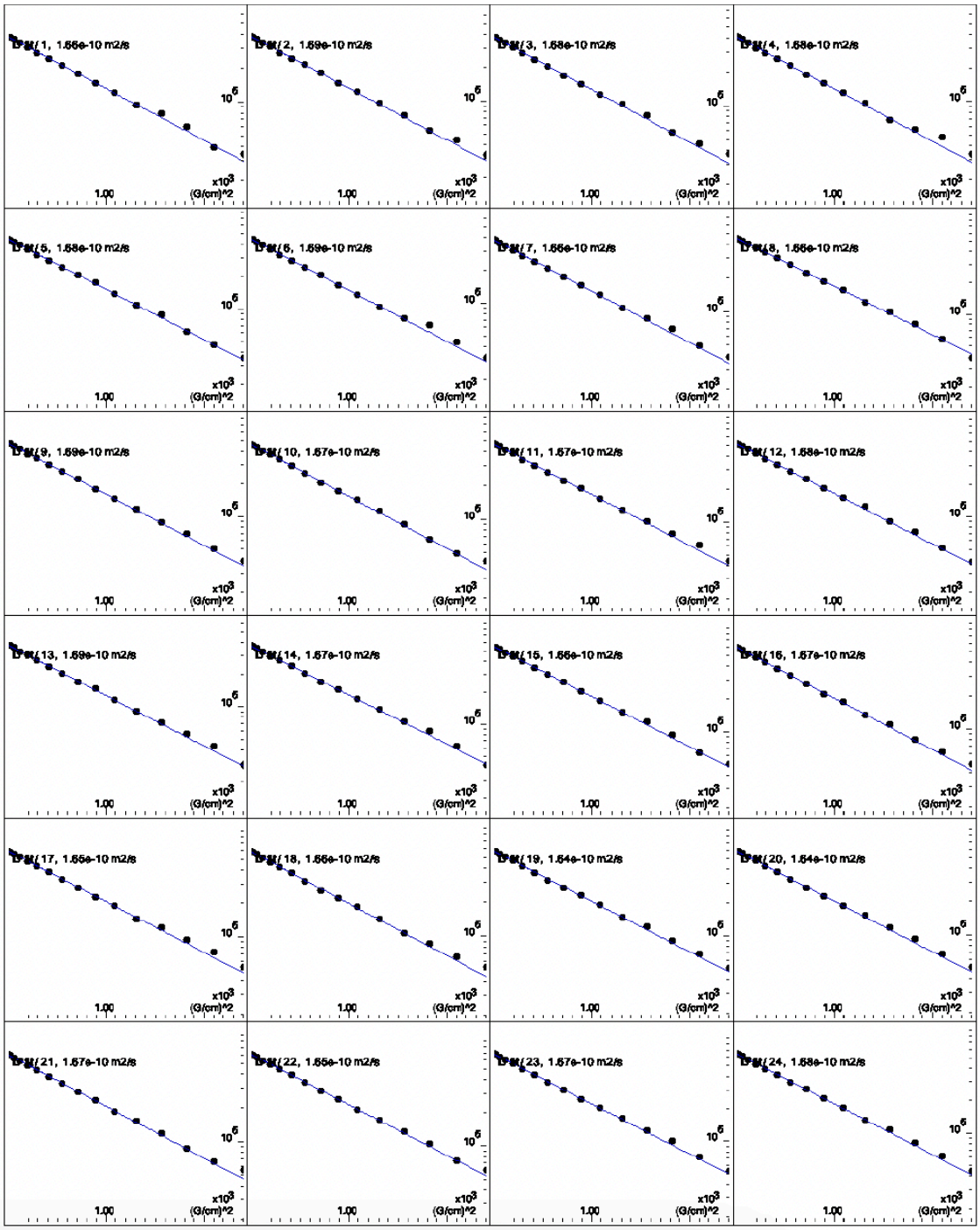
a) Reduced W_1M

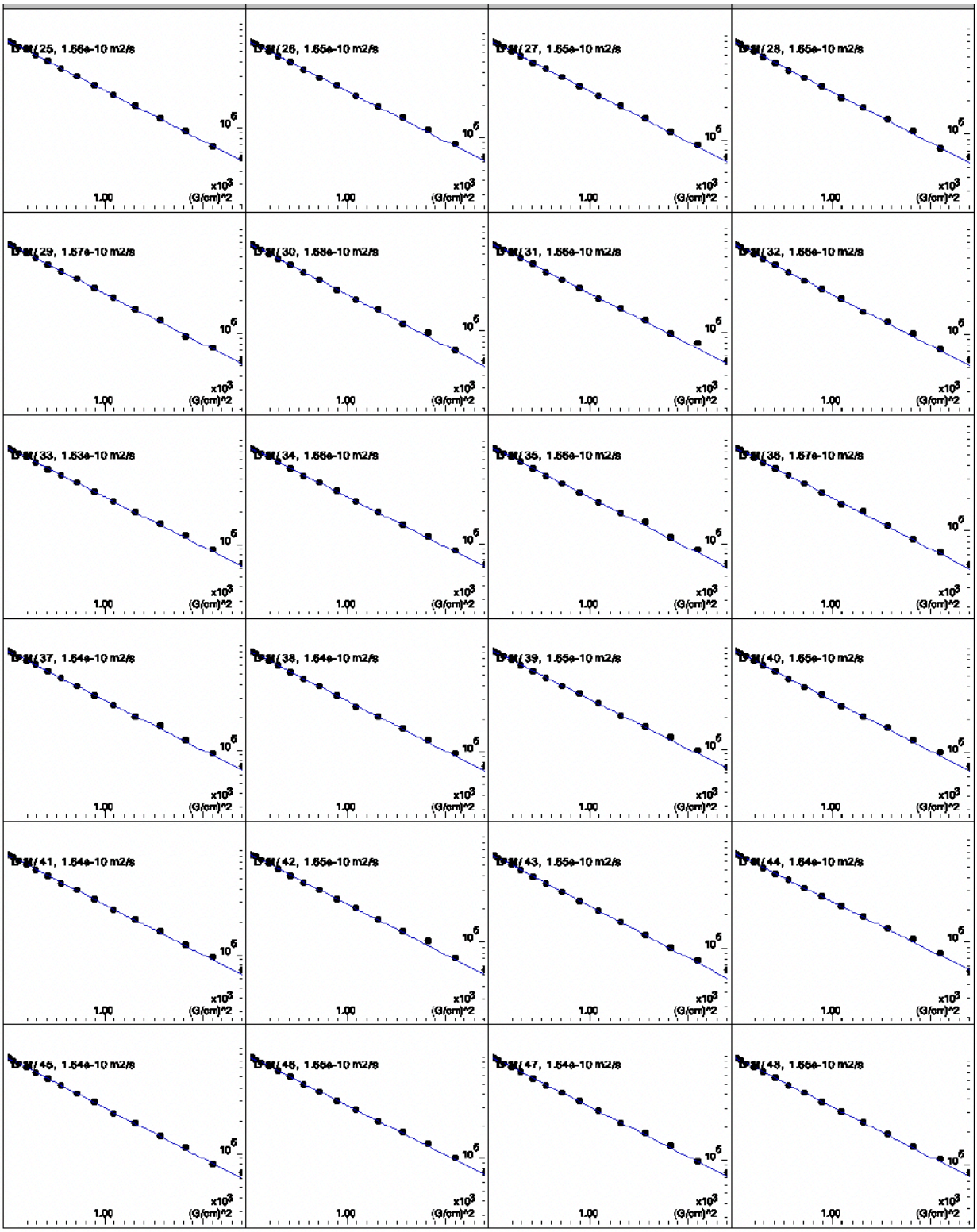
#	F2 [ppm]	D [m^2/s]	Error	fitInfo
1	8.565	1.66E-10	4.73E-12	Done
2	8.546	1.69E-10	4.26E-12	Done
3	8.538	1.68E-10	4.47E-12	Done
4	8.507	1.68E-10	5.64E-12	Done
5	8.488	1.68E-10	3.84E-12	Done
6	8.48	1.69E-10	4.97E-12	Done
7	8.456	1.66E-10	3.87E-12	Done
8	8.448	1.66E-10	2.02E-12	Done
9	8.443	1.69E-10	4.12E-12	Done
10	8.438	1.67E-10	2.97E-12	Done
11	8.43	1.67E-10	3.66E-12	Done
12	8.422	1.68E-10	4.42E-12	Done
13	8.413	1.69E-10	3.92E-12	Done
14	8.401	1.67E-10	3.24E-12	Done
15	8.393	1.66E-10	2.96E-12	Done
16	8.384	1.67E-10	3.08E-12	Done
17	8.373	1.65E-10	3.52E-12	Done
18	8.369	1.66E-10	4.16E-12	Done
19	8.363	1.64E-10	3.72E-12	Done
20	8.355	1.64E-10	3.52E-12	Done
21	8.35	1.67E-10	2.93E-12	Done
22	8.342	1.65E-10	2.51E-12	Done
23	8.335	1.67E-10	3.54E-12	Done
24	8.326	1.68E-10	4.14E-12	Done
25	8.321	1.66E-10	2.27E-12	Done
26	8.313	1.65E-10	2.50E-12	Done
27	8.304	1.65E-10	2.39E-12	Done
28	8.296	1.65E-10	2.77E-12	Done
29	8.284	1.67E-10	2.38E-12	Done
30	8.276	1.68E-10	3.21E-12	Done
31	8.253	1.66E-10	2.73E-12	Done

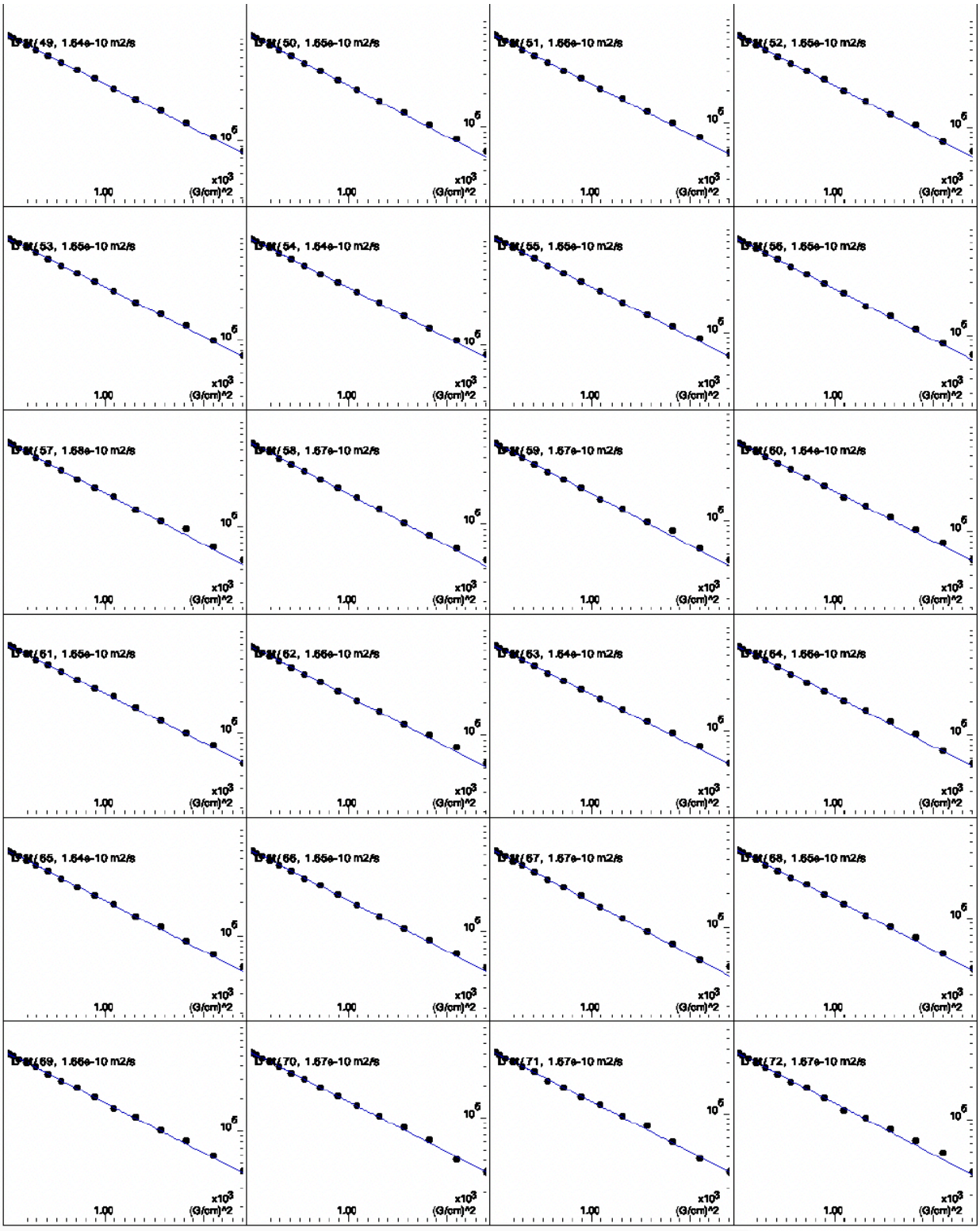
32	8.25	1.66E-10	2.87E-12	Done
33	8.244	1.63E-10	2.17E-12	Done
34	8.236	1.66E-10	2.43E-12	Done
35	8.226	1.66E-10	2.97E-12	Done
36	8.218	1.67E-10	3.51E-12	Done
37	8.194	1.64E-10	2.62E-12	Done
38	8.191	1.64E-10	2.43E-12	Done
39	8.187	1.65E-10	2.48E-12	Done
40	8.178	1.65E-10	2.23E-12	Done
41	8.167	1.64E-10	2.78E-12	Done
42	8.16	1.65E-10	3.50E-12	Done
43	8.149	1.65E-10	2.56E-12	Done
44	8.147	1.64E-10	2.65E-12	Done
45	8.142	1.64E-10	2.78E-12	Done
46	8.133	1.65E-10	1.80E-12	Done
47	8.128	1.64E-10	2.71E-12	Done
48	8.12	1.65E-10	2.17E-12	Done
49	8.111	1.64E-10	1.94E-12	Done
50	8.104	1.65E-10	2.47E-12	Done
51	8.091	1.66E-10	2.83E-12	Done
52	8.083	1.65E-10	2.47E-12	Done
53	8.073	1.65E-10	1.65E-12	Done
54	8.064	1.64E-10	1.69E-12	Done
55	8.055	1.65E-10	2.13E-12	Done
56	8.05	1.65E-10	2.27E-12	Done
57	8.039	1.68E-10	3.75E-12	Done
58	8.031	1.67E-10	3.82E-12	Done
59	8.023	1.67E-10	3.75E-12	Done
60	8.017	1.64E-10	2.84E-12	Done
61	8.012	1.65E-10	2.95E-12	Done
62	8.008	1.66E-10	3.55E-12	Done
63	8.003	1.64E-10	2.96E-12	Done
64	7.995	1.66E-10	3.04E-12	Done
65	7.988	1.64E-10	3.08E-12	Done
66	7.957	1.65E-10	2.92E-12	Done
67	7.949	1.67E-10	3.96E-12	Done
68	7.939	1.65E-10	3.07E-12	Done
69	7.931	1.66E-10	3.07E-12	Done
70	7.922	1.67E-10	3.56E-12	Done

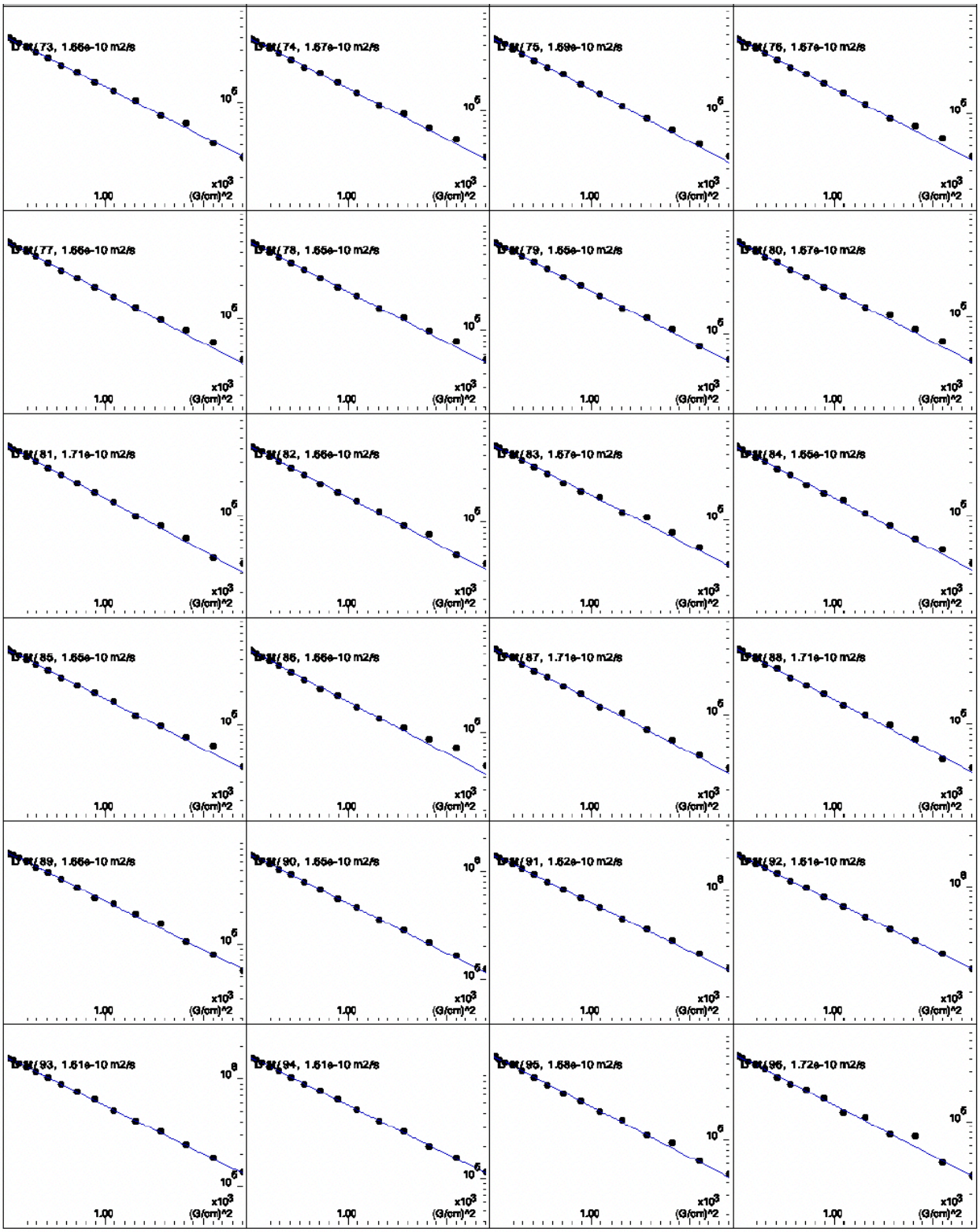
71	7.903	1.67E-10	4.04E-12	Done
72	7.874	1.67E-10	4.73E-12	Done
73	7.818	1.66E-10	3.41E-12	Done
74	7.542	1.67E-10	3.70E-12	Done
75	7.533	1.69E-10	4.13E-12	Done
76	7.526	1.67E-10	3.92E-12	Done
77	7.518	1.66E-10	3.84E-12	Done
78	7.503	1.65E-10	3.96E-12	Done
79	7.495	1.65E-10	2.06E-12	Done
80	7.472	1.67E-10	5.62E-12	Done
81	7.461	1.71E-10	4.08E-12	Done
82	7.449	1.66E-10	3.90E-12	Done
83	7.443	1.67E-10	5.94E-12	Done
84	7.43	1.65E-10	4.80E-12	Done
85	7.405	1.65E-10	4.82E-12	Done
86	7.393	1.66E-10	5.72E-12	Done
87	7.382	1.71E-10	5.23E-12	Done
88	7.365	1.71E-10	5.11E-12	Done
89	7.356	1.66E-10	5.24E-12	Done
90	7.326	1.65E-10	2.13E-12	Done
91	7.32	1.62E-10	1.62E-12	Done
92	7.317	1.61E-10	1.69E-12	Done
93	7.311	1.61E-10	2.04E-12	Done
94	7.308	1.61E-10	1.93E-12	Done
95	7.293	1.68E-10	4.57E-12	Done
96	7.286	1.72E-10	7.10E-12	Done
97	7.272	1.63E-10	2.91E-12	Done
98	7.264	1.64E-10	2.19E-12	Done
99	7.259	1.64E-10	3.45E-12	Done
100	7.229	1.63E-10	2.80E-12	Done
101	7.22	1.62E-10	1.67E-12	Done
102	7.214	1.62E-10	1.64E-12	Done
103	7.205	1.62E-10	2.49E-12	Done
104	7.174	1.63E-10	2.78E-12	Done
105	7.165	1.64E-10	2.58E-12	Done
106	7.125	1.64E-10	3.38E-12	Done
107	7.114	1.64E-10	2.27E-12	Done
108	7.075	1.62E-10	1.89E-12	Done
109	7.068	1.61E-10	1.73E-12	Done

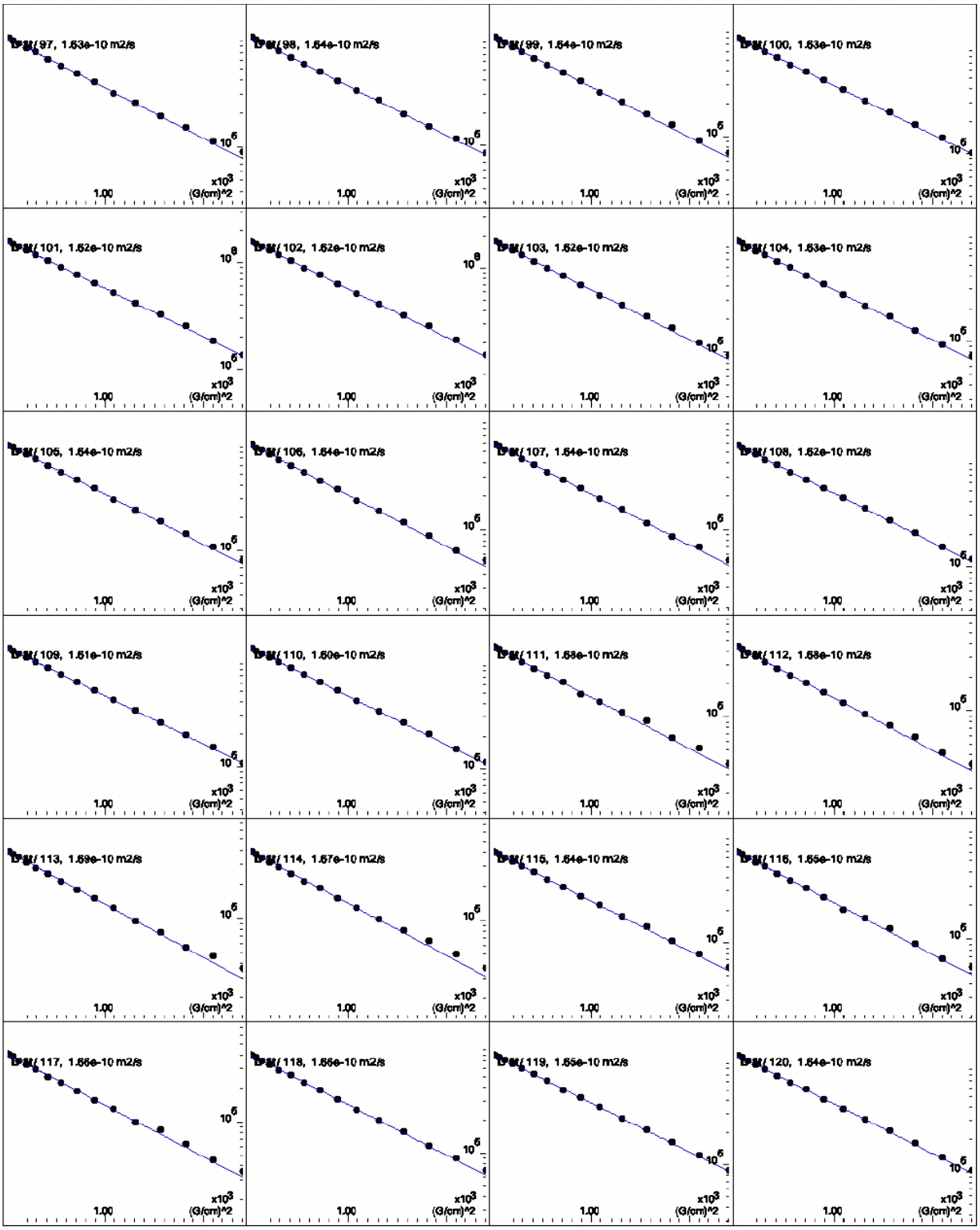
110	7.065	1.60E-10	1.69E-12	Done
111	6.977	1.68E-10	5.61E-12	Done
112	6.969	1.68E-10	4.30E-12	Done
113	6.948	1.69E-10	4.71E-12	Done
114	6.894	1.67E-10	5.82E-12	Done
115	6.857	1.64E-10	2.98E-12	Done
116	6.838	1.65E-10	3.83E-12	Done
117	6.824	1.66E-10	4.66E-12	Done
118	6.796	1.66E-10	2.22E-12	Done
119	6.786	1.65E-10	1.54E-12	Done
120	6.775	1.64E-10	2.04E-12	Done
121	6.763	1.63E-10	1.79E-12	Done
122	6.752	1.64E-10	1.66E-12	Done
123	7.981	1.65E-10	4.92E-12	Done
124	7.554	1.66E-10	6.42E-12	Done
125	6.813	1.68E-10	5.52E-12	Done
126	8.648	1.72E-10	4.71E-12	Done
127	8.362	1.63E-10	2.89E-12	Done
128	8.441	1.67E-10	3.78E-12	Done
129	7.552	1.64E-10	5.95E-12	Done
130	8.252	1.63E-10	3.15E-12	Done
131	8.237	1.65E-10	2.59E-12	Done
132	8.192	1.62E-10	1.81E-12	Done
133	7.875	1.66E-10	3.50E-12	Done
134	7.474	1.66E-10	6.39E-12	Done
135	6.941	1.70E-10	5.25E-12	Done
136	7.459	1.69E-10	5.70E-12	Done
137	8.399	1.66E-10	4.29E-12	Done
138	8.319	1.66E-10	2.92E-12	Done
139	7.863	1.68E-10	5.82E-12	Done
Average		**1.66E-10**		
Dc		**0.83 E-10**	**3.40E-12**	

