STRUCTURAL CHARACTERIZATION OF MINOR AMPULLATE SPIDROIN DOMAINS AND THEIR DISTINCT ROLES IN FIBROIN SOLUBILITY AND FIBER FORMATION

GAO ZHENWEI

NATIONAL UNIVERSITY OF SINGAPORE

2013

STRUCTURAL CHARACTERIZATION OF MINOR AMPULLATE SPIDROIN DOMAINS AND THEIR DISTINCT ROLES IN FIBROIN SOLUBILITY AND FIBER FORMATION

GAO ZHENWEI

(B.Sc XMU)

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2013

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

gao zhenwei / 19 August 2013

GAO ZHENWEI
19 August 2013

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my supervisor Prof. Yang Daiwen for his continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my Ph.D study.

I would like to thank Associate Professor Mok Yu-Keung, Henry for his valuable suggestions during our group meetings. I am also grateful to Dr. Song jianxing, Dr. Ge Ruowen and Dr. Sivaraman for examining me with their critical and supportive suggestions during my PhD qualifying examination.

I highly appreciate Dr. Lin Zhi for his sincere help and guidance. Thank Dr. Fan Jingsong for his NMR training and advices on NMR experiments. Thank Dr. Wang shujing for her experimental training when I just join Prof. Yang's lab and her advices all through the four years. Thank Lai Chong Cheong for providing the sequences. Thank Liu Xiao for providing the software. Thank Dr. Zhang Jingfeng, Dr. Balakrishna Chandrababu Karthik, Dr. Zheng Yu, Dr. Lim Jack-wee, Liang Chen for their kindly helps in experiments, data analysis and software using.

I also want to express my gratitude to my fellow graduates, postdoctoral fellows and friends from structural biology labs: Dr. Long Dong, Dr. Zhou Zhiming, Dr. Hu Wentao, Dr. Tan Kang Wei, Dr. Nguyen Van Sang, Dr. Qin Haina, Dr. Huan Xuelu, Wang Wei, Yu Binhan, Lei Cheng, Xiao Tianshu, Rusha Chakraborty and Iman Fabim Hameed,

I take this opportunity to thank all my family members for their love and support.

At last, the research scholarship provided by National University of Singapore is gratefully acknowledged.

Table of Contents

Summary

Spider silk is protein fibers with extraordinary mechanical properties. Up to now, it is still poorly understood how silk proteins are kept in a soluble form before spinning into fibers and how the protein molecules are aligned orderly to form fibers. Minor ampullate spidroin is one of the seven types of silk proteins, which consists of four types of domains: N-terminal domain (NTD), C-terminal domain (CTD), repetitive domain (RP) and linker domain (LK). Current studies on MiSp only focus on gene sequence identification. To understand the mechanism of MiSp storage and its fiber formation, both structural and functional studies are necessary.

In this study, we characterized the tertiary structures of CTD and RP, secondary structures of NTD and LK in aqueous solution, and their distinct roles in protein stability, solubility and fiber formation. The stability and solubility of individual domains are dramatically different and can be explained by their distinct structures. In the multi-domain protein fragment, CTD cannot stabilize the entire protein against chemical and thermal denaturation, although the CTD itself possesses high stability. Nevertheless, both NTD and CTD have the effect on maintaining the entire multi-domain protein fragment in a highly water-soluble state. In the presence of shear force, protein aggregation is greatly accelerated and the aggregation rate is determined by the stability of folded domains and the solubility of the disordered domains. During the shear-induced aggregation process, only the longer protein fragments, RP-LK-CTD_{Mi} and NTD-LK-RP-LK-CTD_{Mi}, could form well defined silk-like fibers, indicating that different domains play distinct roles in fiber formation. Taken together, the functional roles of MiSp domains in silk protein storage and fiber formation were proposed. During protein storage, NTD and CTD could maintain the entire protein soluble at high concentration. Upon assembly, NTD serves as a salt sensor and start to oligomerize to gather more molecules. In the presence of shear force, with the assistance of RP domain, LK domain serves as a nucleation site to assemble different molecules

together, while CTD directs the alignment of molecules.

This study provides new insights into both structural and functional roles of MiSp domains and the mechanism of silk protein storage and fiber assembly, which can shed light on the studies on other types of spider silk proteins.

List of Figures

List of Tables

List of Abbreviations

1D/2D/3D/4D	One-/Two-/Three-/Four-dimensional
aa	Amino acid
AcSp	Aciniform Spidroin
ANS	8-anilinonaphthalene-1-sulphonic acid
cDNA	Complementary Deoxyribonucleic acid
CD	Circular Dichroism
CSI	Chemical Shift Index
CTD	C-terminal Domain
DLS	Dynamic Lights Scattering
DPC	Dodecyl Phosphocholine
E. coli	*Escherichia Coli*
ER	Endoplasmic Reticulum
HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	kilobases
kDa	kiloDalton
LB	Luria Bertani
LK	Linker
M(mM)	Mole/L (Milimole/L)
MaSp	Major ampullate Spidroin
min	Minute
MiSp	Minor ampullate Spidroin
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Enhancement
NOESY	Nuclear Overhauser Enhancement Spectroscopy
NTD	N-terimnal Domain
OD	Optical density

PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
ppm	parts per million
RMSD	Root Mean Square Deviation
SDS	Soduim dodecyl sulphate
T_1	longitudinal relaxation time
T_2	transverse relaxation time
TOCSY	Total Correlation Spectroscopy
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
RP	Repetitive Domain
TuSp	Tubuliform Spidroin
UV	Ultraviolet

1 Chapter 1: Introduction

The entire research in this project involved the characterization of structure-based functions of different domains of spider silk proteins in protein storage and fiber assembly. This thesis starts with a general introduction describing the background and the on-going research in the field of spider silk proteins.

1.1 Biological background of spider silk

Spider silk has received vast attention due to its outstanding mechanical properties since ancient times. However, only in the past two decades, an understanding of how the spider silk is formed and why the spider silk possesses such extraordinary properties has emerged. Limited by the cannibalistic and territorial behaviors of spider (Greenwood MJ 2010), farming spider is not practical which leads to that mass production of spider silk from nature is impossible. Thus, production of spider silk by recombinant biotechnology is one of the most promising alternatives (Heim, Keerl et al. 2009; Teule, Cooper et al. 2009; Spiess, Lammel et al. 2010). To achieve this goal, understanding the molecular structures, self-assembly mechanism and fiber formation of spider silk proteins is necessary. The biological background of spider silk will be reviewed in detail in the following sessions.

1.1.1 Different types of spider silks

Spider silk is protein-based nanomaterial (Lewis 2006). After millions of years of evolution, the female orb-weaving spiders obtain the ability to produce up to seven different spider silks with various tensile strength and elasticity(Heim, Keerl et al. 2009). These spider silks are produced in separate specialized glands and each of the silk fulfills a certain task (Vollrath 2000) (Figure 1.1).

The orb-weaving spiders live off the ground and utilize the orb-shaped web to capture prey. Thus, the web needs to be strong enough to stop the flying prey on the surface without breaking and sticky enough to immobilize it before the spider reaches it. This combination of design and silk properties has been shown to be nearly

perfect(Gosline, DeMont et al. 1986). Among the seven types of silks, the dragline silk, which is composed of two major ampullate spidroins (MaSp), possesses great toughness. Therefore, it provides shape and stability for the web by constructing the frame and radii of an orb web. In addition, it serves as the lifeline of spider, which is the reason why it is called dragline silk. Silk constituted by minor ampullate spidroins (MiSp) is used to form a temporary auxiliary spiral which supports the emerging web and serves as a template for the capture spiral(Romer and Scheibel 2008; Blackledge, Kuntner et al. 2011). The capture spiral is built by a single flagelliform silk which is strong enough to dissipate the impact of prey. The stickiness of the capture spiral comes from an additional adhesive protein, synthesized by the aggregate silk gland, coating the flagelliform silk. Silk proteins from the piriform gland act like “attachment cement” to connect the different silk types and adhere the web to the environment(Gosline, DeMont et al. 1986). The other two types of spider silks are utilized to protect the offspring: tubulliform spidroins (TuSp) form the tough outer layer of egg case to protect offspring from injury, while aciniform spidrions (AcSp) form a soft inner layer of egg case to provide further protection(Hayashi, Blackledge et al. 2004; Vasanthavada, Hu et al. 2007; Romer and Scheibel 2008).