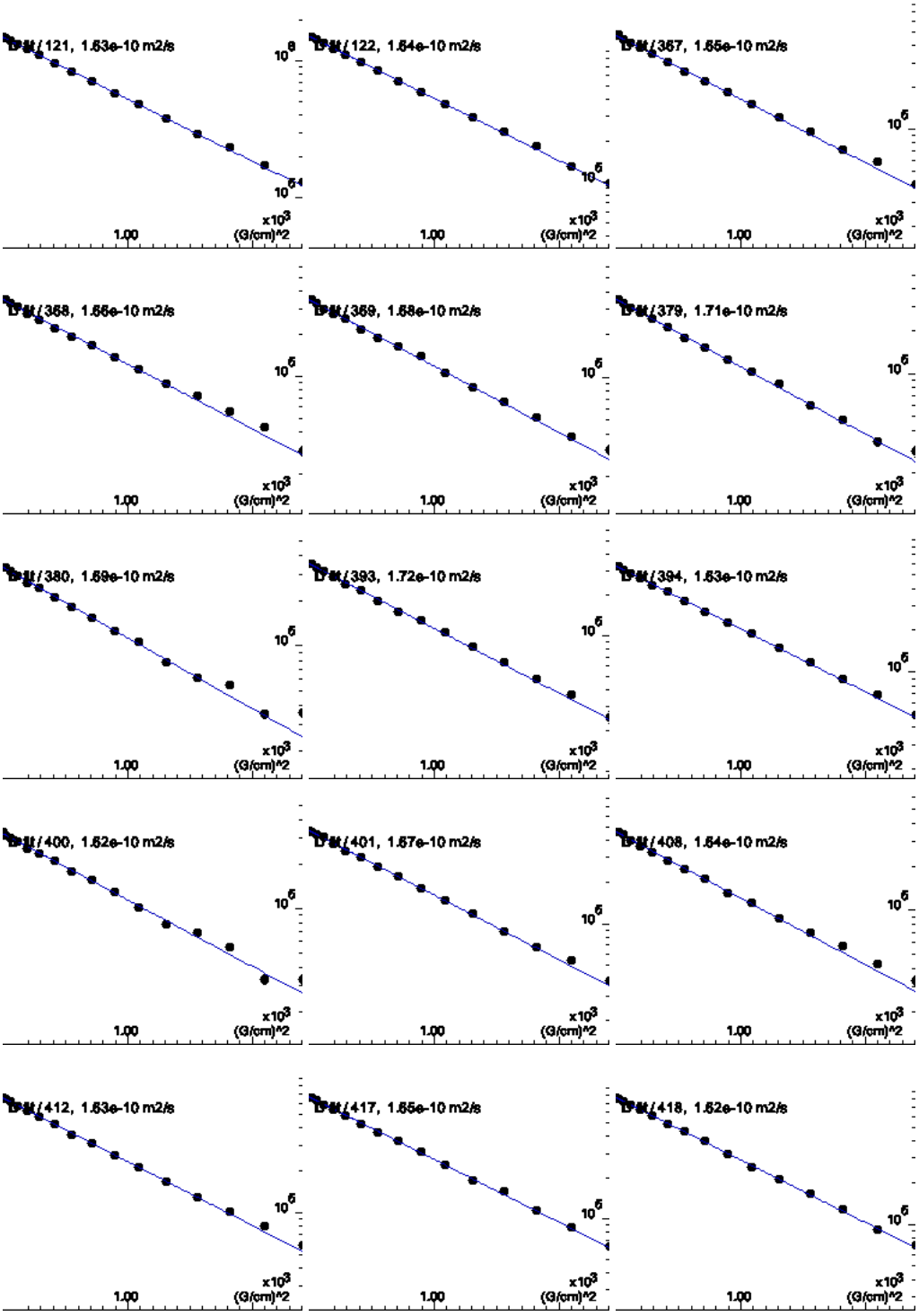

Dat(1, 1.55e-10 m2/s
Dat(2, 1.59e-10 m2/s
Dat(3, 1.58e-10 m2/s
Dat(4, 1.58e-10 m2/s
Dat(5, 1.58e-10 m2/s
Dat(6, 1.59e-10 m2/s
Dat(7, 1.56e-10 m2/s
Dat(8, 1.56e-10 m2/s
Dat(9, 1.59e-10 m2/s
Dat(10, 1.57e-10 m2/s
Dat(11, 1.57e-10 m2/s
Dat(12, 1.58e-10 m2/s
Dat(13, 1.59e-10 m2/s
Dat(14, 1.57e-10 m2/s
Dat(15, 1.56e-10 m2/s
Dat(16, 1.57e-10 m2/s
Dat(17, 1.55e-10 m2/s
Dat(18, 1.56e-10 m2/s
Dat(19, 1.54e-10 m2/s
Dat(20, 1.54e-10 m2/s
Dat(21, 1.57e-10 m2/s
Dat(22, 1.55e-10 m2/s
Dat(23, 1.57e-10 m2/s
Dat(24, 1.58e-10 m2/s
10^5
1.00
x10^3
(G/cm)^2

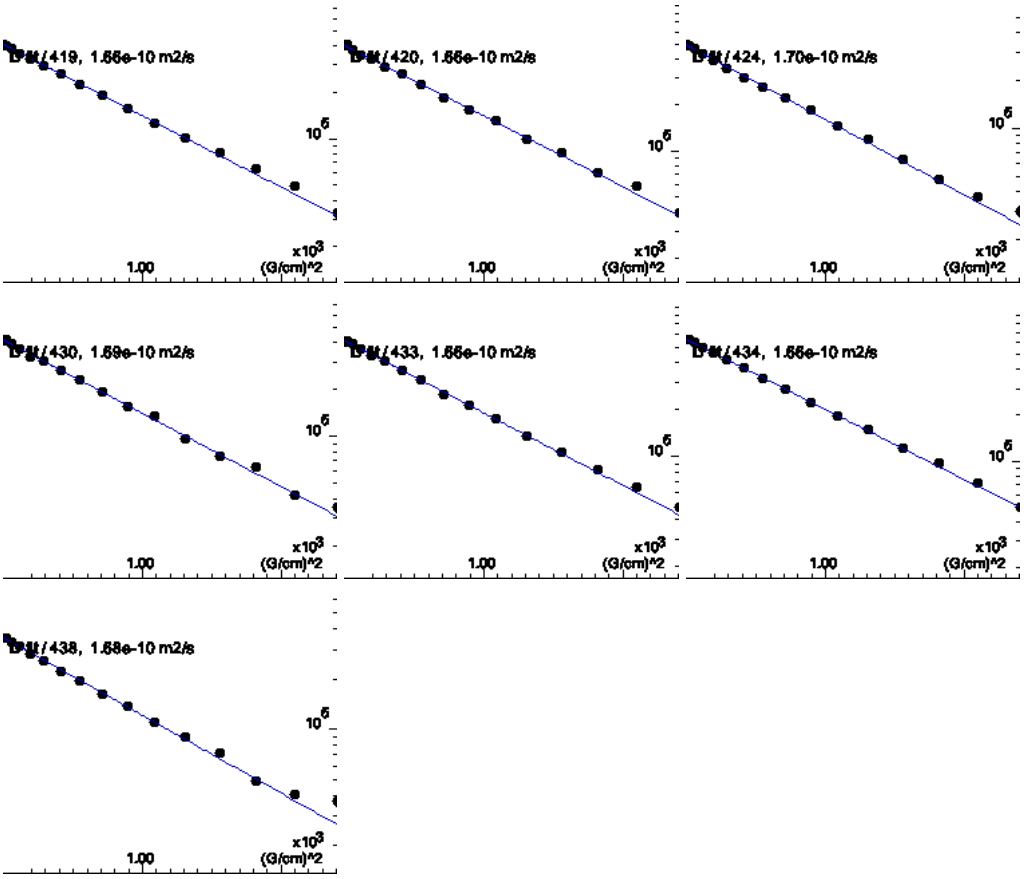

DSI(25, 1.66e-10 m2/s
DSI(26, 1.65e-10 m2/s
DSI(27, 1.65e-10 m2/s
DSI(28, 1.65e-10 m2/s
DSI(29, 1.67e-10 m2/s
DSI(30, 1.68e-10 m2/s
DSI(31, 1.66e-10 m2/s
DSI(32, 1.66e-10 m2/s
DSI(33, 1.63e-10 m2/s
DSI(34, 1.66e-10 m2/s
DSI(35, 1.66e-10 m2/s
DSI(36, 1.67e-10 m2/s
DSI(37, 1.64e-10 m2/s
DSI(38, 1.64e-10 m2/s
DSI(39, 1.65e-10 m2/s
DSI(40, 1.65e-10 m2/s
DSI(41, 1.64e-10 m2/s
DSI(42, 1.65e-10 m2/s
DSI(43, 1.65e-10 m2/s
DSI(44, 1.64e-10 m2/s
DSI(45, 1.64e-10 m2/s
DSI(46, 1.65e-10 m2/s
DSI(47, 1.64e-10 m2/s
DSI(48, 1.65e-10 m2/s
10^5
x10^3
1.00
(G/cm)^2


Dat(49, 1.64e-10 m2/s
Dat(50, 1.65e-10 m2/s
Dat(51, 1.66e-10 m2/s
Dat(52, 1.65e-10 m2/s
Dat(53, 1.65e-10 m2/s
Dat(54, 1.64e-10 m2/s
Dat(55, 1.65e-10 m2/s
Dat(56, 1.65e-10 m2/s
Dat(57, 1.68e-10 m2/s
Dat(58, 1.67e-10 m2/s
Dat(59, 1.67e-10 m2/s
Dat(60, 1.64e-10 m2/s
Dat(61, 1.65e-10 m2/s
Dat(62, 1.66e-10 m2/s
Dat(63, 1.64e-10 m2/s
Dat(64, 1.66e-10 m2/s
Dat(65, 1.64e-10 m2/s
Dat(66, 1.65e-10 m2/s
Dat(67, 1.67e-10 m2/s
Dat(68, 1.65e-10 m2/s
Dat(69, 1.66e-10 m2/s
Dat(70, 1.67e-10 m2/s
Dat(71, 1.67e-10 m2/s
Dat(72, 1.67e-10 m2/s
10^5
x10^3
1.00
(G/cm)^2


Dav(73, 1.66e-10 m2/s
Dav(74, 1.67e-10 m2/s
Dav(75, 1.69e-10 m2/s
Dav(76, 1.67e-10 m2/s
Dav(77, 1.66e-10 m2/s
Dav(78, 1.65e-10 m2/s
Dav(79, 1.65e-10 m2/s
Dav(80, 1.67e-10 m2/s
Dav(81, 1.71e-10 m2/s
Dav(82, 1.66e-10 m2/s
Dav(83, 1.67e-10 m2/s
Dav(84, 1.65e-10 m2/s
Dav(85, 1.65e-10 m2/s
Dav(86, 1.66e-10 m2/s
Dav(87, 1.71e-10 m2/s
Dav(88, 1.71e-10 m2/s
Dav(89, 1.66e-10 m2/s
Dav(90, 1.65e-10 m2/s
Dav(91, 1.52e-10 m2/s
Dav(92, 1.51e-10 m2/s
Dav(93, 1.51e-10 m2/s
Dav(94, 1.51e-10 m2/s
Dav(95, 1.58e-10 m2/s
Dav(96, 1.72e-10 m2/s
1.00
x10^3
(G/cm)^2


DSE(97, 1.53e-10 m2/s
DSE(98, 1.54e-10 m2/s
DSE(99, 1.54e-10 m2/s
DSE(100, 1.53e-10 m2/s
DSE(101, 1.52e-10 m2/s
DSE(102, 1.52e-10 m2/s
DSE(103, 1.52e-10 m2/s
DSE(104, 1.53e-10 m2/s
DSE(105, 1.54e-10 m2/s
DSE(106, 1.54e-10 m2/s
DSE(107, 1.54e-10 m2/s
DSE(108, 1.52e-10 m2/s
DSE(109, 1.51e-10 m2/s
DSE(110, 1.50e-10 m2/s
DSE(111, 1.58e-10 m2/s
DSE(112, 1.58e-10 m2/s
DSE(113, 1.59e-10 m2/s
DSE(114, 1.57e-10 m2/s
DSE(115, 1.54e-10 m2/s
DSE(116, 1.55e-10 m2/s
DSE(117, 1.56e-10 m2/s
DSE(118, 1.56e-10 m2/s
DSE(119, 1.55e-10 m2/s
DSE(120, 1.54e-10 m2/s
1.00
x10^3
(G/cm)^2


DW (121, 1.53e-10 m2/s
DW (122, 1.54e-10 m2/s
DW (367, 1.55e-10 m2/s
DW (368, 1.56e-10 m2/s
DW (369, 1.58e-10 m2/s
DW (379, 1.71e-10 m2/s
DW (380, 1.59e-10 m2/s
DW (393, 1.72e-10 m2/s
DW (394, 1.53e-10 m2/s
DW (400, 1.52e-10 m2/s
DW (401, 1.57e-10 m2/s
DW (408, 1.54e-10 m2/s
DW (412, 1.53e-10 m2/s
DW (417, 1.55e-10 m2/s
DW (418, 1.52e-10 m2/s
1.00
x10^3
(G/cm)^2



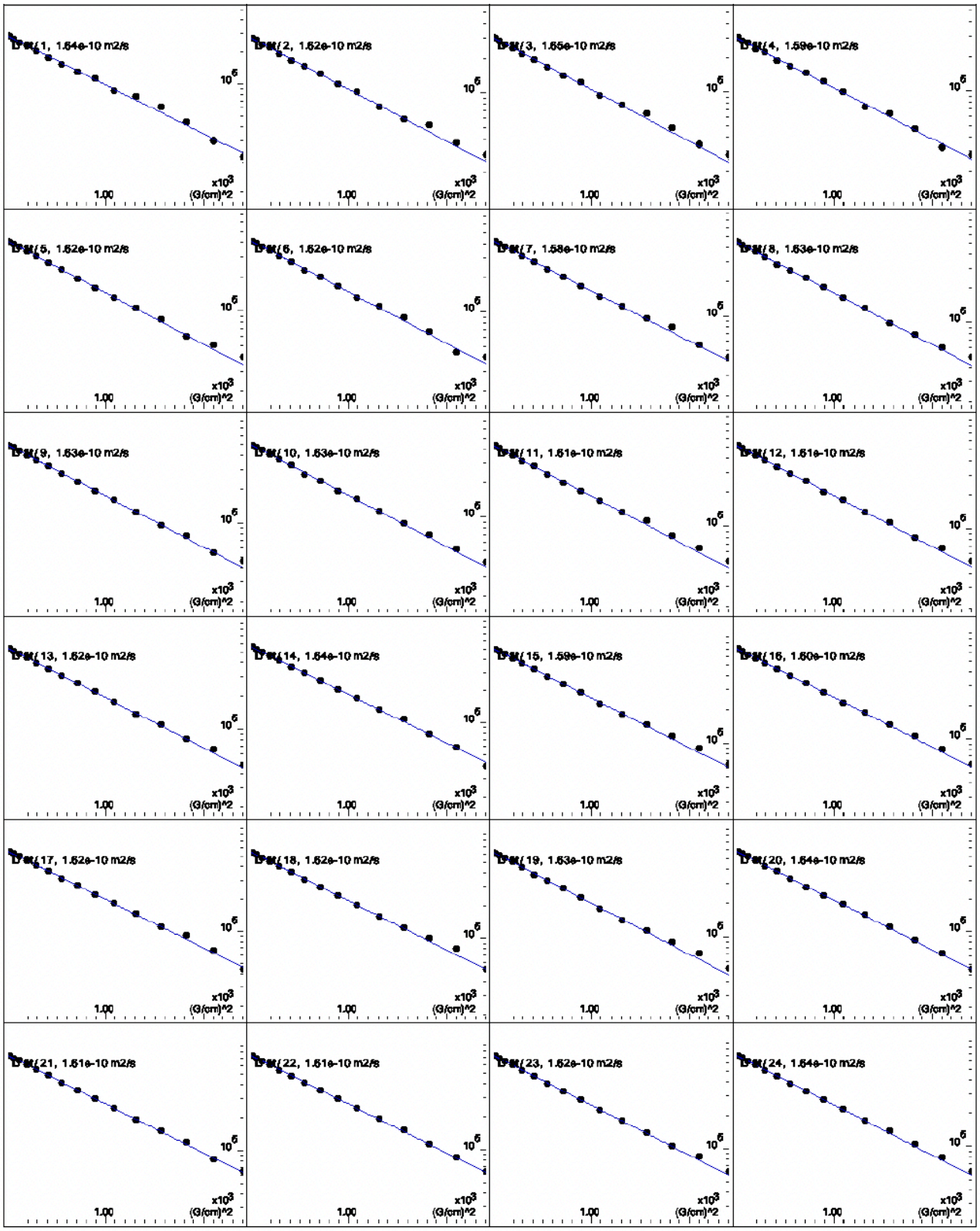
b) Stapled W_1M

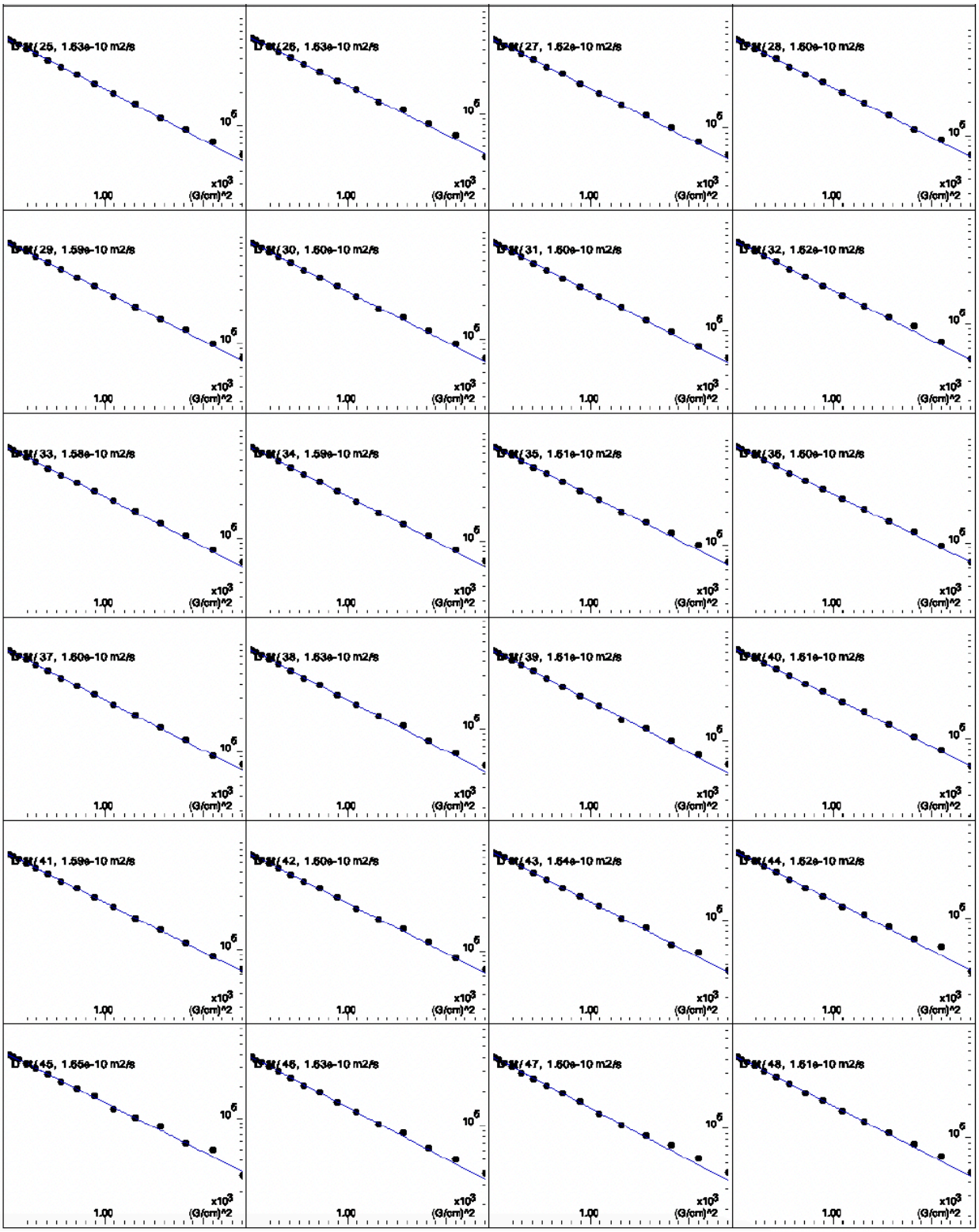
#	F2 [ppm]	D [m2/s]	error	fitInfo
1	8.444	1.64E-10	5.45E-12	Done
2	8.438	1.62E-10	5.79E-12	Done
3	8.431	1.65E-10	5.23E-12	Done
4	8.426	1.59E-10	5.84E-12	Done
5	8.396	1.62E-10	2.94E-12	Done
6	8.389	1.62E-10	3.78E-12	Done
7	8.38	1.58E-10	3.59E-12	Done
8	8.369	1.63E-10	3.61E-12	Done
9	8.359	1.63E-10	2.48E-12	Done
10	8.351	1.63E-10	3.13E-12	Done
11	8.343	1.61E-10	4.16E-12	Done
12	8.337	1.61E-10	3.26E-12	Done
13	8.33	1.62E-10	2.06E-12	Done
14	8.323	1.64E-10	2.72E-12	Done
15	8.308	1.59E-10	2.52E-12	Done