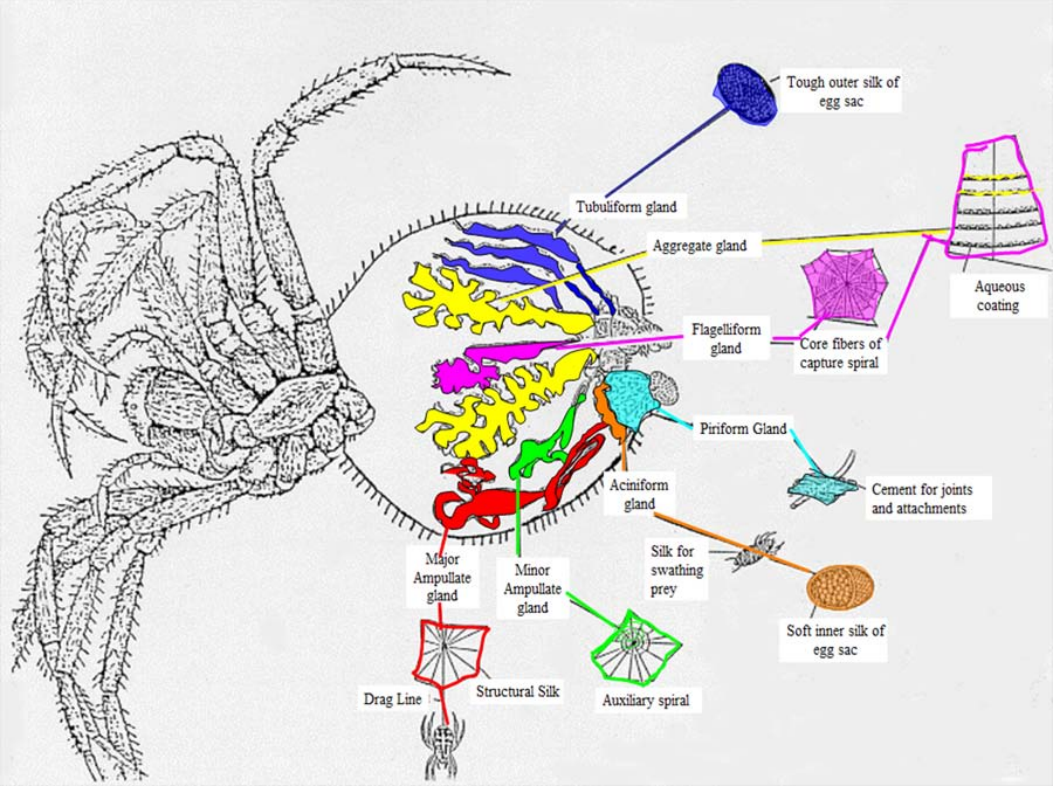


Figure 1.1 Spider silk glands and functions of spider silks(Lewis R.V., 2006).

Unlike the orb-weaving spiders, the non-orb-weaving spiders use fewer silks in different ways. For example, their silks could be used for above ground shelters, lining underground burrows and so forth.

1.1.2 Production of spider silk proteins

Spider silk proteins are synthesized in spider silk glands. The seven silk glands have different shape and size, but their functional organizations are similar (Figure 1.2). The majority of research attention has focused on the major ampullate gland due to its large size and ease of study. Thus, most of the knowledge about the mechanism of silk production is based on the study of this gland. Meanwhile, studies of other glands further support the insights gained from the major ampullate gland(Lewis 2006).

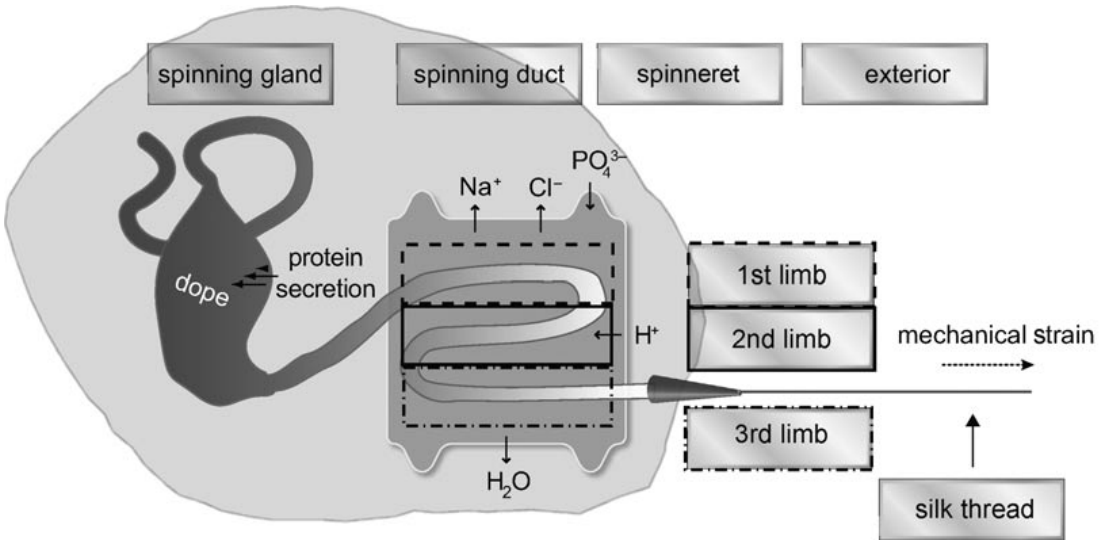


Figure 1.2 Diagram of major ampullate gland(Heim, Keerl et al. 2009)**.**

Although the genes encoding spider silk proteins are diverse, they all share similar gene configuration. One of the common features of this gene family is two non-repetitive terminal regions flanking an extremely repetitive core region constituted by consensus motifs. Nevertheless, the organization of the repetitive motifs varies significantly in different types of silk genes. This variation in gene sequence is believed to result in different mechanical properties of different type of silks. The detailed gene information of spider silk proteins will be reviewed in a later section (1.2). The tall columnar endothelial cells locating in the most upper part of a spider's silk gland is the place where translation from gene into spider silk proteins takes place(Vollrath and Knight 1999). The extensive endoplasmic reticulum (ER) and the large amount of secretory vesicles in the columnar endothelial cells facilitate the synthesis and transportation of spider silk proteins(Bell and Peakall 1969; Plazaola and C. Candelas 1991). In the case of major ampullate gland, there are at least two different types of columnar endothelial cells synthesizing major ampullate spidroins (MaSp1 and MaSp2). After being synthesized in the endothelial cells, MaSp1 and MaSp2 are secreted into the lumen of gland and stored there before spinning. Spider silk proteins could be stored in the lumen of the gland at an extremely high concentration without premature aggregation. For instance, MaSps

could remain soluble in the lumen when the protein concentration reaches up to 50% w/v(Hijirida, Do et al. 1996). Even today, the mechanism behind spidroin's high concentration storage is not thoroughly understood.

1.1.3 Assembly of silk protein

Upon usage, the silk proteins in the lumen of the gland pass through the spinning duct where they encounter both biochemical and physical environment changes and finally process into fibers (Figure 1.2). This assembly of fiber is definitely one of the most fantastic natural processes. With years of study on major ampullate gland of spiders, two theories (liquid crystal theory and micelle theory) have been developed to explain the assembly pathways of natural spider silk proteins (Figure 1.3).