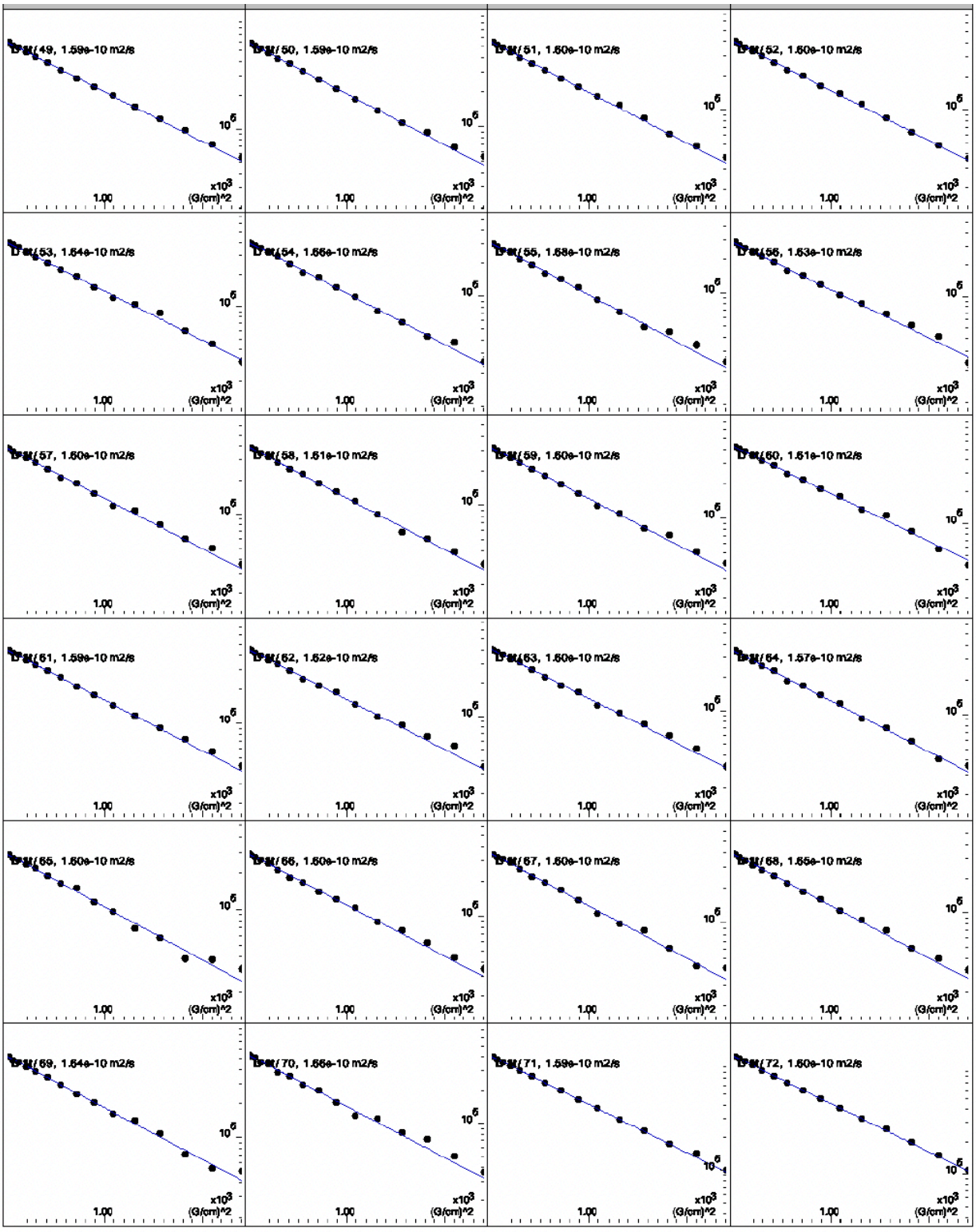
16	8.301	1.60E-10	3.17E-12	Done
17	8.284	1.62E-10	3.23E-12	Done
18	8.277	1.62E-10	3.15E-12	Done
19	8.271	1.63E-10	3.28E-12	Done
20	8.263	1.64E-10	3.00E-12	Done
21	8.246	1.61E-10	2.14E-12	Done
22	8.241	1.61E-10	2.13E-12	Done
23	8.234	1.62E-10	2.93E-12	Done
24	8.226	1.64E-10	2.86E-12	Done
25	8.219	1.63E-10	3.14E-12	Done
26	8.213	1.63E-10	3.35E-12	Done
27	8.2	1.62E-10	2.57E-12	Done
28	8.19	1.60E-10	2.24E-12	Done
29	8.184	1.59E-10	1.86E-12	Done
30	8.176	1.60E-10	2.40E-12	Done
31	8.167	1.60E-10	2.74E-12	Done
32	8.159	1.62E-10	3.91E-12	Done
33	8.146	1.58E-10	3.92E-12	Done
34	8.139	1.59E-10	2.31E-12	Done
35	8.13	1.61E-10	2.80E-12	Done
36	8.126	1.60E-10	1.48E-12	Done
37	8.121	1.60E-10	2.22E-12	Done
38	8.095	1.63E-10	3.06E-12	Done
39	8.087	1.61E-10	3.02E-12	Done
40	8.073	1.61E-10	1.51E-12	Done
41	8.066	1.59E-10	1.58E-12	Done
42	8.063	1.60E-10	2.15E-12	Done
43	8.048	1.64E-10	2.73E-12	Done
44	8.041	1.62E-10	4.10E-12	Done
45	8.034	1.65E-10	4.26E-12	Done
46	8.027	1.63E-10	4.49E-12	Done
47	8.022	1.60E-10	5.10E-12	Done
48	8.015	1.61E-10	3.17E-12	Done
49	8.005	1.59E-10	3.05E-12	Done
50	7.997	1.59E-10	3.87E-12	Done
51	7.956	1.60E-10	4.25E-12	Done
52	7.947	1.60E-10	4.13E-12	Done
53	7.937	1.64E-10	4.97E-12	Done
54	7.871	1.66E-10	4.61E-12	Done

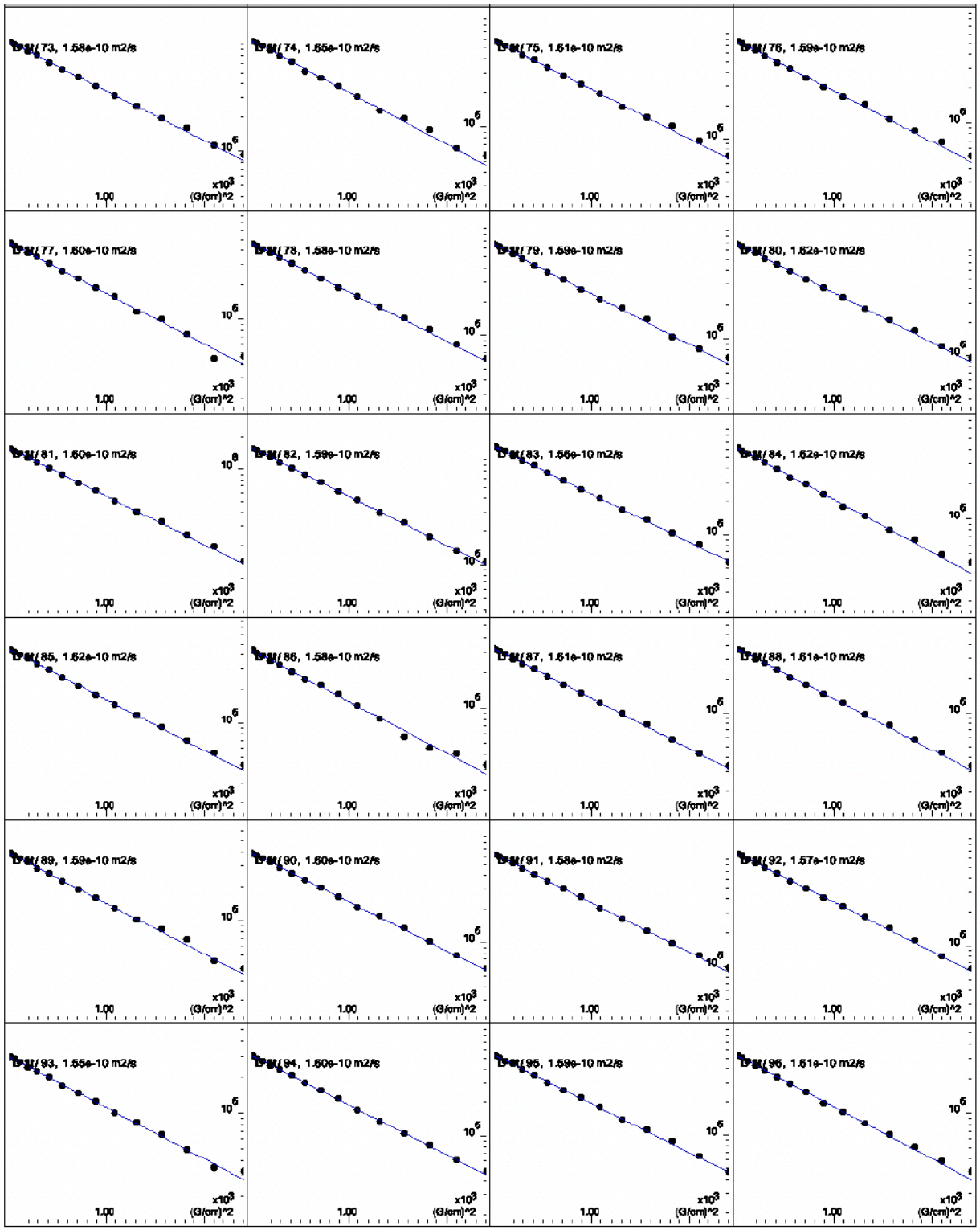
55	7.863	1.68E-10	5.49E-12	Done
56	7.555	1.63E-10	6.32E-12	Done
57	7.543	1.60E-10	4.40E-12	Done
58	7.532	1.61E-10	4.62E-12	Done
59	7.516	1.60E-10	4.41E-12	Done
60	7.506	1.61E-10	3.95E-12	Done
61	7.474	1.59E-10	3.80E-12	Done
62	7.461	1.62E-10	5.11E-12	Done
63	7.447	1.60E-10	4.45E-12	Done
64	7.431	1.57E-10	6.01E-12	Done
65	7.422	1.60E-10	7.10E-12	Done
66	7.405	1.60E-10	5.51E-12	Done
67	7.393	1.60E-10	4.84E-12	Done
68	7.368	1.65E-10	6.11E-12	Done
69	7.359	1.64E-10	5.11E-12	Done
70	7.333	1.66E-10	7.93E-12	Done
71	7.311	1.59E-10	1.99E-12	Done
72	7.307	1.60E-10	2.00E-12	Done
73	7.297	1.58E-10	3.20E-12	Done
74	7.286	1.65E-10	4.70E-12	Done
75	7.276	1.61E-10	3.00E-12	Done
76	7.267	1.59E-10	4.03E-12	Done
77	7.257	1.60E-10	3.82E-12	Done
78	7.248	1.58E-10	2.24E-12	Done
79	7.244	1.59E-10	3.14E-12	Done
80	7.226	1.62E-10	2.18E-12	Done
81	7.216	1.60E-10	1.87E-12	Done
82	7.207	1.59E-10	1.70E-12	Done
83	7.199	1.56E-10	2.08E-12	Done
84	7.183	1.62E-10	4.79E-12	Done
85	7.174	1.62E-10	2.86E-12	Done
86	7.15	1.58E-10	6.35E-12	Done
87	7.142	1.61E-10	3.74E-12	Done
88	7.12	1.61E-10	3.98E-12	Done
89	7.11	1.59E-10	4.71E-12	Done
90	7.074	1.60E-10	1.92E-12	Done
91	7.065	1.58E-10	1.74E-12	Done
92	7.057	1.57E-10	2.32E-12	Done
93	7.017	1.55E-10	5.25E-12	Done

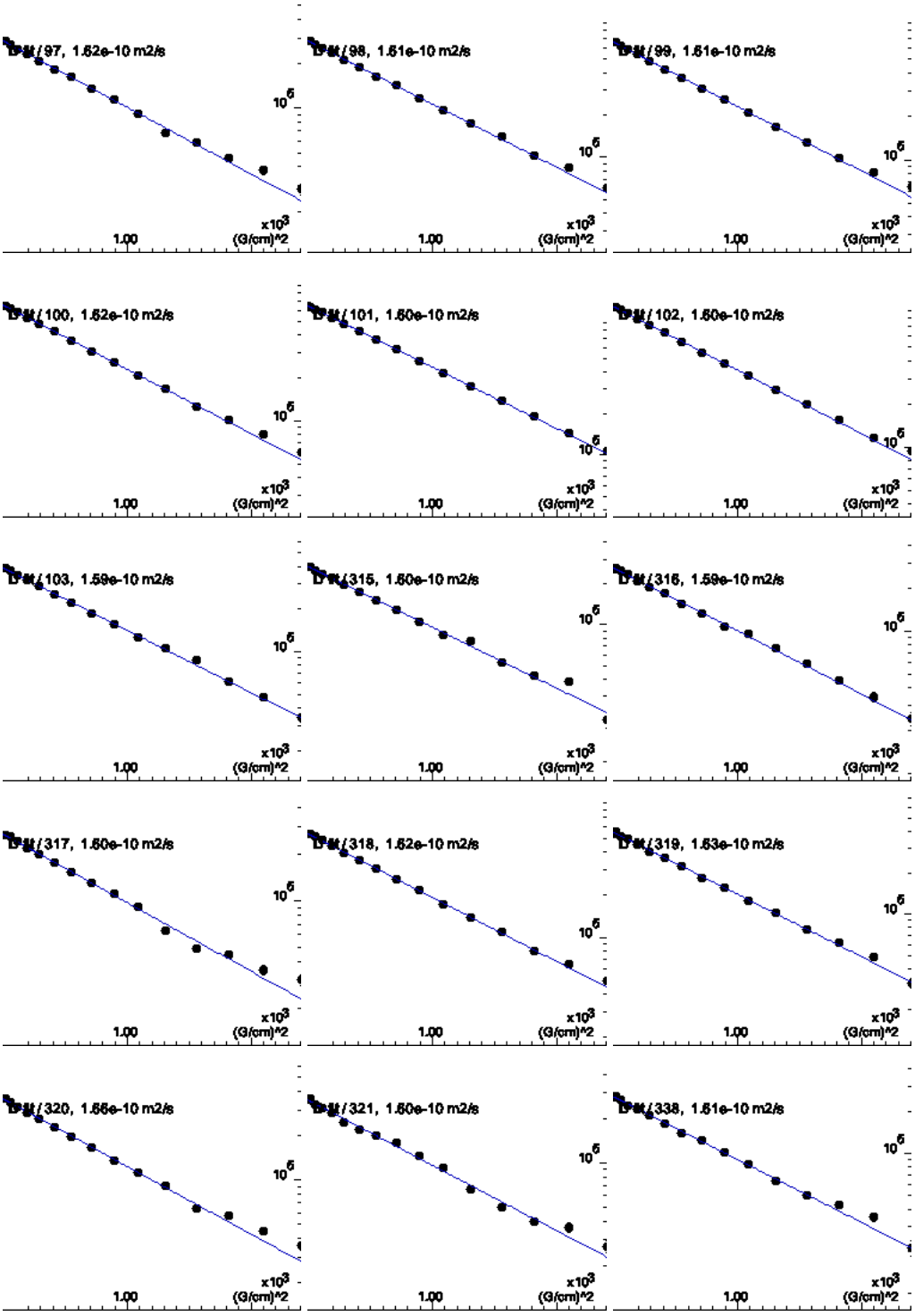
94	6.851	1.60E-10	2.97E-12	Done
95	6.836	1.59E-10	2.83E-12	Done
96	6.818	1.61E-10	3.84E-12	Done
97	6.807	1.62E-10	4.75E-12	Done
98	6.792	1.61E-10	2.98E-12	Done
99	6.782	1.61E-10	3.50E-12	Done
100	6.769	1.62E-10	2.58E-12	Done
101	6.759	1.60E-10	9.81E-13	Done
102	6.748	1.60E-10	2.33E-12	Done
103	6.724	1.59E-10	3.42E-12	Done
104	8.458	1.60E-10	5.07E-12	Done
105	8.454	1.59E-10	5.13E-12	Done
106	8.412	1.60E-10	6.40E-12	Done
107	8.318	1.62E-10	2.44E-12	Done
108	8.037	1.63E-10	4.52E-12	Done
109	7.452	1.66E-10	4.96E-12	Done
110	7.414	1.60E-10	9.27E-12	Done
111	8.416	1.61E-10	5.74E-12	Done
112	8.202	1.62E-10	3.83E-12	Done
113	7.903	1.59E-10	4.95E-12	Done
114	7.805	1.61E-10	6.48E-12	Done
115	7.491	1.63E-10	3.64E-12	Done
Average		1.61E-10		
Dc		**0.805 E-10**	**3.73E-12**	

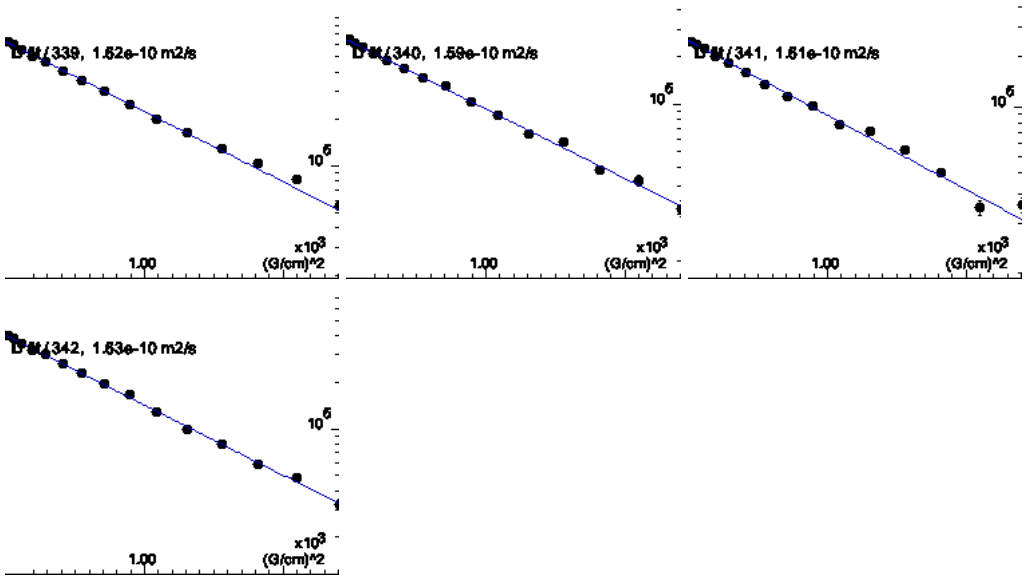

Dat(1, 1.64e-10 m2/s
Dat(2, 1.62e-10 m2/s
Dat(3, 1.65e-10 m2/s
Dat(4, 1.59e-10 m2/s
Dat(5, 1.62e-10 m2/s
Dat(6, 1.62e-10 m2/s
Dat(7, 1.58e-10 m2/s
Dat(8, 1.63e-10 m2/s
Dat(9, 1.63e-10 m2/s
Dat(10, 1.63e-10 m2/s
Dat(11, 1.61e-10 m2/s
Dat(12, 1.61e-10 m2/s
Dat(13, 1.62e-10 m2/s
Dat(14, 1.64e-10 m2/s
Dat(15, 1.59e-10 m2/s
Dat(16, 1.60e-10 m2/s
Dat(17, 1.62e-10 m2/s
Dat(18, 1.62e-10 m2/s
Dat(19, 1.63e-10 m2/s
Dat(20, 1.64e-10 m2/s
Dat(21, 1.61e-10 m2/s
Dat(22, 1.61e-10 m2/s
Dat(23, 1.62e-10 m2/s
Dat(24, 1.64e-10 m2/s
10^5
1.00
x10^3
(G/cm)^2


Dat(25, 1.53e-10 m2/s)
Dat(26, 1.53e-10 m2/s)
Dat(27, 1.52e-10 m2/s)
Dat(28, 1.50e-10 m2/s)
Dat(29, 1.59e-10 m2/s)
Dat(30, 1.60e-10 m2/s)
Dat(31, 1.60e-10 m2/s)
Dat(32, 1.62e-10 m2/s)
Dat(33, 1.58e-10 m2/s)
Dat(34, 1.59e-10 m2/s)
Dat(35, 1.61e-10 m2/s)
Dat(36, 1.60e-10 m2/s)
Dat(37, 1.60e-10 m2/s)
Dat(38, 1.63e-10 m2/s)
Dat(39, 1.61e-10 m2/s)
Dat(40, 1.61e-10 m2/s)
Dat(41, 1.59e-10 m2/s)
Dat(42, 1.60e-10 m2/s)
Dat(43, 1.64e-10 m2/s)
Dat(44, 1.62e-10 m2/s)
Dat(45, 1.55e-10 m2/s)
Dat(46, 1.53e-10 m2/s)
Dat(47, 1.60e-10 m2/s)
Dat(48, 1.61e-10 m2/s)
10^6
1.00
x10^3
(G/cm)^2


Dat(49, 1.59e-10 m2/s
Dat(50, 1.59e-10 m2/s
Dat(51, 1.60e-10 m2/s
Dat(52, 1.60e-10 m2/s
Dat(53, 1.64e-10 m2/s
Dat(54, 1.66e-10 m2/s
Dat(55, 1.58e-10 m2/s
Dat(56, 1.63e-10 m2/s
Dat(57, 1.60e-10 m2/s
Dat(58, 1.61e-10 m2/s
Dat(59, 1.60e-10 m2/s
Dat(60, 1.61e-10 m2/s
Dat(61, 1.59e-10 m2/s
Dat(62, 1.62e-10 m2/s
Dat(63, 1.60e-10 m2/s
Dat(64, 1.57e-10 m2/s
Dat(65, 1.60e-10 m2/s
Dat(66, 1.60e-10 m2/s
Dat(67, 1.60e-10 m2/s
Dat(68, 1.65e-10 m2/s
Dat(69, 1.64e-10 m2/s
Dat(70, 1.66e-10 m2/s
Dat(71, 1.59e-10 m2/s
Dat(72, 1.60e-10 m2/s
10^6
1.00
x10^3
(G/cm)^2


Dat(73, 1.58e-10 m2/s)
Dat(74, 1.55e-10 m2/s)
Dat(75, 1.61e-10 m2/s)
Dat(76, 1.59e-10 m2/s)
Dat(77, 1.60e-10 m2/s)
Dat(78, 1.58e-10 m2/s)
Dat(79, 1.59e-10 m2/s)
Dat(80, 1.62e-10 m2/s)
Dat(81, 1.60e-10 m2/s)
Dat(82, 1.59e-10 m2/s)
Dat(83, 1.56e-10 m2/s)
Dat(84, 1.62e-10 m2/s)
Dat(85, 1.62e-10 m2/s)
Dat(86, 1.58e-10 m2/s)
Dat(87, 1.61e-10 m2/s)
Dat(88, 1.61e-10 m2/s)
Dat(89, 1.59e-10 m2/s)
Dat(90, 1.60e-10 m2/s)
Dat(91, 1.58e-10 m2/s)
Dat(92, 1.57e-10 m2/s)
Dat(93, 1.55e-10 m2/s)
Dat(94, 1.60e-10 m2/s)
Dat(95, 1.59e-10 m2/s)
Dat(96, 1.61e-10 m2/s)
1.00
x10^3
(G/cm)^2


Dav(97, 1.62e-10 m2/s
Dav(98, 1.61e-10 m2/s
Dav(99, 1.61e-10 m2/s
Dav(100, 1.62e-10 m2/s
Dav(101, 1.60e-10 m2/s
Dav(102, 1.60e-10 m2/s
Dav(103, 1.59e-10 m2/s
Dav(315, 1.60e-10 m2/s
Dav(316, 1.59e-10 m2/s
Dav(317, 1.60e-10 m2/s
Dav(318, 1.62e-10 m2/s
Dav(319, 1.63e-10 m2/s
Dav(320, 1.66e-10 m2/s
Dav(321, 1.60e-10 m2/s
Dav(338, 1.61e-10 m2/s
10^5
x10^3
1.00
(G/cm)^2


D#(339, 1.52e-10 m2/s
D#(340, 1.59e-10 m2/s
D#(341, 1.51e-10 m2/s
D#(342, 1.53e-10 m2/s
10^5
1.00
x10^3
(G/cm)^2

***APPENDIX D**: Chemical shift assignments (ppm) for mutant W units aciniform protein.*

a) Reduced W_1M

	H	N	C	Cα	Cβ
1 Ala			174.11	51.94	19.32
2 Gly	8.51	108.51	171.75	44.55	
3 Pro			177.37	63.27	32.08
4 Gln	8.57	120.88	176.56	56.04	29.43
5 Gly	8.37	109.99	174.42	45.33	
6 Gly	8.15	108.31	174.05	45.12	
7 Phe	8.20	119.80	176.50	58.12	39.49
8 Gly	8.40	110.34	173.98	45.34	
9 Ala	8.14	123.77	178.21	52.71	19.30
10 Thr	8.12	112.20	175.31	61.93	69.81
11 Gly	8.34	110.78	174.78	45.50	
12 Gly	8.28	108.65	174.53	45.41	
13 Ala	8.33	123.94	178.29	53.05	19.35
14 Ser	8.20	114.06	174.85	58.84	63.52
15 Ala	8.16	124.68	180.02	55.03	18.36
16 Gly	8.38	106.04	176.13	46.59	
17 Leu	7.53	122.96	177.86	57.65	41.95
18 Ile	7.52	118.40	180.34	64.78	37.87
19 Ser	8.05	115.39	175.71	61.65	62.81
20 Arg	7.86	123.13	178.68	59.99	30.71
21 Val	7.80	118.21	177.19	66.93	31.80
22 Ala	8.28	121.02	178.99	55.73	18.52
23 Asn	8.60	116.11	177.25	55.94	38.30
24 Ala	7.42	121.12	179.74	54.82	17.77
25 Leu	7.82	115.01	179.07	57.05	43.06
26 Ala	8.84	121.12	176.79	54.49	18.32
27 Asn	7.01	113.35	175.54	52.68	39.12
28 Thr	7.35	109.29	174.07	60.68	70.39
29 Cys	10.35	126.83	179.49	62.55	26.59
30 Thr	10.67	119.41	175.62	68.66	
31 Leu	7.52	118.42	178.25	60.00	39.80
32 Arg	7.61	114.54	177.81	59.06	30.67
33 Thr	7.87	111.76	174.64	65.53	69.42
34 Val	7.41	118.42	175.04	65.51	33.87
35 Leu	8.13	119.68	172.93	51.99	39.40
36 Arg	6.24	116.40	174.59	53.82	31.73
37 Thr	8.31	117.22	175.74	65.12	69.16
38 Gly	8.93	115.10	174.86	45.05	
39 Val	8.08	122.83	174.72	62.10	32.23
40 Ser	8.42	123.12	175.21	57.89	64.52
41 Gln	9.06	123.80	178.85	59.42	28.98
42 Gln	8.57	118.80	178.92	59.45	28.19
43 Ile	7.47	121.29	177.54	63.56	36.41
44 Ala	8.51	121.44	179.26	55.85	18.51
45 Ser	8.32	110.13	176.30	62.46	63.02
46 Ser	7.64	116.98	176.71	62.34	63.48
47 Val	8.65	120.58	177.50	66.91	31.41
48 Val	8.18	116.30	177.01	66.91	30.89
49 Gln	8.25	119.62	177.85	60.31	29.12
50 Arg	8.04	117.33	179.73	59.03	30.07

51	Ala	9.01	122.08	179.41	55.50	18.81
52	Ala	8.53	119.27	178.96	55.55	17.80
53	Gln	7.93	116.07	179.37	59.30	28.02
54	Ser	8.31	116.20	176.96	61.76	62.84
55	Leu	8.48	123.23	178.08	58.39	41.50
56	Ala	8.17	121.81	179.64	55.55	18.56
57	Ser	8.00	112.89	177.31	61.33	62.88
58	Thr	8.07	117.26	175.77	65.83	68.96
59	Leu	7.88	118.18	177.28	55.01	43.50
60	Gly	8.02	108.83	175.11	46.58	
61	Val	7.56	111.91	174.92	58.97	33.93
62	Asp	8.35	119.76	177.93	54.89	41.30
63	Gly	8.86	115.72	174.90	46.89	
64	Asn	8.37	120.53	177.91	56.59	37.97
65	Asn	8.43	119.39	176.72	56.10	38.35
66	Leu	7.96	119.04	177.88	58.68	41.30
67	Ala	8.53	118.94	178.70	55.76	18.30
68	Arg	7.77	117.56	179.41	59.93	29.65
69	Phe	8.04	116.78	179.20	61.71	39.15
70	Ala	8.73	123.02	178.29	55.85	17.67
71	Val	8.75	117.78	179.30	66.82	31.54
72	Gln	8.13	120.61	178.61	59.34	28.34
73	Ala	7.84	119.92	180.85	55.16	18.81
74	Val	8.07	114.96	178.37	65.12	31.67
75	Ser	8.36	117.08	174.53	61.32	63.04
76	Arg	6.96	117.46	176.52	55.77	30.70
77	Leu	7.32	121.77	175.40	52.93	40.37
78	Pro			177.58	61.91	32.43
79	Ala	8.76	125.87	178.88	53.56	18.09
80	Gly	8.94	110.51	175.06	45.62	
81	Ser	7.54	116.99	171.58	61.47	64.07
82	Asp	8.30	121.11	176.68	52.07	43.56
83	Thr	8.27	115.01	176.36	67.25	69.58
84	Ser	8.62	115.97	175.98	62.57	61.92
85	Ala	7.81	125.85	180.93	54.97	18.54
86	Tyr	7.62	117.78	178.24	62.92	38.81
87	Ala	8.86	121.85	179.56	55.79	18.95
88	Gln	8.23	119.10	177.90	59.34	27.41
89	Ala	7.94	122.98	179.63	55.19	18.37
90	Phe	8.42	117.32	178.45	60.41	38.99
91	Ser	8.40	111.56	176.88	61.88	63.76
92	Ser	8.73	112.92	176.56	62.57	63.73
93	Ala	7.94	123.64	178.63	55.89	17.74
94	Leu	8.06	116.40	179.35	58.45	42.42
95	Phe	7.77	117.20	180.20	60.68	38.96
96	Asn	9.24	122.23	176.70	56.05	37.47
97	Ala	8.10	117.32	177.45	52.25	19.58
98	Gly	7.52	105.11	174.23	45.53	
99	Val	7.62	119.03	175.94	64.84	31.99
100	Leu	6.46	113.05	174.47	51.57	45.56

101	Asn	8.66	114.52	173.94	52.19	40.35
102	Ala	8.88	118.80	178.93	55.16	18.65
103	Ser	8.41	111.44	175.62	60.02	63.83
104	Asn	8.01	118.68	178.92	51.95	39.55
105	Ile	7.70	121.55	176.59	64.41	37.85
106	Asp	8.36	119.53	177.11	56.76	40.45
107	Thr	7.58	110.23	176.15	63.05	69.54
108	Leu	8.73	122.22	177.83	54.71	44.38
109	Gly	7.99	107.60	174.85	49.00	
110	Ser	8.90	115.97	176.89	62.24	
111	Arg	7.84	124.10	179.36	59.26	30.20
112	Val	8.39	120.14	177.30	66.60	31.32
113	Leu	8.57	118.49	178.85	58.68	39.85
114	Ser	8.20	113.87	176.20	61.82	62.68
115	Ala	7.74	123.02	181.06	54.85	18.26
116	Leu	8.66	121.67	178.35	59.17	42.20
117	Leu	8.55	121.02	179.30	58.88	41.50
118	Asn	8.18	117.30	177.98	56.90	38.65
119	Gly	8.28	109.35	175.80	47.23	
120	Val	8.97	123.30	176.77	67.10	31.28
121	Ser	8.16	113.67	176.76	61.86	62.43
122	Ser	8.02	116.77	177.67	60.80	62.87
123	Ala	8.21	124.55	179.98	54.56	20.27
124	Ala	8.50	119.51	179.64	55.37	18.79
125	Gln	8.15	118.69	180.67	59.37	28.51
126	Gly	7.84	106.62	174.77	46.55	
127	Leu	7.49	120.07	176.91	54.36	42.84
128	Gly	7.76	107.45	174.09	45.91	
129	Ile	7.96	122.63	174.39	60.67	39.04
130	Asn	8.40	124.79	174.40	53.74	39.27
131	Val	7.99	123.53	174.81	61.20	33.71
132	Asp	8.62	126.72	177.36	54.40	41.97
133	Ser	8.93	122.37	177.18	61.64	62.73
134	Gly	8.74	111.34	176.87	46.81	
135	Ser	7.85	118.91	175.22	61.49	62.88
136	Val	7.58	122.07	178.11	66.94	31.55
137	Gln	7.90	116.90	179.30	59.22	27.93
138	Ser	8.02	115.99	177.24	61.52	62.53
139	Asp	8.55	124.46	178.56	57.13	41.47
140	Ile	8.53	122.28	179.08	66.27	38.12
141	Ser	8.03	116.04	177.46	61.75	62.71
142	Ser	8.52	118.01	177.21	61.56	62.94
143	Cys	8.48	120.50	176.26	63.81	27.00
144	Ser	8.10	114.80	174.64	61.76	62.81
146	Phe	8.19	122.03	177.69	60.65	39.27
147	Leu	7.99	118.79	177.79	56.42	41.75
148	Ser	7.84	114.02	175.59	59.70	63.67
149	Thr	8.06	114.93	175.08	62.73	69.63
150	Ser	8.11	117.13	174.73		63.76

151	Ser	8.10	117.20	174.93	58.72	63.76
152	Ser	8.31	117.57	174.81	58.66	63.81
153	Ser	8.21	117.54	174.49	58.64	63.79
154	Ala	8.19	125.75	177.80	52.84	19.13
155	Ser	8.09	114.33	174.34	58.52	63.75
156	Tyr	8.01	121.78	175.85	58.00	38.73
157	Ser	8.08	117.13	174.38	58.23	63.92
158	Gln	8.27	122.31	175.89	56.03	29.40
159	Ala	8.25	124.89	177.92	52.88	19.14
160	Ser	8.19	114.79	174.53	58.35	63.91
161	Ala	8.27	125.89	177.84	52.84	19.23
162	Ser	8.21	114.36	174.83	58.44	63.86
163	Ser	8.25	117.53	175.02	58.56	63.86
164	Thr	8.17	115.01	174.75	61.87	69.61
165	Ser	8.23	117.53	175.00	58.62	63.90
166	Gly	8.37	110.89	173.93	45.31	
167	Ala	8.17	123.68	178.16	52.64	19.23
168	Gly	8.33	107.89	173.73	45.20	
169	Tyr	8.04	120.67	176.10	58.16	38.99
170	Thr	8.10	117.29	174.16	61.34	69.84
171	Gly	7.28	110.42	171.50	44.64	
172	Pro			177.37	63.33	32.22
173	Ser	8.43	116.03	174.52	58.32	64.05
174	Gly	8.14	110.47	171.69	44.66	
175	Pro			177.30	63.21	32.21
176	Ser	8.47	116.07	174.88	58.42	63.77
177	Thr	8.09	114.77	174.70	61.69	69.85
178	Gly	8.17	110.84	171.92	44.70	
179	Pro			177.39	63.45	32.10
180	Ser	8.41	115.68	174.97	58.61	63.83
181	Gly	8.23	110.15	173.35	45.08	
182	Tyr	8.03	121.38	174.41	55.65	38.49
183	Pro			176.86	63.33	31.99
184	Gly	7.53	108.25	171.76	44.65	
185	Pro			177.50	63.35	32.10
186	Leu	8.39	121.89	178.19	55.42	42.10
187	Gly	8.32	109.53	174.75	45.46	
188	Gly	8.21	108.49	174.59	45.32	
189	Gly	8.23	108.45	173.40	44.92	
190	Ala	8.09	124.59	175.71	50.55	18.30
191	Pro			176.64	63.18	31.80
192	Phe	8.17	119.73	176.43	57.92	39.43
193	Gly	8.20	110.67	174.08	45.37	
194	Gln	8.12	119.73	176.13	55.82	29.60
195	Ser	8.40	116.87	174.90	58.67	63.84
196	Gly	8.32	110.44	173.85	45.22	
197	Phe	8.14	120.05	176.23	57.89	39.62
198	Gly	8.36	111.66	173.53	45.32	
199	Gly	7.49	114.60	178.87	45.97	

b) Stapled W_1M

	H	N	C	C_α	C_β
1 Ala			174.06	51.95	19.32
2 Gly	8.50	108.83	171.74	44.52	
3 Pro			177.33	63.26	32.06
4 Gln	8.56	121.22	176.54	56.04	29.45
5 Gly	8.37	110.33	174.40	45.31	
6 Gly	8.15	108.64	174.03	45.11	
7 Phe	8.21	120.14	176.47	58.12	39.53
8 Gly	8.39	110.65	173.95	45.33	
9 Ala	8.13	124.06	178.18	52.70	19.31
10 Thr	8.11	112.55	175.29	61.92	69.81
11 Gly	8.34	111.14	174.74	45.47	
12 Gly	8.27	108.95	174.45	45.37	
13 Ala	8.29	124.24	178.19	52.98	19.33
14 Ser	8.15	114.60	174.81	58.76	63.58
15 Ala	8.18	125.03	179.97	54.93	18.36
16 Gly	8.36	106.39	176.11	46.51	
17 Leu	7.51	123.37	177.79	57.70	41.73
18 Ile	7.51	118.58	180.43	64.64	38.11
19 Ser	8.10	116.10	175.75	61.69	62.76
20 Arg	7.89	123.29	179.68	59.78	30.50
21 Val	7.85	120.48	177.14	66.68	31.94
22 Ala	8.32	121.96	179.17	55.83	18.40
23 Asn	8.58	116.17	177.27	56.15	38.27
24 Ala	7.46	121.26	180.65	54.85	18.06
25 Leu	8.04	116.70	178.98	57.37	43.23
26 Ala	8.77	120.63	176.71	54.60	18.44
27 Asn	7.24	113.23	175.65	52.82	39.37
28 Thr	7.53	110.08	175.04	61.01	71.42
29 Cys	8.98	122.53	178.03	57.85	38.77
30 Thr	8.77	116.65	174.95	67.35	68.96
31 Leu	7.42	119.29	178.31	59.46	40.83
32 Arg	7.77	113.89	177.27	58.58	30.93
33 Thr	7.59	109.39	174.80	64.66	69.62
34 Val	7.35	118.89	175.02	65.30	33.79
35 Leu	7.96	119.27	172.89	52.27	39.59
36 Arg	6.39	117.16	174.53	53.77	31.87
37 Thr	8.32	117.08	175.50	65.18	69.22
38 Gly	8.88	115.06	174.90	45.02	
39 Val	8.00	122.87	174.98	62.17	32.35
40 Ser	8.45	123.59	175.07	57.88	64.55
41 Gln	9.05	124.00	178.83	59.46	29.08
42 Gln	8.59	119.14	178.88	59.43	28.21
43 Ile	7.47	121.58	177.64	63.35	36.35
44 Ala	8.58	122.05	179.28	55.84	18.50
45 Ser	8.26	110.72	176.26	62.43	62.98
46 Ser	7.62	117.35	176.64	62.34	63.48
47 Val	8.62	121.38	177.46	66.86	31.48
48 Val	8.23	116.87	177.02	66.87	30.97
49 Gln	8.22	119.87	177.81	60.28	29.16
50 Arg	8.02	117.64	179.69	59.06	30.21

51	Ala	8.99	122.53	179.45	55.53	18.79
52	Ala	8.48	119.58	178.92	55.56	17.86
53	Gln	7.95	116.33	179.36	59.28	28.03
54	Ser	8.27	116.49	176.81	61.73	62.84
55	Leu	8.29	122.87	178.21	57.97	41.56
56	Ala	8.17	121.75	179.64	55.54	18.61
57	Ser	7.97	112.99	177.08	61.26	62.93
58	Thr	8.00	117.04	175.77	65.61	69.00
59	Leu	7.92	118.46	177.36	54.97	43.67
60	Gly	7.98	109.03	175.06	46.55	
61	Val	7.47	112.33	174.91	59.02	33.92
62	Asp	8.35	120.22	177.82	54.77	41.18
63	Gly	8.81	115.81	174.88	46.86	
64	Asn	8.38	120.79	177.89	56.55	38.00
65	Asn	8.41	119.72	176.71	56.07	38.41
66	Leu	7.95	119.41	177.87	58.66	41.41
67	Ala	8.51	119.22	178.66	55.73	18.33
68	Arg	7.78	117.95	179.46	59.92	29.69
69	Phe	8.01	117.04	179.18	61.71	39.17
70	Ala	8.70	123.47	178.28	55.84	17.72
71	Val	8.77	118.20	179.25	66.82	31.58
72	Gln	8.11	120.83	178.66	59.33	28.39
73	Ala	7.83	120.35	180.85	55.16	18.85
74	Val	8.11	115.68	178.30	65.07	31.64
75	Ser	8.33	117.48	174.48	61.37	63.07
76	Arg	6.96	117.82	176.52	55.71	30.72
77	Leu	7.35	122.05	175.40	52.94	40.47
78	Pro			177.53	61.90	32.44
79	Ala	8.74	126.06	178.83	53.50	18.15
80	Gly	8.94	110.81	175.13	45.69	
81	Ser	7.53	117.33	171.54	61.48	64.10
82	Asp	8.29	121.39	176.64	51.99	43.59
83	Thr	8.38	115.15	176.37	67.33	69.61
84	Ser	8.61	116.15	175.93	62.56	
85	Ala	7.79	126.20	180.84	54.97	18.59
86	Tyr	7.57	117.94	178.17	62.89	38.65
87	Ala	8.87	122.08	179.50	55.76	19.16
88	Gln	8.22	119.42	177.86	59.37	27.47
89	Ala	7.88	123.28	179.64	55.19	18.30
90	Phe	8.40	117.15	178.28	60.60	39.24
91	Ser	8.42	112.02	176.77	61.76	63.73
92	Ser	8.69	113.09	176.54	62.56	
93	Ala	7.95	123.84	178.60	55.89	17.81
94	Leu	8.14	117.11	179.41	58.49	42.39
95	Phe	7.79	117.46	180.22	60.89	38.93
96	Asn	9.19	122.52	176.67	56.08	37.44
97	Ala	8.13	117.67	177.44	52.24	19.54
98	Gly	7.50	105.43	174.24	45.52	
99	Val	7.61	119.32	175.93	64.83	32.12
100	Leu	6.42	113.29	174.45	51.49	45.34

101	Asn	8.67	114.94	173.91	52.15	40.37
102	Ala	8.87	119.06	178.90	55.14	18.66
103	Ser	8.39	111.78	175.67	60.08	63.78
104	Asn	8.01	119.16	178.97	51.98	39.53
105	Ile	7.66	121.59	176.66	64.15	37.84
106	Asp	8.33	120.35	177.18	56.88	40.72
107	Thr	7.63	111.28	176.07	63.12	69.34
108	Leu	8.66	122.49	177.86	54.88	44.16
109	Gly	7.89	107.25	174.89	48.95	
110	Ser	8.65	115.91	176.85	62.21	62.82
111	Arg	7.82	123.64	179.15	59.11	30.22
112	Val	8.28	120.15	177.06	66.55	31.50
113	Leu	8.54	119.02	178.75	58.78	40.52
114	Ser	8.05	113.37	176.72	62.10	62.77
115	Ala	7.82	123.60	180.97	54.93	18.25
116	Leu	8.66	122.46	178.56	58.73	41.95
117	Leu	8.34	120.27	179.33	58.44	40.91
118	Asn	8.14	118.80	178.16	56.36	38.17
119	Gly	8.20	109.57	175.83	47.21	
120	Val	8.64	124.18	176.79	67.09	31.16
121	Ser	8.32	115.03	177.29	62.17	
122	Ser	8.42	116.16	177.58	61.61	62.81
123	Ala	8.21	125.06	179.55	54.57	20.27
124	Ala	8.33	119.57	179.38	55.23	18.71
125	Gln	8.04	118.07	180.20	58.98	28.48
126	Gly	8.08	107.73	175.16	46.59	
127	Leu	7.59	120.04	177.15	54.74	43.19
128	Gly	7.81	108.33	174.50	46.14	
129	Ile	7.83	120.39	175.08	61.00	38.50
130	Asn	8.48	124.47	175.00	53.26	38.63
131	Val	8.10	120.85	175.27	61.68	32.64
132	Asp	8.26	123.52	176.42	54.16	41.29
133	Ser	8.31	117.91	175.39	59.16	63.71
134	Gly	8.48	110.85	174.52	45.73	
135	Ser	8.05	115.74	174.78	58.66	63.93
136	Val	7.95	121.07	176.33	62.78	32.44
137	Gln	8.31	122.91	176.17	56.30	29.16
138	Ser	8.16	116.34	174.30	58.90	63.89
139	Asp	8.30	121.70	176.68	54.48	40.94
140	Ile	8.02	119.74	176.59	61.71	38.49
141	Ser	8.24	118.40	175.10	59.59	63.65
142	Ser	8.17	116.78	174.33	58.73	63.65
143	Cys	8.15	119.11	174.08	55.14	41.37
144	Ser	8.52	116.97	174.87	58.44	64.29
145	Ser	8.48	119.16	174.04	58.79	63.84
146	Phe	8.08	121.66	175.35	57.98	39.72
147	Leu	8.11	122.92	176.98	55.10	42.53
148	Ser	8.20	116.75	174.92	58.37	63.89
149	Thr	8.13	115.30	174.74	61.90	69.63
150	Ser	8.19	117.87	174.97	58.63	63.94

151	Ser					
152	Ser					
153	Ser			174.40	58.57	63.82
154	Ala	8.20	126.02	177.71	52.80	19.17
155	Ser	8.09	114.72	174.32	58.49	63.78
156	Tyr	8.03	122.16	175.82	58.09	38.74
157	Ser	8.08	117.43	174.36	58.26	63.93
158	Gln	8.26	122.63	175.86	56.09	29.38
159	Ala	8.24	125.11	177.88	52.87	19.15
160	Ser	8.17	115.01	174.52	58.35	63.91
161	Ala	8.26	126.17	177.82	52.82	19.23
162	Ser	8.20	114.69	174.81	58.43	63.89
163	Ser	8.27	117.86	174.99	58.55	63.87
164	Thr	8.14	115.26	174.74	61.85	69.66
165	Ser	8.24	117.89	174.38	58.61	63.87
166	Gly	8.30	110.71	173.88	45.28	
167	Ala	8.14	123.99	178.14	52.70	19.26
168	Gly	8.32	108.17	173.71	45.19	
169	Tyr	8.03	120.96	176.08	58.16	39.00
170	Thr	8.08	117.50	174.15	61.35	69.85
171	Gly	7.27	110.71	171.51	44.64	
172	Pro			177.30	63.34	32.21
173	Ser	8.41	116.30	174.49	58.30	64.06
174	Gly	8.13	110.77	171.69	44.66	
175	Pro			177.24	63.23	32.20
176	Ser	8.45	116.37	174.84	58.41	63.80
177	Thr	8.08	115.09	174.68	61.67	69.87
178	Gly	8.16	111.14	171.91	44.68	
179	Pro			177.35	63.43	32.09
180	Ser	8.39	116.01	174.95	58.60	63.88
181	Gly	8.22	110.45	173.32	45.08	
182	Tyr	8.02	121.69	174.39	55.63	38.54
183	Pro			176.80	63.32	31.99
184	Gly	7.51	108.54	171.76	44.58	
185	Pro			177.44	63.31	32.11
186	Leu	8.38	122.12	178.11	55.39	42.11
187	Gly	8.27	109.57	174.67	45.42	
188	Gly	8.20	108.76	174.55	45.38	
189	Gly	8.21	108.75	173.36	44.90	
190	Ala	8.08	124.88	175.70	50.55	18.34
191	Pro			177.44	63.35	32.10
192	Phe	8.34	121.87	178.11	55.40	42.10
193	Gly	8.30	109.79	174.73	45.41	
194	Gln	8.11	120.04	176.10	55.83	29.62
195	Ser	8.38	117.13	174.86	58.66	63.88
196	Gly	8.31	110.71	173.83	45.22	
197	Phe	8.14	120.32	176.16	57.92	39.64
198	Gly	8.35	111.91	173.49	45.31	
199	Gly	7.49	114.89	178.85	45.96	