1.1.3.1 Liquid crystal theory

According to the study of Vollrath and Knight, the state of freshly synthesized, rod-shaped spider silk protein is liquid-crystalline during storage in the lumen of gland. This state is believed to prevent premature fiber formation before passing through the duct. The long axes of the protein molecules are parallel to each other and vertical to the secreting epithelium(Vollrath and Knight 2001). Once the protein starts to move, the long axes turn until they become parallel to the epithelial walls. When the protein reach the second limb (Figure 1.2), minor stress force promotes the protein to organize in the cellular optical texture(Bunning and Lydon 1996), which means the protein molecules are arranged in bilayered disks with their long axes vertical to the plane of the disk (Figure 1.3). Meanwhile, a slight acidification in the second limb further facilitates the conformational change of the silk protein(Vollrath and Knight 1999; Terry, Knight et al. 2004; Dicko, Kenney et al. 2006). The negatively charged residues of silk protein are neutralized by the acidification. This could reduce the electronic repulsion force between protein molecules, which in turn promote hydrophobic interactions.

In the third limb, the random-coil and helix-like conformations of silk protein is

transformed to mainly β-sheet-rich structures under elongational flow and shear forces(Vollrath and Knight 1999). In addition, the epithelial cells in the third limb resorb water from the protein solution which additionally increases the protein concentration. The thin cuticle of the third limb additionally promotes this water resorbing process. Finally, the lips of the spigot at the end of the abdomen of spider remove the remaining water before the silk fiber exits the spider's abdomen.

1.1.3.2 Micelle theory

The micelle theory of silk assembly is established mainly based on in vitro experiments. The two common approaches used to investigate the silk assembly in vitro are regenerating silk protein solution by dissolving native silk fiber in certain solvents and expressing recombinant silk proteins with some model hosts like *Escherichia coli.*

It has been found that the fracture surfaces of native silkworm silk often exhibit globular structures in their core area(Jin and Kaplan 2003). In addition, Lin has observed the micelle structures in both recombinant tubuliform spider silk protein and native silk droplet isolated from tubuliform gland(Lin, Huang et al. 2009). Moreover, a sequence feature shared by silkworm and spider silk proteins is that short alternating hydrophobic and hydrophilic amino acid motifs are flanked by hydrophilic terminal domains. This type of amphiphilic sequence provides these molecules the ability to form micelles(Scheibel 2004; Exler, Hümmerich et al. 2007).