***APPENDIX E**: PANAV verification of CS assignments for W_1M proteins.*

a) reduced W_1M

Seq	C	CA	CB	N	H	HA	Beta sheet probability	Coil probability	Helix probability	SecStr
A1	174.1058	51.9397	19.3224				0.83	0.17	0.00	B
G2	171.7465	44.5457		108.5064	8.5111		0.30	0.70	0.00	C
P3	177.3747	63.2725	32.0794				0.14	0.85	0.01	C
Q4	176.5603	56.0427	29.4331	120.8846	8.5710		0.19	0.78	0.02	C
G5	174.4242	45.3281		109.9944	8.3725		0.05	0.93	0.01	C
G6	174.0464	45.1214		108.3108	8.1539		0.07	0.91	0.01	C
F7	176.5007	58.1245	39.4888	119.7971	8.1985		0.07	0.65	0.28	C
G8	173.9757	45.3421		110.3443	8.3958		0.06	0.93	0.01	C
A9	178.2127	52.7104	19.2951	123.7738	8.1436		0.04	0.89	0.07	C
T10	175.3098	61.9311	69.8127	112.1968	8.1151		0.52	0.45	0.03	B
G11	174.7802	45.4999		110.7769	8.3397		0.04	0.92	0.04	C
G12	174.5319	45.4058		108.6497	8.2811		0.06	0.91	0.04	C
A13	178.2928	53.0478	19.3453	123.9413	8.3256		0.02	0.81	0.17	C
S14	174.8531	58.8395	63.5163	114.0578	8.2000		0.16	0.78	0.06	C
A15	180.0200	55.0294	18.3572	124.6776	8.1577		0.00	0.01	0.99	H
G16	176.1288	46.5907		106.0387	8.3820		0.02	0.41	0.58	H
L17	177.8592	57.6499	41.9496	122.9578	7.5272		0.00	0.03	0.97	H
I18	180.3361	64.7765	37.8661	118.4029	7.5206		0.00	0.00	1.00	H
S19	175.7142	61.6475	62.8061	115.3911	8.0539		0.00	0.04	0.96	H
R20	178.6782	59.9927	30.7090	123.1340	7.8630		0.00	0.01	0.99	H
V21	177.1906	66.9333	31.8020	118.2138	7.7999		0.00	0.00	1.00	H
A22	178.9895	55.7320	18.5163	121.0200	8.2832		0.00	0.00	1.00	H
N23	177.2499	55.9413	38.2954	116.1083	8.5973		0.00	0.04	0.96	H
A24	179.7356	54.8183	17.7701	121.1156	7.4174		0.00	0.01	0.99	H
L25	179.0685	57.0485	43.0607	115.0105	7.8219		0.00	0.08	0.92	H
A26	176.7860	54.4868	18.3178	121.1193	8.8435		0.00	0.42	0.58	H
N27	175.5385	52.6781	39.1242	113.3483	7.0125		0.01	0.98	0.01	C
T28	174.0709	60.6759	70.3947	109.2888	7.3508		0.49	0.51	0.00	C
C29	179.4854	62.5471	26.5899	126.8303	10.3535		0.02	0.54	0.45	C
T30	175.6170	68.6640		119.4136	10.6732		0.01	0.98	0.01	C
L31	178.2496	59.9960	39.7962	118.4226	7.5231		0.00	0.03	0.97	H
R32	177.8146	59.0602	30.6735	114.5413	7.6129		0.00	0.09	0.91	H
T33	174.6407	65.5300	69.4159	111.7580	7.8750		0.01	0.21	0.78	H
V34	175.0380	65.5076	33.8660	118.4187	7.4142		0.02	0.97	0.01	C
L35	172.9300	51.9903	39.3957	119.6804	8.1256		0.07	0.93	0.00	C
R36	174.5924	53.8199	31.7305	116.4045	6.2405		0.20	0.80	0.00	C
T37	175.7406	65.1246	69.1616	117.2220	8.3079		0.01	0.03	0.97	H

G38	174.8624	45.0506		115.1038	8.9271		0.10	0.90	0.00	C
V39	174.7226	62.0978	32.2303	122.8253	8.0774		0.63	0.37	0.00	B
S40	175.2106	57.8944	64.5176	123.1214	8.4208		0.44	0.56	0.00	C
Q41	178.8469	59.4244	28.9806	123.8049	9.0582		0.00	0.02	0.98	H
Q42	178.9202	59.4537	28.1919	118.8008	8.5673		0.00	0.01	0.99	H
I43	177.5392	63.5561	36.4089	121.2874	7.4704		0.00	0.03	0.97	H
A44	179.2643	55.8485	18.5119	121.4400	8.5121		0.00	0.01	0.99	H
S45	176.3009	62.4589	63.0152	110.1281	8.3247		0.00	0.03	0.97	H
S46	176.7143	62.3389	63.4778	116.9825	7.6357		0.00	0.04	0.96	H
V47	177.5009	66.9112	31.4106	120.5798	8.6480		0.00	0.00	1.00	H
V48	177.0072	66.9125	30.8850	116.3027	8.1753		0.00	0.03	0.97	H
Q49	177.8534	60.3057	29.1182	119.6215	8.2470		0.00	0.01	0.99	H
R50	179.7302	59.0321	30.0671	117.3253	8.0447		0.00	0.00	1.00	H
A51	179.4060	55.4954	18.8080	122.0831	9.0117		0.00	0.04	0.96	H
A52	178.9565	55.5546	17.7976	119.2718	8.5317		0.00	0.02	0.98	H
Q53	179.3674	59.3003	28.0216	116.0714	7.9277		0.00	0.00	1.00	H
S54	176.9602	61.7578	62.8391	116.1993	8.3139		0.00	0.01	0.99	H
L55	178.0786	58.3894	41.4963	123.2296	8.4766		0.00	0.02	0.98	H
A56	179.6438	55.5476	18.5628	121.8122	8.1683		0.00	0.01	0.99	H
S57	177.3100	61.3306	62.8816	112.8948	7.9965		0.00	0.02	0.98	H
T58	175.7743	65.8327	68.9606	117.2649	8.0698		0.00	0.02	0.98	H
L59	177.2753	55.0094	43.4980	118.1772	7.8824		0.03	0.95	0.03	C
G60	175.1123	46.5796		108.8331	8.0154		0.03	0.79	0.18	C
V61	174.9159	58.9702	33.9297	111.9070	7.5588		0.78	0.22	0.00	B
D62	177.9274	54.8880	41.3033	119.7639	8.3462		0.05	0.57	0.37	C
G63	174.9002	46.8898		115.7190	8.8589		0.04	0.95	0.01	C
N64	177.9149	56.5862	37.9690	120.5301	8.3711		0.00	0.01	0.99	H
N65	176.7229	56.0973	38.3457	119.3866	8.4330		0.00	0.05	0.95	H
L66	177.8786	58.6763	41.2993	119.0390	7.9582		0.00	0.02	0.98	H
A67	178.7041	55.7612	18.2951	118.9361	8.5323		0.00	0.04	0.96	H
R68	179.4133	59.9252	29.6542	117.5609	7.7725		0.00	0.00	1.00	H
F69	179.2016	61.7142	39.1485	116.7777	8.0366		0.00	0.01	0.99	H
A70	178.2901	55.8537	17.6656	123.0225	8.7308		0.00	0.06	0.94	H
V71	179.3012	66.8155	31.5403	117.7773	8.7544		0.00	0.00	1.00	H
Q72	178.6056	59.3355	28.3442	120.6079	8.1308		0.00	0.00	1.00	H
A73	180.8489	55.1560	18.8147	119.9172	7.8440		0.00	0.00	1.00	H
V74	178.3741	65.1215	31.6667	114.9644	8.0739		0.00	0.02	0.97	H
S75	174.5340	61.3236	63.0393	117.0778	8.3646		0.00	0.24	0.76	H
R76	176.5216	55.7672	30.6999	117.4561	6.9585		0.12	0.85	0.03	C
L77	175.3954	52.9336	40.3671	121.7712	7.3235		0.26	0.74	0.00	C
P78	177.5849	61.9137	32.4270				0.22	0.78	0.00	C

A79	178.8799	53.5574	18.0862	125.8730	8.7608		0.01	0.58	0.42	C
G80	175.0631	45.6205		110.5102	8.9413		0.05	0.93	0.02	C
S81	171.5800	61.4666	64.0696	116.9898	7.5431		0.04	0.91	0.06	C
D82	176.6758	52.0728	43.5621	121.1052	8.2999		0.43	0.57	0.00	C
T83	176.3632	67.2500	69.5758	115.0095	8.2725		0.00	0.00	1.00	H
S84	175.9847	62.5736	61.9184	115.9699	8.6197		0.00	0.04	0.96	H
A85	180.9256	54.9716	18.5377	125.8495	7.8133		0.00	0.01	0.99	H
Y86	178.2428	62.9214	38.8054	117.7813	7.6216		0.00	0.01	0.99	H
A87	179.5553	55.7930	18.9494	121.8471	8.8581		0.00	0.03	0.97	H
Q88	177.9005	59.3388	27.4133	119.1036	8.2251		0.00	0.01	0.99	H
A89	179.6285	55.1889	18.3675	122.9847	7.9409		0.00	0.01	0.99	H
F90	178.4468	60.4093	38.9932	117.3176	8.4209		0.00	0.03	0.97	H
S91	176.8835	61.8838	63.7587	111.5604	8.4006		0.00	0.03	0.97	H
S92	176.5553	62.5677	63.7263	112.9237	8.7312		0.00	0.11	0.88	H
A93	178.6320	55.8888	17.7417	123.6439	7.9438		0.00	0.03	0.97	H
L94	179.3503	58.4455	42.4156	116.4015	8.0599		0.00	0.01	0.99	H
F95	180.1989	60.6785	38.9570	117.2018	7.7677		0.00	0.01	0.99	H
N96	176.6958	56.0536	37.4695	122.2296	9.2412		0.02	0.38	0.60	H
A97	177.4545	52.2479	19.5769	117.3242	8.1035		0.15	0.85	0.00	C
G98	174.2338	45.5325		105.1079	7.5188		0.04	0.93	0.03	C
V99	175.9398	64.8407	31.9916	119.0345	7.6217		0.01	0.42	0.57	H
L100	174.4660	51.5675	45.5552	113.0498	6.4587		0.00	1.00	0.00	C
N101	173.9405	52.1908	40.3456	114.5191	8.6553		0.23	0.77	0.00	C
A102	178.9335	55.1576	18.6531	118.8029	8.8791		0.00	0.13	0.87	H
S103	175.6162	60.0180	63.8293	111.4363	8.4128		0.02	0.47	0.51	H
N104	178.9188	51.9511	39.5521	118.6825	8.0138		0.08	0.62	0.31	C
I105	176.5945	64.4068	37.8498	121.5453	7.7021		0.00	0.06	0.94	H
D106	177.1117	56.7602	40.4493	119.5261	8.3585		0.00	0.08	0.92	H
T107	176.1476	63.0482	69.5350	110.2306	7.5837		0.05	0.60	0.35	C
L108	177.8298	54.7068	44.3766	122.2180	8.7328		0.38	0.61	0.01	C
G109	174.8544	49.0028		107.5990	7.9872		0.00	0.09	0.90	H
S110	176.8946	62.2365		115.9721	8.9035		0.00	0.05	0.95	H
R111	179.3572	59.2559	30.1992	124.0966	7.8423		0.00	0.01	0.99	H
V112	177.3040	66.5995	31.3212	120.1424	8.3901		0.00	0.00	1.00	H
L113	178.8523	58.6766	39.8519	118.4858	8.5722		0.00	0.00	1.00	H
S114	176.2003	61.8231	62.6753	113.8682	8.1969		0.00	0.03	0.97	H
A115	181.0586	54.8547	18.2579	123.0238	7.7440		0.00	0.00	1.00	H
L116	178.3455	59.1657	42.2043	121.6697	8.6612		0.00	0.03	0.97	H
L117	179.2978	58.8823	41.5045	121.0206	8.5529		0.00	0.01	0.99	H
N118	177.9770	56.8962	38.6514	117.3042	8.1750		0.00	0.00	1.00	H
G119	175.7971	47.2349		109.3527	8.2782		0.01	0.31	0.68	H

V120	176.7708	67.0961	31.2824	123.3044	8.9687		0.00	0.01	0.99	H
S121	176.7619	61.8625	62.4278	113.6721	8.1606		0.00	0.01	0.99	H
S122	177.6718	60.8027	62.8700	116.7747	8.0211		0.00	0.04	0.96	H
A123	179.9799	54.5617	20.2734	124.5480	8.2140		0.00	0.08	0.92	H
A124	179.6363	55.3740	18.7851	119.5076	8.5039		0.00	0.02	0.98	H
Q125	180.6719	59.3739	28.5115	118.6925	8.1526		0.00	0.00	1.00	H
G126	174.7685	46.5505		106.6160	7.8446		0.01	0.58	0.40	C
L127	176.9093	54.3611	42.8414	120.0670	7.4858		0.14	0.86	0.00	C
G128	174.0873	45.9092		107.4487	7.7571		0.04	0.90	0.06	C
I129	174.3915	60.6730	39.0436	122.6253	7.9648		0.62	0.38	0.00	B
N130	174.3997	53.7403	39.2741	124.7889	8.3967		0.34	0.65	0.00	C
V131	174.8080	61.1952	33.7056	123.5267	7.9912		0.85	0.15	0.00	B
D132	177.3592	54.4014	41.9665	126.7206	8.6171		0.27	0.73	0.00	C
S133	177.1752	61.6400	62.7302	122.3662	8.9285		0.00	0.05	0.95	H
G134	176.8718	46.8070		111.3391	8.7390		0.01	0.29	0.70	H
S135	175.2246	61.4910	62.8834	118.9066	7.8488		0.00	0.11	0.89	H
V136	178.1132	66.9363	31.5547	122.0706	7.5770		0.00	0.00	1.00	H
Q137	179.2975	59.2167	27.9310	116.9045	7.8983		0.00	0.00	1.00	H
S138	177.2433	61.5220	62.5256	115.9851	8.0227		0.00	0.01	0.99	H
D139	178.5632	57.1264	41.4670	124.4583	8.5481		0.00	0.02	0.97	H
I140	179.0814	66.2688	38.1158	122.2764	8.5293		0.00	0.00	1.00	H
S141	177.4561	61.7464	62.7111	116.0410	8.0279		0.00	0.01	0.99	H
S142	177.2110	61.5566	62.9385	118.0135	8.5210		0.00	0.03	0.97	H
C143	176.2603	63.8096	27.0037	120.4998	8.4831		0.00	0.15	0.85	H
S144	174.6421	61.7643	62.8127	114.7987	8.0963		0.00	0.25	0.75	H
S145							0.39	0.47	0.14	C
F146	177.6861	60.6498	39.2690	122.0267	8.1882		0.00	0.03	0.97	H
L147	177.7860	56.4210	41.7453	118.7931	7.9876		0.00	0.11	0.89	H
S148	175.5917	59.6993	63.6670	114.0195	7.8374		0.02	0.51	0.47	C
T149	175.0795	62.7328	69.6349	114.9301	8.0592		0.37	0.55	0.08	C
S150	174.7250		63.7572	117.1275	8.1069		0.13	0.65	0.22	C
S151	174.9347	58.7197	63.7608	117.2007	8.0980		0.14	0.83	0.03	C
S152	174.8145	58.6630	63.8078	117.5664	8.3113		0.17	0.81	0.03	C
S153	174.4944	58.6361	63.7932	117.5362	8.2143		0.18	0.80	0.02	C
A154	177.7981	52.8392	19.1282	125.7527	8.1934		0.05	0.89	0.06	C
S155	174.3429	58.5194	63.7499	114.3286	8.0928		0.26	0.73	0.01	C
Y156	175.8507	58.0001	38.7282	121.7836	8.0101		0.23	0.63	0.13	C
S157	174.3792	58.2250	63.9198	117.1333	8.0752		0.30	0.69	0.01	C
Q158	175.8927	56.0332	29.3968	122.3143	8.2655		0.18	0.76	0.06	C
A159	177.9193	52.8756	19.1395	124.8945	8.2515		0.02	0.95	0.03	C

S160	174.5331	58.3535	63.9065	114.7893	8.1885		0.31	0.68	0.01	C
A161	177.8350	52.8392	19.2284	125.8874	8.2704		0.05	0.89	0.06	C
S162	174.8348	58.4426	63.8594	114.3561	8.2128		0.26	0.73	0.02	C
S163	175.0226	58.5646	63.8622	117.5265	8.2450		0.16	0.81	0.03	C
T164	174.7509	61.8734	69.6097	115.0063	8.1716		0.56	0.43	0.01	B
S165	174.9953	58.6233	63.9035	117.5271	8.2308		0.13	0.81	0.05	C
G166	173.9257	45.3098		110.8861	8.3657		0.05	0.93	0.01	C
A167	178.1621	52.6429	19.2313	123.6770	8.1664		0.04	0.90	0.05	C
G168	173.7252	45.2017		107.8948	8.3336		0.07	0.92	0.01	C
Y169	176.1014	58.1608	38.9921	120.6701	8.0435		0.30	0.55	0.15	C
T170	174.1625	61.3447	69.8382	117.2942	8.0955		0.80	0.20	0.00	B
G171	171.5042	44.6385		110.4166	7.2845		0.22	0.78	0.00	C
P172	177.3694	63.3282	32.2156				0.14	0.85	0.01	C
S173	174.5214	58.3200	64.0518	116.0313	8.4260		0.47	0.53	0.01	C
G174	171.6946	44.6572		110.4708	8.1398		0.24	0.76	0.00	C
P175	177.2968	63.2074	32.2136				0.15	0.84	0.01	C
S176	174.8799	58.4185	63.7659	116.0667	8.4665		0.38	0.60	0.02	C
T177	174.7046	61.6897	69.8506	114.7731	8.0884		0.57	0.42	0.01	B
G178	171.9197	44.7011		110.8351	8.1714		0.20	0.80	0.00	C
P179	177.3902	63.4480	32.1005				0.12	0.86	0.02	C
S180	174.9662	58.6138	63.8275	115.6810	8.4052		0.33	0.65	0.03	C
G181	173.3544	45.0779		110.1458	8.2337		0.08	0.92	0.01	C
Y182	174.4091	55.6518	38.4852	121.3813	8.0303		0.61	0.39	0.00	B
P183	176.8649	63.3310	31.9871				0.17	0.82	0.01	C
G184	171.7595	44.6487		108.2462	7.5282		0.30	0.70	0.00	C
P185	177.4969	63.3542	32.1022				0.12	0.86	0.02	C
L186	178.1854	55.4152	42.1047	121.8887	8.3877		0.08	0.71	0.21	C
G187	174.7482	45.4562		109.5338	8.3205		0.05	0.94	0.01	C
G188	174.5911	45.3245		108.4863	8.2114		0.06	0.91	0.03	C
G189	173.3957	44.9198		108.4458	8.2264		0.11	0.89	0.01	C
A190	175.7142	50.5474	18.3035	124.5908	8.0854		0.57	0.43	0.00	B
P191	176.6366	63.1777	31.7996				0.20	0.79	0.00	C
F192	176.4260	57.9185	39.4329	119.7305	8.1684		0.15	0.56	0.28	C
G193	174.0837	45.3703		110.6658	8.2015		0.06	0.93	0.01	C
Q194	176.1312	55.8195	29.5996	119.7279	8.1175		0.14	0.84	0.03	C
S195	174.9005	58.6697	63.8362	116.8712	8.3978		0.17	0.78	0.05	C
G196	173.8536	45.2183		110.4387	8.3205		0.06	0.93	0.01	C
F197	176.2282	57.8894	39.6211	120.0492	8.1403		0.10	0.70	0.20	C
G198	173.5254	45.3204		111.6630	8.3634		0.08	0.91	0.01	C
G199	178.8703	45.9736		114.5967	7.4895		0.02	0.41	0.57	H

CSI and Sec Structure (No_Deviant_Ref_Calibrated)

Seq	C	CA	CB	N	H	HA	Beta sheet probability	Coil probability	Helix probability	SecStr
A1	173.75	51.45	19.55				0.92	0.08	0.00	B
G2	171.39	44.06		108.86	8.5111		0.44	0.56	0.00	C
P3	177.01	62.78	32.31				0.22	0.78	0.00	C
Q4	176.20	55.55	29.66	121.23	8.5710		0.37	0.63	0.00	C
G5	174.06	44.84		110.34	8.3725		0.07	0.92	0.00	C
G6	173.69	44.63		108.66	8.1539		0.10	0.90	0.00	C
F7	176.14	57.63	39.72	120.15	8.1985		0.16	0.70	0.14	C
G8	173.62	44.85		110.69	8.3958		0.09	0.91	0.00	C
A9	177.85	52.22	19.53	124.12	8.1436		0.10	0.89	0.01	C
T10	174.95	61.44	70.04	112.55	8.1151		0.66	0.34	0.01	B
G11	174.42	45.01		111.13	8.3397		0.05	0.94	0.01	C
G12	174.17	44.92		109.00	8.2811		0.07	0.92	0.01	C
A13	177.93	52.56	19.58	124.29	8.3256		0.08	0.90	0.02	C
S14	174.49	58.35	63.75	114.41	8.2000		0.29	0.70	0.01	C
A15	179.66	54.54	18.59	125.03	8.1577		0.00	0.03	0.97	H
G16	175.77	46.10		106.39	8.3820		0.03	0.67	0.30	C
L17	177.50	57.16	42.18	123.31	7.5272		0.00	0.07	0.93	H
I18	179.98	64.29	38.10	118.75	7.5206		0.00	0.00	1.00	H
S19	175.35	61.16	63.04	115.74	8.0539		0.00	0.10	0.90	H
R20	178.32	59.50	30.94	123.48	7.8630		0.00	0.03	0.97	H
V21	176.83	66.44	32.03	118.56	7.7999		0.00	0.01	0.99	H
A22	178.63	55.24	18.75	121.37	8.2832		0.00	0.01	0.99	H
N23	176.89	55.45	38.53	116.46	8.5973		0.00	0.10	0.90	H
A24	179.38	54.33	18.00	121.47	7.4174		0.00	0.03	0.97	H
L25	178.71	56.56	43.29	115.36	7.8219		0.00	0.19	0.81	H
A26	176.43	54.00	18.55	121.47	8.8435		0.02	0.72	0.26	C
N27	175.18	52.19	39.35	113.70	7.0125		0.02	0.98	0.00	C
T28	173.71	60.19	70.62	109.64	7.3508		0.55	0.45	0.00	B
C29	179.13	62.06	26.82	127.18	10.3535		0.05	0.63	0.32	C
T30	175.26	68.17		119.76	10.6732		0.01	0.98	0.01	C
L31	177.89	59.51	40.03	118.77	7.5231		0.00	0.03	0.97	H
R32	177.45	58.57	30.90	114.89	7.6129		0.00	0.18	0.82	H
T33	174.28	65.04	69.65	112.11	7.8750		0.07	0.44	0.49	H
V34	174.68	65.02	34.10	118.77	7.4142		0.04	0.95	0.00	C
L35	172.57	51.50	39.63	120.03	8.1256		0.08	0.92	0.00	C
R36	174.23	53.33	31.96	116.75	6.2405		0.28	0.72	0.00	C
T37	175.38	64.63	69.39	117.57	8.3079		0.03	0.08	0.89	H

G38	174.50	44.56		115.45	8.9271		0.14	0.86	0.00	C
V39	174.36	61.61	32.46	123.18	8.0774		0.74	0.26	0.00	B
S40	174.85	57.40	64.75	123.47	8.4208		0.56	0.44	0.00	B
Q41	178.49	58.93	29.21	124.15	9.0582		0.00	0.06	0.94	H
Q42	178.56	58.96	28.42	119.15	8.5673		0.00	0.02	0.98	H
I43	177.18	63.07	36.64	121.64	7.4704		0.00	0.07	0.93	H
A44	178.90	55.36	18.74	121.79	8.5121		0.00	0.02	0.98	H
S45	175.94	61.97	63.25	110.48	8.3247		0.00	0.05	0.95	H
S46	176.35	61.85	63.71	117.33	7.6357		0.00	0.09	0.91	H
V47	177.14	66.42	31.64	120.93	8.6480		0.00	0.01	0.99	H
V48	176.65	66.42	31.12	116.65	8.1753		0.00	0.03	0.97	H
Q49	177.49	59.82	29.35	119.97	8.2470		0.00	0.02	0.98	H
R50	179.37	58.54	30.30	117.68	8.0447		0.00	0.01	0.99	H
A51	179.05	55.01	19.04	122.43	9.0117		0.00	0.09	0.91	H
A52	178.60	55.06	18.03	119.62	8.5317		0.00	0.04	0.96	H
Q53	179.01	58.81	28.25	116.42	7.9277		0.00	0.00	1.00	H
S54	176.60	61.27	63.07	116.55	8.3139		0.00	0.04	0.96	H
L55	177.72	57.90	41.73	123.58	8.4766		0.00	0.04	0.96	H
A56	179.28	55.06	18.79	122.16	8.1683		0.00	0.02	0.98	H
S57	176.95	60.84	63.11	113.24	7.9965		0.00	0.05	0.95	H
T58	175.41	65.34	69.19	117.61	8.0698		0.01	0.06	0.94	H
L59	176.92	54.52	43.73	118.53	7.8824		0.07	0.93	0.00	C
G60	174.75	46.09		109.18	8.0154		0.04	0.91	0.05	C
V61	174.56	58.48	34.16	112.26	7.5588		0.82	0.18	0.00	B
D62	177.57	54.40	41.53	120.11	8.3462		0.11	0.83	0.07	C
G63	174.54	46.40		116.07	8.8589		0.04	0.96	0.00	C
N64	177.55	56.10	38.20	120.88	8.3711		0.00	0.03	0.97	H
N65	176.36	55.61	38.58	119.74	8.4330		0.01	0.13	0.86	H
L66	177.52	58.19	41.53	119.39	7.9582		0.00	0.04	0.96	H
A67	178.34	55.27	18.53	119.29	8.5323		0.00	0.06	0.94	H
R68	179.05	59.44	29.88	117.91	7.7725		0.00	0.00	1.00	H
F69	178.84	61.22	39.38	117.13	8.0366		0.00	0.02	0.98	H
A70	177.93	55.36	17.90	123.37	8.7308		0.00	0.10	0.90	H
V71	178.94	66.33	31.77	118.13	8.7544		0.00	0.00	1.00	H
Q72	178.25	58.85	28.57	120.96	8.1308		0.00	0.01	0.99	H
A73	180.49	54.67	19.04	120.27	7.8440		0.00	0.01	0.99	H
V74	178.01	64.63	31.90	115.31	8.0739		0.00	0.08	0.92	H
S75	174.17	60.83	63.27	117.43	8.3646		0.01	0.50	0.48	C
R76	176.16	55.28	30.93	117.81	6.9585		0.22	0.77	0.01	C
L77	175.04	52.44	40.60	122.12	7.3235		0.32	0.68	0.00	C
P78	177.22	61.42	32.66				0.29	0.71	0.00	C

A79	178.52	53.07	18.32	126.22	8.7608		0.04	0.87	0.10	C
G80	174.70	45.13		110.86	8.9413		0.07	0.93	0.00	C
S81	171.22	60.98	64.30	117.34	7.5431		0.12	0.86	0.02	C
D82	176.32	51.58	43.79	121.46	8.2999		0.52	0.48	0.00	B
T83	176.00	66.76	69.81	115.36	8.2725		0.00	0.01	0.99	H
S84	175.62	62.08	62.15	116.32	8.6197		0.00	0.06	0.94	H
A85	180.57	54.48	18.77	126.20	7.8133		0.00	0.02	0.98	H
Y86	177.88	62.43	39.04	118.13	7.6216		0.00	0.02	0.98	H
A87	179.20	55.30	19.18	122.20	8.8581		0.00	0.07	0.93	H
Q88	177.54	58.85	27.64	119.45	8.2251		0.00	0.03	0.97	H
A89	179.27	54.70	18.60	123.33	7.9409		0.00	0.04	0.96	H
F90	178.09	59.92	39.22	117.67	8.4209		0.00	0.05	0.95	H
S91	176.52	61.39	63.99	111.91	8.4006		0.00	0.07	0.93	H
S92	176.20	62.08	63.96	113.27	8.7312		0.00	0.22	0.78	H
A93	178.27	55.40	17.97	123.99	7.9438		0.00	0.04	0.96	H
L94	178.99	57.96	42.65	116.75	8.0599		0.00	0.01	0.99	H
F95	179.84	60.19	39.19	117.55	7.7677		0.00	0.02	0.98	H
N96	176.34	55.56	37.70	122.58	9.2412		0.05	0.62	0.33	C
A97	177.09	51.76	19.81	117.67	8.1035		0.30	0.70	0.00	C
G98	173.87	45.04		105.46	7.5188		0.06	0.93	0.01	C
V99	175.58	64.35	32.22	119.38	7.6217		0.05	0.75	0.20	C
L100	174.11	51.08	45.79	113.40	6.4587		0.00	1.00	0.00	C
N101	173.58	51.70	40.58	114.87	8.6553		0.35	0.65	0.00	C
A102	178.57	54.67	18.88	119.15	8.8791		0.00	0.23	0.77	H
S103	175.26	59.53	64.06	111.79	8.4128		0.06	0.72	0.22	C
N104	178.56	51.46	39.78	119.03	8.0138		0.14	0.76	0.10	C
I105	176.23	63.92	38.08	121.90	7.7021		0.00	0.17	0.83	H
D106	176.75	56.27	40.68	119.88	8.3585		0.00	0.26	0.74	H
T107	175.79	62.56	69.77	110.58	7.5837		0.12	0.75	0.13	C
L108	177.47	54.22	44.61	122.57	8.7328		0.61	0.39	0.00	B
G109	174.49	48.51		107.95	7.9872		0.01	0.23	0.76	H
S110	176.53	61.75		116.32	8.9035		0.00	0.09	0.91	H
R111	179.00	58.77	30.43	124.45	7.8423		0.00	0.02	0.98	H
V112	176.94	66.11	31.55	120.49	8.3901		0.00	0.01	0.99	H
L113	178.49	58.19	40.08	118.84	8.5722		0.00	0.00	1.00	H
S114	175.84	61.33	62.91	114.22	8.1969		0.00	0.07	0.93	H
A115	180.70	54.36	18.49	123.37	7.7440		0.00	0.01	0.99	H
L116	177.99	58.68	42.43	122.02	8.6612		0.00	0.06	0.94	H
L117	178.94	58.39	41.73	121.37	8.5529		0.00	0.01	0.99	H
N118	177.62	56.41	38.88	117.65	8.1750		0.00	0.01	0.99	H
G119	175.44	46.74		109.70	8.2782		0.02	0.55	0.42	C

V120	176.41	66.61	31.51	123.65	8.9687		0.00	0.01	0.99	H
S121	176.40	61.37	62.66	114.02	8.1606		0.00	0.02	0.98	H
S122	177.31	60.31	63.10	117.12	8.0211		0.00	0.11	0.89	H
A123	179.62	54.07	20.50	124.90	8.2140		0.00	0.26	0.74	H
A124	179.28	54.88	19.02	119.86	8.5039		0.00	0.04	0.96	H
Q125	180.31	58.88	28.74	119.04	8.1526		0.00	0.00	1.00	H
G126	174.41	46.06		106.97	7.8446		0.02	0.78	0.19	C
L127	176.55	53.87	43.07	120.42	7.4858		0.23	0.77	0.00	C
G128	173.73	45.42		107.80	7.7571		0.05	0.93	0.02	C
I129	174.03	60.18	39.27	122.98	7.9648		0.72	0.28	0.00	B
N130	174.04	53.25	39.50	125.14	8.3967		0.48	0.52	0.00	C
V131	174.45	60.71	33.94	123.88	7.9912		0.90	0.10	0.00	B
D132	177.00	53.91	42.20	127.07	8.6171		0.33	0.67	0.00	C
S133	176.82	61.15	62.96	122.72	8.9285		0.00	0.12	0.88	H
G134	176.51	46.32		111.69	8.7390		0.02	0.54	0.45	C
S135	174.86	61.00	63.11	119.26	7.8488		0.00	0.25	0.74	H
V136	177.75	66.45	31.78	122.42	7.5770		0.00	0.00	1.00	H
Q137	178.94	58.73	28.16	117.25	7.8983		0.00	0.00	1.00	H
S138	176.88	61.03	62.76	116.34	8.0227		0.00	0.03	0.97	H
D139	178.20	56.64	41.70	124.81	8.5481		0.02	0.08	0.90	H
I140	178.72	65.78	38.35	122.63	8.5293		0.00	0.00	1.00	H
S141	177.10	61.26	62.94	116.39	8.0279		0.00	0.02	0.98	H
S142	176.85	61.07	63.17	118.36	8.5210		0.00	0.08	0.92	H
C143	175.90	63.32	27.23	120.85	8.4831		0.00	0.20	0.79	H
S144	174.28	61.27	63.04	115.15	8.0963		0.00	0.46	0.54	H
S145							0.39	0.47	0.14	C
F146	177.33	60.16	39.50	122.38	8.1882		0.00	0.06	0.94	H
L147	177.43	55.93	41.98	119.14	7.9876		0.01	0.34	0.65	H
S148	175.23	59.21	63.90	114.37	7.8374		0.05	0.77	0.18	C
T149	174.72	62.24	69.86	115.28	8.0592		0.53	0.45	0.02	B
S150	174.36		63.99	117.48	8.1069		0.18	0.68	0.14	C
S151	174.57	58.23	63.99	117.55	8.0980		0.22	0.77	0.01	C
S152	174.45	58.17	64.04	117.92	8.3113		0.27	0.73	0.01	C
S153	174.13	58.15	64.02	117.89	8.2143		0.29	0.71	0.00	C
A154	177.44	52.35	19.36	126.10	8.1934		0.12	0.88	0.01	C
S155	173.98	58.03	63.98	114.68	8.0928		0.41	0.59	0.00	C
Y156	175.49	57.51	38.96	122.13	8.0101		0.36	0.61	0.04	C
S157	174.02	57.73	64.15	117.48	8.0752		0.44	0.56	0.00	C
Q158	175.53	55.54	29.63	122.66	8.2655		0.35	0.64	0.01	C
A159	177.56	52.39	19.37	125.24	8.2515		0.07	0.93	0.00	C

S160	174.17	57.86	64.14	115.14	8.1885		0.46	0.54	0.00	C
A161	177.47	52.35	19.46	126.24	8.2704		0.13	0.87	0.01	C
S162	174.47	57.95	64.09	114.71	8.2128		0.40	0.60	0.00	C
S163	174.66	58.07	64.09	117.88	8.2450		0.26	0.74	0.01	C
T164	174.39	61.38	69.84	115.36	8.1716		0.68	0.31	0.00	B
S165	174.64	58.13	64.13	117.88	8.2308		0.23	0.76	0.01	C
G166	173.57	44.82		111.24	8.3657		0.07	0.93	0.00	C
A167	177.80	52.15	19.46	124.03	8.1664		0.11	0.89	0.01	C
G168	173.37	44.71		108.24	8.3336		0.10	0.90	0.00	C
Y169	175.74	57.67	39.22	121.02	8.0435		0.46	0.50	0.04	C
T170	173.80	60.85	70.07	117.64	8.0955		0.86	0.14	0.00	B
G171	171.14	44.15		110.77	7.2845		0.35	0.65	0.00	C
P172	177.01	62.84	32.45				0.22	0.78	0.00	C
S173	174.16	57.83	64.28	116.38	8.4260		0.62	0.38	0.00	B
G174	171.33	44.17		110.82	8.1398		0.36	0.64	0.00	C
P175	176.94	62.72	32.44				0.24	0.76	0.00	C
S176	174.52	57.93	64.00	116.42	8.4665		0.54	0.46	0.00	B
T177	174.34	61.20	70.08	115.12	8.0884		0.68	0.32	0.00	B
G178	171.56	44.21		111.19	8.1714		0.31	0.69	0.00	C
P179	177.03	62.96	32.33				0.20	0.79	0.00	C
S180	174.61	58.12	64.06	116.03	8.4052		0.49	0.51	0.00	C
G181	172.99	44.59		110.50	8.2337		0.11	0.89	0.00	C
Y182	174.05	55.16	38.72	121.73	8.0303		0.64	0.35	0.00	B
P183	176.50	62.84	32.22				0.26	0.73	0.00	C
G184	171.40	44.16		108.60	7.5282		0.42	0.58	0.00	C
P185	177.14	62.86	32.33				0.20	0.80	0.00	C
L186	177.83	54.93	42.33	122.24	8.3877		0.20	0.77	0.03	C
G187	174.39	44.97		109.88	8.3205		0.07	0.93	0.00	C
G188	174.23	44.83		108.84	8.2114		0.07	0.92	0.01	C
G189	173.04	44.43		108.80	8.2264		0.15	0.85	0.00	C
A190	175.35	50.06	18.53	124.94	8.0854		0.72	0.28	0.00	B
P191	176.28	62.69	32.03				0.30	0.70	0.00	C
F192	176.07	57.43	39.66	120.08	8.1684		0.32	0.57	0.11	C
G193	173.72	44.88		111.02	8.2015		0.08	0.92	0.00	C
Q194	175.77	55.33	29.83	120.08	8.1175		0.27	0.73	0.01	C
S195	174.54	58.18	64.07	117.22	8.3978		0.28	0.71	0.01	C
G196	173.49	44.73		110.79	8.3205		0.08	0.92	0.00	C
F197	175.87	57.40	39.85	120.40	8.1403		0.21	0.70	0.09	C
G198	173.17	44.83		112.01	8.3634		0.12	0.88	0.00	C
G199	178.51	45.48		114.95	7.4895		0.02	0.74	0.24	C

b) stapled W_1M

Seq	C	CA	CB	N	H	HA	Beta sheet probability	Coil probability	Helix probability	SecStr
A1	174.0631	51.9519	19.3168				0.83	0.17	0.00	B
G2	171.7441	44.5217		108.8337	8.5038		0.31	0.69	0.00	C
P3	177.3306	63.2646	32.0650				0.14	0.84	0.01	C
Q4	176.5383	56.0365	29.4506	121.2181	8.5629		0.20	0.78	0.02	C
G5	174.4013	45.3148		110.3300	8.3654		0.06	0.94	0.01	C
G6	174.0321	45.1150		108.6424	8.1457		0.07	0.91	0.01	C
F7	176.4679	58.1159	39.5284	120.1410	8.2119		0.07	0.62	0.30	C
G8	173.9542	45.3339		110.6514	8.3905		0.06	0.93	0.01	C
A9	178.1843	52.7017	19.3060	124.0582	8.1342		0.04	0.90	0.07	C
T10	175.2871	61.9183	69.8147	112.5504	8.1088		0.54	0.43	0.03	B
G11	174.7389	45.4732		111.1357	8.3357		0.04	0.92	0.04	C
G12	174.4544	45.3666		108.9524	8.2679		0.06	0.91	0.03	C
A13	178.1918	52.9771	19.3304	124.2377	8.2931		0.03	0.84	0.13	C
S14	174.8129	58.7623	63.5847	114.5962	8.1509		0.18	0.77	0.04	C
A15	179.9731	54.9337	18.3551	125.0319	8.1848		0.00	0.01	0.99	H
G16	176.1109	46.5076		106.3891	8.3569		0.02	0.44	0.54	H
L17	177.7937	57.6967	41.7342	123.3711	7.5105		0.00	0.03	0.97	H
I18	180.4272	64.6358	38.1120	118.5797	7.5056		0.00	0.00	1.00	H
S19	175.7496	61.6854	62.7619	116.1030	8.0997		0.00	0.04	0.96	H
R20	179.6812	59.7779	30.5017	123.2871	7.8870		0.00	0.00	1.00	H
V21	177.1378	66.6801	31.9445	120.4823	7.8540		0.00	0.00	1.00	H
A22	179.1743	55.8290	18.4010	121.9562	8.3193		0.00	0.01	0.99	H
N23	177.2713	56.1516	38.2714	116.1681	8.5834		0.00	0.03	0.97	H
A24	180.6547	54.8491	18.0612	121.2581	7.4585		0.00	0.00	1.00	H
L25	178.9783	57.3719	43.2331	116.7026	8.0384		0.00	0.03	0.97	H
A26	176.7060	54.6041	18.4446	120.6300	8.7712		0.00	0.41	0.59	H
N27	175.6469	52.8212	39.3714	113.2262	7.2369		0.02	0.97	0.01	C
T28	175.0430	61.0068	71.4249	110.0816	7.5346		0.37	0.63	0.00	C
C29	178.0345	57.8540	38.7684	122.5275	8.9817		0.15	0.31	0.55	H
T30	174.9483	67.3541	68.9567	116.6542	8.7674		0.00	0.02	0.98	H
L31	178.3146	59.4599	40.8285	119.2900	7.4206		0.00	0.01	0.99	H
R32	177.2721	58.5796	30.9258	113.8941	7.7699		0.00	0.34	0.66	H
T33	174.8031	64.6587	69.6190	109.3870	7.5858		0.04	0.76	0.20	C
V34	175.0191	65.3019	33.7861	118.8927	7.3495		0.02	0.96	0.02	C
L35	172.8883	52.2730	39.5866	119.2745	7.9641		0.04	0.96	0.00	C
R36	174.5278	53.7736	31.8718	117.1649	6.3894		0.28	0.72	0.00	C
T37	175.5017	65.1812	69.2194	117.0806	8.3245		0.01	0.03	0.96	H

G38	174.8973	45.0155		115.0586	8.8848		0.09	0.90	0.00	C
V39	174.9820	62.1725	32.3450	122.8727	8.0027		0.56	0.44	0.00	B
S40	175.0651	57.8765	64.5531	123.5851	8.4534		0.46	0.54	0.00	C
Q41	178.8336	59.4556	29.0791	123.9992	9.0457		0.00	0.03	0.97	H
Q42	178.8826	59.4344	28.2096	119.1361	8.5867		0.00	0.01	0.99	H
I43	177.6391	63.3469	36.3456	121.5839	7.4689		0.00	0.03	0.97	H
A44	179.2850	55.8413	18.5034	122.0513	8.5850		0.00	0.01	0.99	H
S45	176.2588	62.4310	62.9767	110.7194	8.2613		0.00	0.02	0.98	H
S46	176.6383	62.3433	63.4820	117.3496	7.6168		0.00	0.04	0.96	H
V47	177.4624	66.8560	31.4803	121.3800	8.6170		0.00	0.00	1.00	H
V48	177.0170	66.8736	30.9695	116.8663	8.2335		0.00	0.01	0.99	H
Q49	177.8130	60.2761	29.1572	119.8747	8.2187		0.00	0.01	0.99	H
R50	179.6946	59.0615	30.2084	117.6379	8.0190		0.00	0.00	1.00	H
A51	179.4474	55.5323	18.7922	122.5324	8.9902		0.00	0.04	0.96	H
A52	178.9229	55.5620	17.8609	119.5771	8.4814		0.00	0.02	0.98	H
Q53	179.3553	59.2786	28.0330	116.3273	7.9515		0.00	0.00	1.00	H
S54	176.8108	61.7337	62.8408	116.4911	8.2726		0.00	0.02	0.98	H
L55	178.2125	57.9737	41.5646	122.8665	8.2926		0.00	0.02	0.98	H
A56	179.6414	55.5441	18.6129	121.7545	8.1693		0.00	0.01	0.99	H
S57	177.0776	61.2614	62.9300	112.9912	7.9740		0.00	0.03	0.97	H
T58	175.7653	65.6053	69.0022	117.0391	8.0041		0.00	0.03	0.97	H
L59	177.3578	54.9691	43.6684	118.4562	7.9248		0.04	0.94	0.02	C
G60	175.0566	46.5525		109.0330	7.9753		0.03	0.82	0.15	C
V61	174.9053	59.0183	33.9194	112.3271	7.4677		0.74	0.26	0.00	B
D62	177.8243	54.7731	41.1767	120.2203	8.3542		0.07	0.67	0.26	C
G63	174.8778	46.8557		115.8128	8.8118		0.04	0.96	0.01	C
N64	177.8853	56.5527	37.9997	120.7860	8.3803		0.00	0.01	0.99	H
N65	176.7079	56.0727	38.4085	119.7211	8.4118		0.00	0.05	0.95	H
L66	177.8749	58.6611	41.4088	119.4108	7.9536		0.00	0.02	0.98	H
A67	178.6625	55.7340	18.3347	119.2186	8.5145		0.00	0.04	0.96	H
R68	179.4639	59.9194	29.6906	117.9485	7.7771		0.00	0.00	1.00	H
F69	179.1773	61.7082	39.1682	117.0358	8.0122		0.00	0.01	0.99	H
A70	178.2797	55.8430	17.7165	123.4653	8.6987		0.00	0.07	0.93	H
V71	179.2464	66.8169	31.5806	118.2040	8.7665		0.00	0.00	1.00	H
Q72	178.6648	59.3257	28.3891	120.8262	8.1089		0.00	0.00	1.00	H
A73	180.8510	55.1571	18.8474	120.3513	7.8289		0.00	0.00	1.00	H
V74	178.3017	65.0658	31.6421	115.6765	8.1094		0.00	0.02	0.98	H
S75	174.4841	61.3730	63.0743	117.4826	8.3253		0.00	0.28	0.72	H
R76	176.5185	55.7118	30.7157	117.8247	6.9604		0.12	0.85	0.03	C
L77	175.4011	52.9357	40.4687	122.0532	7.3545		0.29	0.71	0.00	C
P78	177.5315	61.8963	32.4403				0.23	0.77	0.00	C

A79	178.8262	53.5023	18.1453	126.0587	8.7394		0.01	0.64	0.35	C
G80	175.1271	45.6900		110.8144	8.9406		0.05	0.93	0.02	C
S81	171.5409	61.4785	64.1042	117.3282	7.5251		0.04	0.91	0.05	C
D82	176.6393	51.9892	43.5902	121.3917	8.2875		0.45	0.55	0.00	C
T83	176.3661	67.3346	69.6100	115.1541	8.3821		0.00	0.00	1.00	H
S84	175.9295	62.5613		116.1469	8.6139		0.00	0.07	0.93	H
A85	180.8449	54.9744	18.5940	126.1960	7.7904		0.00	0.01	0.99	H
Y86	178.1697	62.8906	38.6531	117.9407	7.5711		0.00	0.01	0.99	H
A87	179.5022	55.7643	19.1562	122.0786	8.8723		0.00	0.05	0.95	H
Q88	177.8611	59.3693	27.4669	119.4238	8.2189		0.00	0.02	0.98	H
A89	179.6413	55.1881	18.2981	123.2825	7.8783		0.00	0.02	0.98	H
F90	178.2807	60.6005	39.2428	117.1463	8.3995		0.00	0.03	0.97	H
S91	176.7683	61.7609	63.7316	112.0161	8.4183		0.00	0.03	0.97	H
S92	176.5422	62.5586		113.0862	8.6859		0.00	0.10	0.90	H
A93	178.6050	55.8856	17.8088	123.8354	7.9451		0.00	0.03	0.97	H
L94	179.4098	58.4930	42.3921	117.1143	8.1379		0.00	0.01	0.99	H
F95	180.2184	60.8947	38.9252	117.4631	7.7884		0.00	0.01	0.99	H
N96	176.6733	56.0763	37.4371	122.5210	9.1861		0.02	0.38	0.60	H
A97	177.4416	52.2350	19.5443	117.6712	8.1265		0.15	0.85	0.00	C
G98	174.2387	45.5202		105.4286	7.4983		0.04	0.93	0.03	C
V99	175.9324	64.8257	32.1196	119.3189	7.6069		0.01	0.43	0.56	H
L100	174.4484	51.4890	45.3415	113.2891	6.4200		0.00	1.00	0.00	C
N101	173.9104	52.1511	40.3657	114.9390	8.6677		0.26	0.74	0.00	C
A102	178.9035	55.1372	18.6612	119.0555	8.8684		0.00	0.12	0.88	H
S103	175.6694	60.0831	63.7849	111.7761	8.3914		0.01	0.43	0.56	H
N104	178.9679	51.9769	39.5306	119.1593	8.0110		0.08	0.56	0.36	C
I105	176.6631	64.1526	37.8372	121.5891	7.6650		0.00	0.07	0.93	H
D106	177.1781	56.8815	40.7236	120.3529	8.3308		0.00	0.07	0.93	H
T107	176.0743	63.1220	69.3356	111.2829	7.6294		0.05	0.48	0.47	C
L108	177.8581	54.8794	44.1638	122.4934	8.6637		0.30	0.68	0.01	C
G109	174.8885	48.9502		107.2514	7.8902		0.00	0.09	0.91	H
S110	176.8543	62.2136	62.8219	115.9148	8.6527		0.00	0.02	0.98	H
R111	179.1519	59.1097	30.2242	123.6448	7.8174		0.00	0.01	0.99	H
V112	177.0554	66.5530	31.4988	120.1486	8.2783		0.00	0.00	1.00	H
L113	178.7471	58.7782	40.5169	119.0158	8.5415		0.00	0.00	1.00	H
S114	176.7224	62.1047	62.7707	113.3667	8.0482		0.00	0.01	0.99	H
A115	180.9713	54.9264	18.2488	123.5954	7.8190		0.00	0.00	1.00	H
L116	178.5600	58.7317	41.9468	122.4580	8.6635		0.00	0.04	0.96	H
L117	179.3327	58.4429	40.9085	120.2741	8.3381		0.00	0.00	1.00	H
N118	178.1566	56.3620	38.1729	118.8000	8.1366		0.00	0.01	0.99	H
G119	175.8266	47.2125		109.5748	8.2010		0.01	0.31	0.68	H

V120	176.7902	67.0882	31.1559	124.1778	8.6412		0.00	0.01	0.99	H
S121	177.2883	62.1723		115.0324	8.3207		0.00	0.01	0.99	H
S122	177.5850	61.6094	62.8116	116.1645	8.4170		0.00	0.03	0.97	H
A123	179.5472	54.5688	20.2746	125.0589	8.2140		0.00	0.13	0.87	H
A124	179.3765	55.2323	18.7130	119.5662	8.3271		0.00	0.02	0.98	H
Q125	180.1994	58.9763	28.4784	118.0667	8.0390		0.00	0.00	1.00	H
G126	175.1577	46.5879		107.7260	8.0825		0.01	0.46	0.53	H
L127	177.1457	54.7428	43.1862	120.0386	7.5929		0.13	0.87	0.01	C
G128	174.4989	46.1383		108.3309	7.8080		0.03	0.89	0.08	C
I129	175.0752	60.9950	38.4963	120.3863	7.8313		0.40	0.59	0.00	C
N130	174.9969	53.2552	38.6302	124.4674	8.4773		0.28	0.72	0.00	C
V131	175.2717	61.6837	32.6396	120.8453	8.0980		0.63	0.37	0.00	B
D132	176.4221	54.1629	41.2896	123.5198	8.2638		0.18	0.82	0.00	C
S133	175.3859	59.1636	63.7136	117.9107	8.3135		0.09	0.74	0.17	C
G134	174.5234	45.7331		110.8473	8.4827		0.04	0.91	0.05	C
S135	174.7823	58.6555	63.9299	115.7371	8.0510		0.16	0.81	0.03	C
V136	176.3260	62.7764	32.4421	121.0742	7.9486		0.30	0.69	0.01	C
Q137	176.1656	56.2976	29.1628	122.9090	8.3099		0.09	0.88	0.03	C
S138	174.3019	58.8981	63.8915	116.3387	8.1628		0.15	0.80	0.05	C
D139	176.6786	54.4831	40.9374	121.6952	8.3031		0.11	0.88	0.01	C
I140	176.5902	61.7106	38.4914	119.7359	8.0170		0.15	0.56	0.29	C
S141	175.1033	59.5942	63.6530	118.3994	8.2382		0.07	0.74	0.19	C
S142	174.3268	58.7329	63.6546	116.7789	8.1684		0.18	0.80	0.02	C
C143	174.0814	55.1381	41.3652	119.1120	8.1505		0.74	0.26	0.01	B
S144	174.8664	58.4447	64.2900	116.9707	8.5226		0.37	0.62	0.01	C
S145	174.0388	58.7851	63.8359	119.1601	8.4830		0.20	0.78	0.02	C
F146	175.3466	57.9791	39.7191	121.6615	8.0830		0.25	0.59	0.16	C
L147	176.9849	55.0989	42.5319	122.9246	8.1111		0.34	0.65	0.02	C
S148	174.9158	58.3717	63.8924	116.7537	8.2040		0.21	0.77	0.02	C
T149	174.7415	61.8963	69.6269	115.3004	8.1321		0.57	0.42	0.01	B
S150	174.9682	58.6319	63.9444	117.8685	8.1892		0.13	0.81	0.05	C
S151							0.39	0.47	0.14	C
S152							0.39	0.47	0.14	C
S153	174.4025	58.5651	63.8160				0.25	0.74	0.01	C
A154	177.7128	52.7988	19.1748	126.0163	8.1982		0.05	0.90	0.04	C
S155	174.3189	58.4876	63.7813	114.7228	8.0946		0.28	0.71	0.01	C
Y156	175.8185	58.0928	38.7370	122.1623	8.0287		0.21	0.64	0.14	C
S157	174.3644	58.2627	63.9283	117.4299	8.0759		0.30	0.69	0.01	C
Q158	175.8617	56.0867	29.3795	122.6288	8.2601		0.18	0.76	0.06	C
A159	177.8833	52.8719	19.1539	125.1130	8.2390		0.03	0.95	0.02	C

S160	174.5192	58.3504	63.9129	115.0112	8.1703		0.31	0.68	0.01	C
A161	177.8247	52.8205	19.2335	126.1651	8.2574		0.05	0.90	0.05	C
S162	174.8114	58.4318	63.8851	114.6921	8.2028		0.27	0.72	0.02	C
S163	174.9944	58.5509	63.8748	117.8643	8.2734		0.16	0.81	0.03	C
T164	174.7390	61.8520	69.6574	115.2627	8.1414		0.57	0.42	0.01	B
S165	174.3832	58.6064	63.8714	117.8862	8.2397		0.18	0.80	0.03	C
G166	173.8814	45.2761		110.7108	8.2996		0.05	0.93	0.01	C
A167	178.1432	52.6980	19.2592	123.9945	8.1444		0.04	0.90	0.06	C
G168	173.7071	45.1898		108.1683	8.3187		0.07	0.92	0.01	C
Y169	176.0785	58.1568	39.0002	120.9577	8.0277		0.29	0.56	0.15	C
T170	174.1475	61.3472	69.8495	117.4972	8.0846		0.80	0.20	0.00	B
G171	171.5099	44.6426		110.7054	7.2706		0.22	0.78	0.00	C
P172	177.3035	63.3419	32.2068				0.14	0.85	0.01	C
S173	174.4946	58.3025	64.0614	116.3035	8.4134		0.48	0.52	0.01	C
G174	171.6942	44.6597		110.7675	8.1251		0.24	0.76	0.00	C
P175	177.2408	63.2251	32.1951				0.16	0.83	0.01	C
S176	174.8448	58.4101	63.7958	116.3690	8.4539		0.39	0.59	0.01	C
T177	174.6753	61.6723	69.8673	115.0945	8.0823		0.59	0.40	0.01	B
G178	171.9125	44.6834		111.1405	8.1593		0.21	0.79	0.00	C
P179	177.3520	63.4344	32.0892				0.13	0.85	0.02	C
S180	174.9454	58.5965	63.8775	116.0118	8.3950		0.34	0.63	0.02	C
G181	173.3216	45.0820		110.4544	8.2213		0.07	0.92	0.01	C
Y182	174.3946	55.6295	38.5442	121.6906	8.0207		0.60	0.40	0.00	B
P183	176.8028	63.3209	31.9928				0.18	0.81	0.01	C
G184	171.7632	44.5842		108.5403	7.5138		0.29	0.71	0.00	C
P185	177.4437	63.3062	32.1118				0.13	0.85	0.01	C
L186	178.1091	55.3868	42.1075	122.1244	8.3777		0.09	0.74	0.18	C
G187	174.6714	45.4244		109.5738	8.2652		0.05	0.94	0.01	C
G188	174.5525	45.3836		108.7637	8.1994		0.06	0.91	0.04	C
G189	173.3611	44.9018		108.7534	8.2101		0.10	0.89	0.01	C
A190	175.6980	50.5515	18.3422	124.8774	8.0761		0.58	0.42	0.00	B
P191	177.4379	63.3471	32.1013				0.13	0.86	0.02	C
F192	178.1125	55.3965	42.0992	121.8750	8.3425		0.72	0.28	0.00	B
G193	174.7286	45.4120		109.7941	8.3048		0.05	0.92	0.03	C
Q194	176.1005	55.8301	29.6169	120.0406	8.1126		0.14	0.83	0.03	C
S195	174.8638	58.6614	63.8808	117.1338	8.3830		0.17	0.78	0.04	C
G196	173.8318	45.2151		110.7057	8.3104		0.05	0.93	0.01	C
F197	176.1641	57.9213	39.6447	120.3245	8.1401		0.10	0.68	0.21	C
G198	173.4866	45.3094		111.9075	8.3497		0.09	0.91	0.00	C
G199	178.8479	45.9634		114.8893	7.4861		0.02	0.47	0.52	H

3-Resdue Scan:

AGPQGGFGAT GGASAGLISR VANALANTCT LRTVLRTGVS QQIASSVVQR AAQSLASTLG
VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL NASNIDTLGS RVLSALLNGV
SSAAQGLGIN VDSGSVQSDI SSCSSFLSTS SSSASYSQAS ASSTSGAGYT GPSGPSTGPS
GYPGPLGGGA PFGQSGFGG

Original Assigment and Probability vs Suggested Assignment and Probability

GFG(6-8) 39.05% GLG(126-128) 100%
SRV(19-21) 90.65% SVV(46-48) 100%
GVD(60-62) 76.22% GVL(98-100) 100%
QSD(137-139) 68.54% QSL(53-55) 100%
DIS(139-141) 68.54% LIS(17-19) 100%
SSF(144-146) 88.27% SNI(103-105) 100%
SFL(145-147) 67.69% NID(104-106) 100%
PFG(191-193) 44.14% PLG(185-187) 100%

# of selected frags: 197; # of confirmed frags: 189; CONA Score: 95.94%

4-Resdue Scan:

AGPQGGFGAT GGASAGLISR VANALANTCT LRTVLRTGVS QQIASSVVQR AAQSLASTLG
VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL NASNIDTLGS RVLSALLNGV
SSAAQGLGIN VDSGSVQSDI SSCSSFLSTS SSSASYSQAS ASSTSGAGYT GPSGPSTGPS
GYPGPLGGGA PFGQSGFGG

Original Assigment and Probability vs Suggested Assignment and Probability

SSFL(144-147) 67.69% SNID(103-106) 100%

# of selected frags: 196; # of confirmed frags: 195; CONA Score: 99.49%

5-Resdue Scan:

AGPQGGFGAT GGASAGLISR VANALANTCT LRTVLRTGVS QQIASSVVQR AAQSLASTLG
VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL NASNIDTLGS RVLSALLNGV
SSAAQGLGIN VDSGSVQSDI SSCSSFLSTS SSSASYSQAS ASSTSGAGYT GPSGPSTGPS
GYPGPLGGGA PFGQSGFGG

Original Assigment and Probability vs Suggested Assignment and Probability

# of selected frags: 195; # of confirmed frags: 195; CONA Score: 100.00%

6-Resdue Scan:

AGPQGGFGAT GGASAGLISR VANALANTCT LRTVLRTGVS QQIASSVVQR AAQSLASTLG
VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL NASNIDTLGS RVLSALLNGV
SSAAQGLGIN VDSGSVQSDI SSCSSFLSTS SSSASYSQAS ASSTSGAGYT GPSGPSTGPS
GYPGPLGGGA PFGQSGFGG

Original Assigment and Probability vs Suggested Assignment and Probability

# of selected frags: 194; # of confirmed frags: 194; CONA Score: 100.00%

Reference Offsets and Deviant Shifts

Detected reference offsets
CO: -0.36ppm CA: -0.49ppm CB: 0.23ppm N: 0.35ppm

Number of assignments: 938
Number of deviant assignments: 0

Number of suspicious assignments after ref-calibration when necessary: 1
T30 H: 10.6732

b) Stapled W_1M

CSI and Sec Structure (No_Deviant_Ref_Calibrated)

Seq	C	CA	CB	N	H	HA	Beta sheet probability	Coil probability	Helix probability	SecStr
A1	173.74	51.53	19.52				0.92	0.08	0.00	B
G2	171.42	44.10		108.86	8.5038		0.43	0.57	0.00	C
P3	177.01	62.84	32.27				0.21	0.78	0.00	C
Q4	176.22	55.62	29.65	121.25	8.5629		0.35	0.64	0.00	C
G5	174.08	44.89		110.36	8.3654		0.07	0.93	0.00	C
G6	173.71	44.70		108.67	8.1457		0.10	0.90	0.00	C
F7	176.15	57.70	39.73	120.17	8.2119		0.15	0.70	0.15	C
G8	173.63	44.91		110.68	8.3905		0.08	0.91	0.00	C
A9	177.86	52.28	19.51	124.09	8.1342		0.09	0.90	0.01	C
T10	174.97	61.50	70.01	112.58	8.1088		0.65	0.35	0.01	B
G11	174.42	45.05		111.17	8.3357		0.05	0.94	0.01	C
G12	174.13	44.95		108.98	8.2679		0.07	0.91	0.01	C
A13	177.87	52.56	19.53	124.27	8.2931		0.08	0.90	0.02	C
S14	174.49	58.34	63.78	114.63	8.1509		0.29	0.70	0.01	C
A15	179.65	54.51	18.56	125.06	8.1848		0.00	0.03	0.97	H
G16	175.79	46.09		106.42	8.3569		0.03	0.67	0.30	C
L17	177.47	57.28	41.93	123.40	7.5105		0.00	0.06	0.94	H
I18	180.11	64.22	38.31	118.61	7.5056		0.00	0.00	1.00	H
S19	175.43	61.27	62.96	116.13	8.0997		0.00	0.09	0.91	H
R20	179.36	59.36	30.70	123.32	7.8870		0.00	0.01	0.99	H
V21	176.82	66.26	32.14	120.51	7.8540		0.00	0.01	0.99	H
A22	178.85	55.41	18.60	121.99	8.3193		0.00	0.01	0.99	H
N23	176.95	55.73	38.47	116.20	8.5834		0.00	0.07	0.93	H
A24	180.33	54.43	18.26	121.29	7.4585		0.00	0.01	0.99	H
L25	178.66	56.95	43.43	116.73	8.0384		0.00	0.08	0.92	H
A26	176.39	54.18	18.64	120.66	8.7712		0.01	0.66	0.33	C
N27	175.33	52.40	39.57	113.26	7.2369		0.03	0.97	0.00	C
T28	174.72	60.59	71.62	110.11	7.5346		0.42	0.58	0.00	C
C29	177.71	57.43	38.97	122.56	8.9817		0.24	0.33	0.43	H
T30	174.63	66.93	69.16	116.68	8.7674		0.00	0.04	0.95	H
L31	177.99	59.04	41.03	119.32	7.4206		0.00	0.02	0.98	H
R32	176.95	58.16	31.13	113.92	7.7699		0.00	0.58	0.42	C
T33	174.48	64.24	69.82	109.42	7.5858		0.09	0.85	0.06	C
V34	174.70	64.88	33.99	118.92	7.3495		0.04	0.95	0.00	C
L35	172.57	51.85	39.79	119.30	7.9641		0.04	0.96	0.00	C
R36	174.21	53.35	32.07	117.19	6.3894		0.35	0.65	0.00	C
T37	175.18	64.76	69.42	117.11	8.3245		0.04	0.09	0.87	H

G38	174.58	44.60		115.09	8.8848		0.12	0.88	0.00	C
V39	174.66	61.75	32.55	122.90	8.0027		0.68	0.32	0.00	B
S40	174.75	57.46	64.75	123.62	8.4534		0.57	0.43	0.00	B
Q41	178.51	59.04	29.28	124.03	9.0457		0.00	0.06	0.94	H
Q42	178.56	59.01	28.41	119.17	8.5867		0.00	0.02	0.98	H
I43	177.32	62.93	36.55	121.61	7.4689		0.00	0.07	0.93	H
A44	178.97	55.42	18.70	122.08	8.5850		0.00	0.02	0.98	H
S45	175.94	62.01	63.18	110.75	8.2613		0.00	0.05	0.95	H
S46	176.32	61.92	63.68	117.38	7.6168		0.00	0.08	0.92	H
V47	177.14	66.44	31.68	121.41	8.6170		0.00	0.01	0.99	H
V48	176.70	66.45	31.17	116.90	8.2335		0.00	0.02	0.98	H
Q49	177.49	59.86	29.36	119.90	8.2187		0.00	0.02	0.98	H
R50	179.37	58.64	30.41	117.67	8.0190		0.00	0.01	0.99	H
A51	179.13	55.11	18.99	122.56	8.9902		0.00	0.08	0.92	H
A52	178.60	55.14	18.06	119.61	8.4814		0.00	0.04	0.96	H
Q53	179.04	58.86	28.23	116.36	7.9515		0.00	0.00	1.00	H
S54	176.49	61.31	63.04	116.52	8.2726		0.00	0.04	0.96	H
L55	177.89	57.55	41.76	122.90	8.2926		0.00	0.03	0.97	H
A56	179.32	55.12	18.81	121.78	8.1693		0.00	0.02	0.98	H
S57	176.76	60.84	63.13	113.02	7.9740		0.00	0.06	0.94	H
T58	175.45	65.19	69.20	117.07	8.0041		0.01	0.07	0.93	H
L59	177.04	54.55	43.87	118.49	7.9248		0.07	0.92	0.00	C
G60	174.74	46.13		109.06	7.9753		0.04	0.90	0.06	C
V61	174.59	58.60	34.12	112.36	7.4677		0.78	0.22	0.00	B
D62	177.50	54.35	41.38	120.25	8.3542		0.11	0.84	0.05	C
G63	174.56	46.44		115.84	8.8118		0.04	0.96	0.00	C
N64	177.57	56.13	38.20	120.82	8.3803		0.00	0.03	0.97	H
N65	176.39	55.65	38.61	119.75	8.4118		0.01	0.12	0.87	H
L66	177.55	58.24	41.61	119.44	7.9536		0.00	0.03	0.97	H
A67	178.34	55.31	18.53	119.25	8.5145		0.00	0.06	0.94	H
R68	179.14	59.50	29.89	117.98	7.7771		0.00	0.00	1.00	H
F69	178.86	61.29	39.37	117.07	8.0122		0.00	0.02	0.98	H
A70	177.96	55.42	17.92	123.50	8.6987		0.00	0.10	0.90	H
V71	178.93	66.40	31.78	118.23	8.7665		0.00	0.00	1.00	H
Q72	178.34	58.91	28.59	120.86	8.1089		0.00	0.01	0.99	H
A73	180.53	54.74	19.05	120.38	7.8289		0.00	0.01	0.99	H
V74	177.98	64.65	31.84	115.71	8.1094		0.00	0.06	0.94	H
S75	174.16	60.95	63.27	117.51	8.3253		0.01	0.48	0.51	H
R76	176.20	55.29	30.92	117.85	6.9604		0.21	0.78	0.01	C
L77	175.08	52.52	40.67	122.08	7.3545		0.33	0.67	0.00	C
P78	177.21	61.48	32.64				0.29	0.71	0.00	C

A79	178.51	53.08	18.35	126.09	8.7394		0.04	0.86	0.10	C
G80	174.81	45.27		110.84	8.9406		0.06	0.93	0.01	C
S81	171.22	61.06	64.30	117.36	7.5251		0.10	0.88	0.02	C
D82	176.32	51.57	43.79	121.42	8.2875		0.52	0.48	0.00	B
T83	176.05	66.91	69.81	115.18	8.3821		0.00	0.01	0.99	H
S84	175.61	62.14		116.18	8.6139		0.00	0.12	0.88	H
A85	180.52	54.55	18.79	126.23	7.7904		0.00	0.02	0.98	H
Y86	177.85	62.47	38.85	117.97	7.5711		0.00	0.02	0.98	H
A87	179.18	55.34	19.36	122.11	8.8723		0.00	0.08	0.92	H
Q88	177.54	58.95	27.67	119.45	8.2189		0.00	0.03	0.97	H
A89	179.32	54.77	18.50	123.31	7.8783		0.00	0.03	0.97	H
F90	177.96	60.18	39.44	117.18	8.3995		0.00	0.05	0.95	H
S91	176.45	61.34	63.93	112.05	8.4183		0.00	0.07	0.93	H
S92	176.22	62.14		113.12	8.6859		0.00	0.17	0.83	H
A93	178.28	55.47	18.01	123.87	7.9451		0.00	0.04	0.96	H
L94	179.09	58.07	42.59	117.14	8.1379		0.00	0.01	0.99	H
F95	179.90	60.47	39.13	117.49	7.7884		0.00	0.01	0.99	H
N96	176.35	55.66	37.64	122.55	9.1861		0.04	0.57	0.39	C
A97	177.12	51.81	19.74	117.70	8.1265		0.28	0.72	0.00	C
G98	173.92	45.10		105.46	7.4983		0.05	0.93	0.01	C
V99	175.61	64.41	32.32	119.35	7.6069		0.04	0.75	0.20	C
L100	174.13	51.07	45.54	113.32	6.4200		0.00	1.00	0.00	C
N101	173.59	51.73	40.57	114.97	8.6677		0.35	0.65	0.00	C
A102	178.58	54.72	18.86	119.09	8.8684		0.00	0.22	0.78	H
S103	175.35	59.66	63.98	111.81	8.3914		0.04	0.66	0.29	C
N104	178.65	51.56	39.73	119.19	8.0110		0.13	0.73	0.14	C
I105	176.34	63.73	38.04	121.62	7.6650		0.00	0.18	0.82	H
D106	176.86	56.46	40.92	120.38	8.3308		0.00	0.20	0.79	H
T107	175.75	62.70	69.54	111.31	7.6294		0.12	0.66	0.22	C
L108	177.54	54.46	44.36	122.52	8.6637		0.48	0.52	0.00	C
G109	174.57	48.53		107.28	7.8902		0.01	0.17	0.83	H
S110	176.53	61.79	63.02	115.94	8.6527		0.00	0.05	0.95	H
R111	178.83	58.69	30.42	123.67	7.8174		0.00	0.02	0.98	H
V112	176.74	66.13	31.70	120.18	8.2783		0.00	0.01	0.99	H
L113	178.43	58.36	40.72	119.05	8.5415		0.00	0.00	1.00	H
S114	176.40	61.68	62.97	113.40	8.0482		0.00	0.03	0.97	H
A115	180.65	54.51	18.45	123.63	7.8190		0.00	0.01	0.99	H
L116	178.24	58.31	42.15	122.49	8.6635		0.00	0.06	0.94	H
L117	179.01	58.02	41.11	120.30	8.3381		0.00	0.01	0.99	H
N118	177.84	55.94	38.37	118.83	8.1366		0.00	0.02	0.98	H
G119	175.51	46.79		109.60	8.2010		0.02	0.52	0.46	C

V120	176.47	66.67	31.36	124.21	8.6412		0.00	0.01	0.99	H
S121	176.97	61.75		115.06	8.3207		0.00	0.01	0.99	H
S122	177.27	61.19	63.01	116.19	8.4170		0.00	0.06	0.94	H
A123	179.23	54.15	20.47	125.09	8.2140		0.00	0.32	0.68	H
A124	179.06	54.81	18.91	119.60	8.3271		0.00	0.04	0.96	H
Q125	179.88	58.56	28.68	118.10	8.0390		0.00	0.00	1.00	H
G126	174.84	46.17		107.76	8.0825		0.02	0.68	0.30	C
L127	176.83	54.32	43.39	120.07	7.5929		0.22	0.78	0.00	C
G128	174.18	45.72		108.36	7.8080		0.04	0.93	0.03	C
I129	174.76	60.57	38.70	120.42	7.8313		0.52	0.48	0.00	B
N130	174.68	52.84	38.83	124.50	8.4773		0.36	0.64	0.00	C
V131	174.95	61.26	32.84	120.88	8.0980		0.73	0.27	0.00	B
D132	176.10	53.74	41.49	123.55	8.2638		0.23	0.77	0.00	C
S133	175.07	58.74	63.91	117.94	8.3135		0.16	0.78	0.05	C
G134	174.20	45.31		110.88	8.4827		0.05	0.94	0.02	C
S135	174.46	58.24	64.13	115.77	8.0510		0.25	0.74	0.01	C
V136	176.01	62.36	32.64	121.10	7.9486		0.44	0.56	0.00	C
Q137	175.85	55.88	29.36	122.94	8.3099		0.18	0.81	0.01	C
S138	173.98	58.48	64.09	116.37	8.1628		0.25	0.74	0.01	C
D139	176.36	54.06	41.14	121.73	8.3031		0.14	0.85	0.00	C
I140	176.27	61.29	38.69	119.77	8.0170		0.30	0.61	0.09	C
S141	174.78	59.17	63.85	118.43	8.2382		0.14	0.80	0.06	C
S142	174.01	58.31	63.85	116.81	8.1684		0.28	0.71	0.00	C
C143	173.76	54.72	41.57	119.14	8.1505		0.77	0.22	0.00	B
S144	174.55	58.02	64.49	117.00	8.5226		0.50	0.50	0.00	C
S145	173.72	58.37	64.04	119.19	8.4830		0.31	0.68	0.00	C
F146	175.03	57.56	39.92	121.69	8.0830		0.42	0.52	0.06	C
L147	176.66	54.68	42.73	122.95	8.1111		0.51	0.49	0.00	B
S148	174.60	57.95	64.09	116.78	8.2040		0.31	0.69	0.00	C
T149	174.42	61.48	69.83	115.33	8.1321		0.66	0.33	0.00	B
S150	174.65	58.21	64.14	117.90	8.1892		0.22	0.77	0.01	C
S151							0.39	0.47	0.14	C
S152							0.39	0.47	0.14	C
S153	174.08	58.15	64.02				0.37	0.63	0.00	C
A154	177.39	52.38	19.37	126.05	8.1982		0.12	0.87	0.01	C
S155	174.00	58.07	63.98	114.75	8.0946		0.40	0.60	0.00	C
Y156	175.50	57.67	38.94	122.19	8.0287		0.33	0.62	0.05	C
S157	174.04	57.84	64.13	117.46	8.0759		0.42	0.58	0.00	C
Q158	175.54	55.67	29.58	122.66	8.2601		0.32	0.67	0.01	C
A159	177.56	52.45	19.35	125.14	8.2390		0.06	0.93	0.00	C

S160	174.20	57.93	64.11	115.04	8.1703		0.43	0.56	0.00	C
A161	177.50	52.40	19.43	126.20	8.2574		0.12	0.88	0.01	C
S162	174.49	58.01	64.09	114.72	8.2028		0.39	0.61	0.00	C
S163	174.67	58.13	64.07	117.89	8.2734		0.25	0.74	0.01	C
T164	174.42	61.43	69.86	115.29	8.1414		0.67	0.33	0.00	B
S165	174.06	58.19	64.07	117.92	8.2397		0.28	0.72	0.01	C
G166	173.56	44.86		110.74	8.2996		0.07	0.93	0.00	C
A167	177.82	52.28	19.46	124.02	8.1444		0.09	0.90	0.01	C
G168	173.39	44.77		108.20	8.3187		0.10	0.90	0.00	C
Y169	175.76	57.74	39.20	120.99	8.0277		0.44	0.51	0.05	C
T170	173.83	60.93	70.05	117.53	8.0846		0.85	0.15	0.00	B
G171	171.19	44.22		110.74	7.2706		0.32	0.68	0.00	C
P172	176.98	62.92	32.41				0.21	0.78	0.00	C
S173	174.17	57.88	64.26	116.33	8.4134		0.61	0.39	0.00	B
G174	171.37	44.24		110.80	8.1251		0.34	0.66	0.00	C
P175	176.92	62.81	32.40				0.23	0.77	0.00	C
S176	174.52	57.99	64.00	116.40	8.4539		0.52	0.47	0.00	B
T177	174.36	61.25	70.07	115.12	8.0823		0.68	0.32	0.00	B
G178	171.59	44.26		111.17	8.1593		0.30	0.70	0.00	C
P179	177.03	63.01	32.29				0.20	0.80	0.00	C
S180	174.63	58.18	64.08	116.04	8.3950		0.48	0.52	0.01	C
G181	173.00	44.66		110.48	8.2213		0.10	0.90	0.00	C
Y182	174.07	55.21	38.74	121.72	8.0207		0.64	0.35	0.00	B
P183	176.48	62.90	32.19				0.26	0.74	0.00	C
G184	171.44	44.16		108.57	7.5138		0.41	0.59	0.00	C
P185	177.12	62.89	32.31				0.20	0.80	0.00	C
L186	177.79	54.97	42.31	122.15	8.3777		0.20	0.77	0.04	C
G187	174.35	45.00		109.60	8.2652		0.07	0.93	0.00	C
G188	174.23	44.96		108.79	8.1994		0.07	0.92	0.01	C
G189	173.04	44.48		108.78	8.2101		0.15	0.85	0.00	C
A190	175.38	50.13	18.54	124.91	8.0761		0.70	0.30	0.00	B
P191	177.12	62.93	32.30				0.20	0.80	0.00	C
F192	177.79	54.98	42.30	121.91	8.3425		0.74	0.26	0.00	B
G193	174.41	44.99		109.82	8.3048		0.06	0.93	0.01	C
Q194	175.78	55.41	29.82	120.07	8.1126		0.25	0.74	0.01	C
S195	174.54	58.24	64.08	117.16	8.3830		0.27	0.72	0.01	C
G196	173.51	44.80		110.74	8.3104		0.07	0.92	0.00	C
F197	175.84	57.50	39.84	120.35	8.1401		0.20	0.71	0.09	C
G198	173.17	44.89		111.94	8.3497		0.12	0.88	0.00	C
G199	178.53	45.54		114.92	7.4861		0.02	0.72	0.26	C

3-Resdue Scan:

AGPQGGFGAT GGASAGLISR VANALANTCT LRTVLRTGVS QQIASSVVQR AAQSLASTLG
VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL NASNIDTLGS RVLSALLNGV
SSAAQGLGIN VDSGSVQSDI SSCSSFLSTS SSSASYSQAS ASSTSGAGYT GPSGPSTGPS
GYPGPLGGGA PFGQSGFGG

Original Assigment and Probability vs Suggested Assignment and Probability

GVD(60-62) 81.43% GVL(98-100) 100%
SSC(141-143) 0.00% SSF(144-146) 100%

# of selected frags: 195; # of confirmed frags: 193; CONA Score: 98.97%

4-Resdue Scan:

AGPQGGFGAT GGASAGLISR VANALANTCT LRTVLRTGVS QQIASSVVQR AAQSLASTLG
VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL NASNIDTLGS RVLSALLNGV
SSAAQGLGIN VDSGSVQSDI SSCSSFLSTS SSSASYSQAS ASSTSGAGYT GPSGPSTGPS
GYPGPLGGGA PFGQSGFGG

Original Assigment and Probability vs Suggested Assignment and Probability

ISSC(140-143) 0.02% CSSF(143-146) 100%
CSSF(143-146) 9.65% ISSC(140-143) 100%

# of selected frags: 196; # of confirmed frags: 194; CONA Score: 98.98%

5-Resdue Scan:

AGPQGGFGAT GGASAGLISR VANALANTCT LRTVLRTGVS QQIASSVVQR AAQSLASTLG
VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL NASNIDTLGS RVLSALLNGV
SSAAQGLGIN VDSGSVQSDI SSCSSFLSTS SSSASYSQAS ASSTSGAGYT GPSGPSTGPS
GYPGPLGGGA PFGQSGFGG

Original Assigment and Probability vs Suggested Assignment and Probability

# of selected frags: 195; # of confirmed frags: 195; CONA Score: 100.00%

6-Resdue Scan:

AGPQGGFGAT GGASAGLISR VANALANTCT LRTVLRTGVS QQIASSVVQR AAQSLASTLG
VDGNNLARFA VQAVSRLPAG SDTSAYAQAF SSALFNAGVL NASNIDTLGS RVLSALLNGV
SSAAQGLGIN VDSGSVQSDI SSCSSFLSTS SSSASYSQAS ASSTSGAGYT GPSGPSTGPS
GYPGPLGGGA PFGQSGFGG

Original Assigment and Probability vs Suggested Assignment and Probability

# of selected frags: 194; # of confirmed frags: 194; CONA Score: 100.00%

Reference Offsets and Deviant Shifts

Detected reference offsets
CO: -0.32ppm CA: -0.42ppm CB: 0.20ppm N: 0.03ppm

Number of assignments: 931
Number of deviant assignments: 0

Number of suspicious assignments after ref-calibration when necessary: 0

APPENDIX F: *Data collected for ^{15}N relaxation NMR experiments at 18.8 T on two different magnets.*

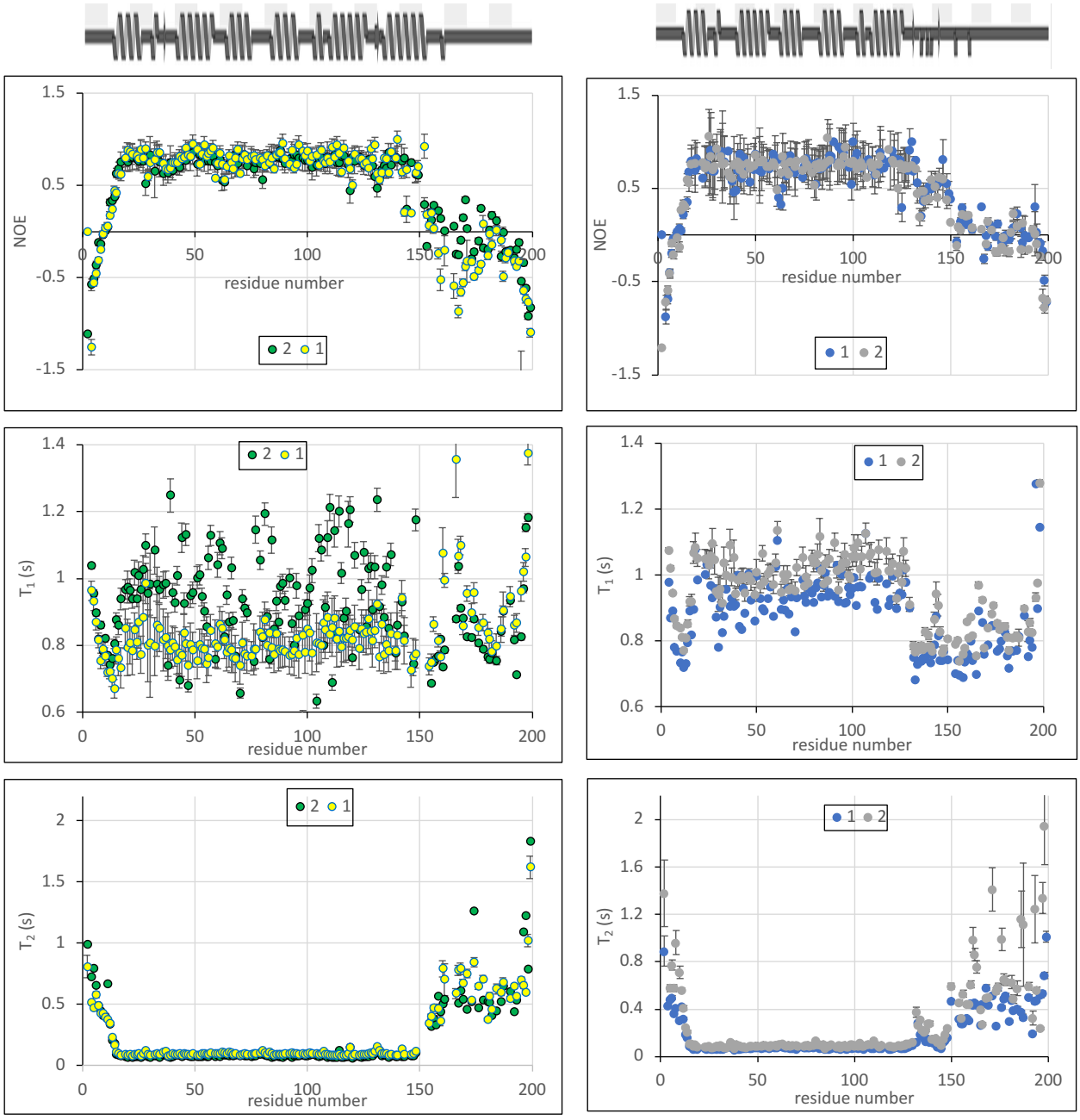


Fig. F.1. *^{1}H- ^{15}N NOE (top), ^{15}N-T_1 (middle)and ^{15}N-T_2 (bottom) relaxation time constants as a function of amino acid for reduced (left) and stapled (right) W_1M collected 18.8 T on two different magnets – 1: Kay lab, Toronto and 2: QANUC, Montreal, respectively. (Sample concentrations: reduced- 1) 500 µM and 2) 560 µM, and stapled – 1) 400 µM and 2) 580 µM.)*

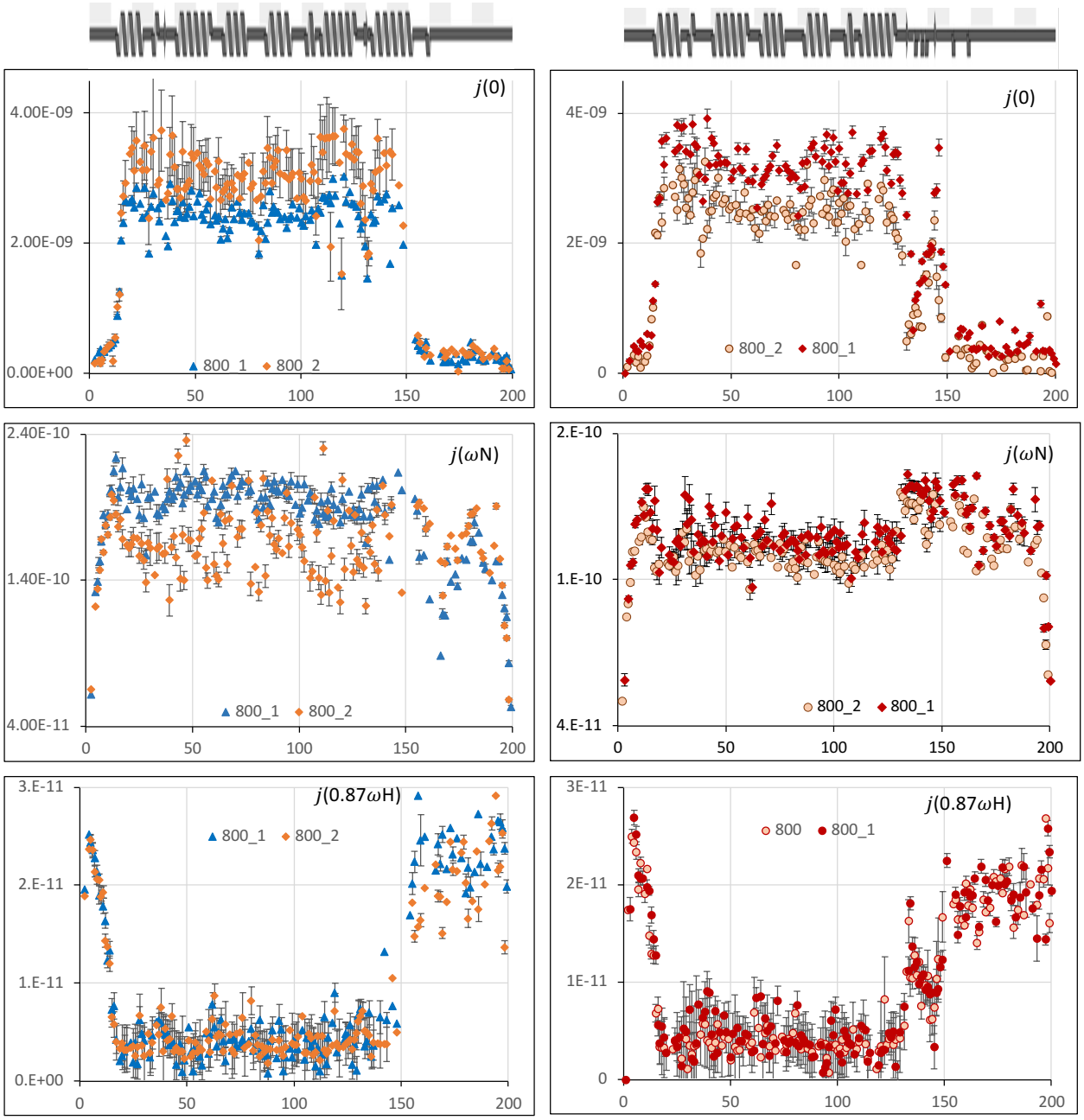


Fig.F.2. *Reduced spectral densities at zero (top), ω_N (middle) and $0.87\omega_H$* (bottom row) *frequencies as a function of amino acid for reduced (left) and stapled (right) W_1M collected 18.8 T on two different magnets – 1: Kay lab, Toronto and 2: QANUC, Montreal, respectively.(Sample concentrations: reduced- 1)500 µM and 2) 560 µM, and stapled – 1) 400 µM and 2) 580 µM.)*

***APPENDIX G**: Copyright licence agreements*

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS
Dec 13, 2021

This Agreement between Ms. Anamika Sulekha ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	5207070491315
License date	Dec 13, 2021
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Angewandte Chemie International Edition
Licensed Content Title	Spider Silk: From Soluble Protein to Extraordinary Fiber
Licensed Content Author	Markus Heim, David Keerl, Thomas Scheibel
Licensed Content Date	Apr 28, 2009
Licensed Content Volume	48
Licensed Content Issue	20
Licensed Content Pages	13
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	Disulphide locking: Contrasting effects on disparate proteins
Institution name	Dalhousie University
Expected presentation date	Dec 2021
Order reference number	theory1234
Portions	figure 4 on page 3590
Requestor Location	Ms. Anamika Sulekha Tupper medical building 10N1 Dalhousie university Halifax, NS B3H4R2

	Canada Attn: Dalhousie University
Publisher Tax ID	EU826007151
Total	0.00 CAD

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). 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 the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at http://myaccount.copyright.com).

Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order,** is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third-party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior

permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the STM Permissions Guidelines only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN

CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.

- 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 prevail.

- WILEY expressly 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.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

Creative Commons Attribution Non-Commercial License

The Creative Commons Attribution Non-Commercial (CC-BY-NC)License permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

Creative Commons Attribution-Non-Commercial-NoDerivs License

The Creative Commons Attribution Non-Commercial-NoDerivs License (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

Use by commercial "for-profit" organizations

ELSEVIER LICENSE
TERMS AND CONDITIONS
Dec 15, 2021

This Agreement between Ms. Anamika Sulekha ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5206831491115
License date	Dec 12, 2021
Licensed Content Publisher	Elsevier
Licensed Content Publication	Trends in Biochemical Sciences
Licensed Content Title	Modification in reverse: the SUMO proteases
Licensed Content Author	Debaditya Mukhopadhyay,Mary Dasso
Licensed Content Date	Jun 1, 2007
Licensed Content Volume	32
Licensed Content Issue	6
Licensed Content Pages	10
Start Page	286
End Page	295
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	Disulphide locking: contrasting effects on disparate proteins
Institution name	Dalhousie University
Expected presentation date	Dec 2021
Order reference number	SUMO1234
Portions	Figure 3.
Requestor Location	Ms. Anamika Sulekha Tupper medical building 10N1 Dalhousie university

	Halifax, NS B3H4R2 Canada Attn: Dalhousie University
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. 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).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier's permissions helpdesk here). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher 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.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher 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.

11. 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 publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher 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 publisher (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.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information

provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of

articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
    - via their non-commercial person homepage or blog
    - by updating a preprint in arXiv or RePEc with the accepted manuscript
    - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
    - directly by providing copies to their students or to research collaborators for their personal use
    - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
    - via non-commercial hosting platforms such as their institutional repository
    - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. **Other Conditions**:

v1.10

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

CCC | Marketplace™

Order Date	15-Dec-2021	Type of Use	Republish in a thesis/dissertation
Order License ID	1169162-1		
ISSN	0066-4170	Publisher	Annual Reviews
		Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Annual Review of Entomology	Publication Type	Journal
		Start Page	443
Article Title	Physicochemical Property Variation in Spider Silk: Ecology, Evolution, and Synthetic Production.	End Page	460
		Issue	1
		Volume	62
Date	01/01/1956		
Language	English		
Rightsholder	Annual Reviews, Inc.		

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Canada
Number of images / photos / illustrations	1	Translation	Original language of publication
Format (select all that apply)	Print, Electronic	Copies for the disabled?	No
		Minor editing privileges?	No
Who will republish the content?	Academic institution	Incidental promotional use?	No
Duration of Use	Life of current edition	Currency	CAD
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	Disulphide locking: Contrasting effects on disparate proteins	Institution name	Dalhousie University
		Expected presentation date	2021-12-17
Instructor name	Jan K Rainey		

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	Anamika Sulekha

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 1	**Title of the article/chapter the portion is from**	Physicochemical Property Variation in Spider Silk: Ecology, Evolution, and Synthetic Production.
Editor of portion(s)	Blamires, Sean J.; Blackledge, Todd A.; Tso, I-Min	**Author of portion(s)**	Blamires, Sean J.; Blackledge, Todd A.; Tso, I-Min
Volume of serial or monograph	62	**Issue, if republishing an article from a serial**	1
Page or page range of portion	443-460	**Publication date of portion**	2017-01-31

CCC Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

   3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

   3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

   3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of

use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4. In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5. Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. " Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7.

Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

8.1. User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2. Use of User-related information collected through the Service is governed by CCC's privacy policy, available online here:https://marketplace.copyright.com/rs-ui-web/mp/privacy-policy

8.3. The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.

8.4. No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5. The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.

v 1.1