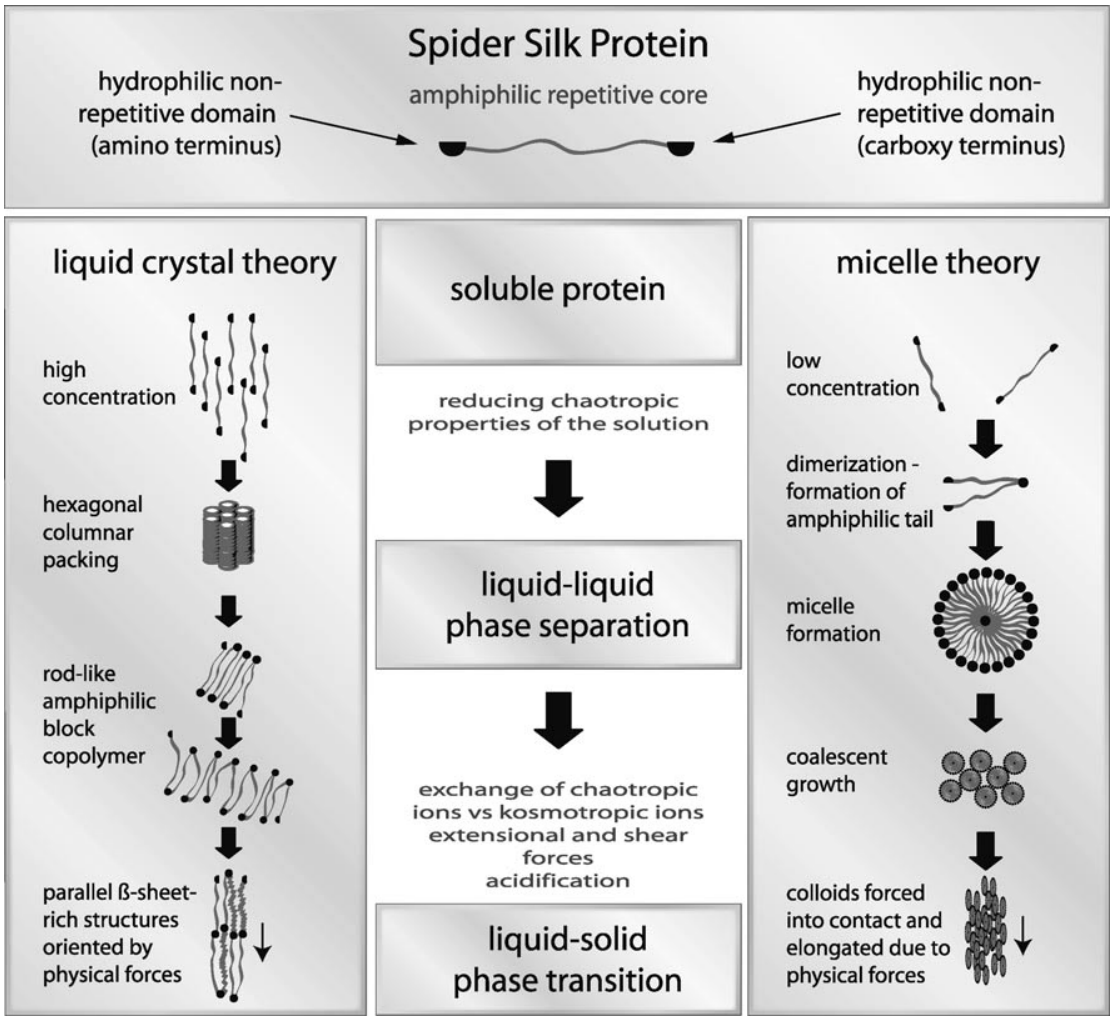


Figure 1.3 Two theories explaining spider silk formation(Vollrath and Knight 2001; Jin and Kaplan 2003)

Combining all the experimental data and sequence information, a model of silk assembly has been developed(Jin and Kaplan 2003). In this model, fibroin molecules in the gland form irregular sized micelles by hydrophobic interactions and chain folding (Figure 1.3). The hydrophilic blocks which are exposed on the surface of micelles could promote their solubility in water, and prevent premature aggregation. As protein concentration increases, micelles interact with each other to coalescence and form globules even gel states, which have been observed in the dissected glands. After experiencing the elongation force at the third limb (Figure 1.2), the globular structures are transformed into fibrillar morphologies, which might serve as the precursors of the spider silk fiber(van Beek, Hess et al. 2002).

As it passing through the spinning duct, the spider silk protein encounters

several environment changes. Besides the pH-shift, elongation and shear force mentioned in the liquid crystal theory, the ion strength change of protein solution in the first limb is noted (Figure 1.2). According to the study of major ampullate gland, sodium chloride is found in the storage dope. After entering the first limb, sodium chloride is removed from the protein solution and potassium phosphate is added into the solution(Heim, Keerl et al. 2009). Based on the studies of Hofmeister, chloride ion is “salting-in” ion which could stabilize soluble proteins, while phosphate ion is “salting-out” ion which could trigger protein aggregation and structure formation(CACACE, LANDAU et al. 1997; Slotta, Rammensee et al. 2008). Therefore, Hofmeister’s theory might be able to explain why spiders need sodium chloride in protein solution during storage and replace it with potassium phosphate upon fiber formation.

1.1.4 Mechanical properties of spider silks

The outstanding mechanical properties of spider silk are the main reason why spider silk could attract researchers. The web constructed by orb-weaving spiders is used to catch prey. It is necessary for the web to stop a fast flying insect almost instantly and trap it on the web without bouncing it away. To achieve this, the spider silks constructing the web have to absorb the energy of the prey without breaking.

Mechanical properties vary with the type of silk significantly (Table 1.1). As a structural silk (Figure 1.1) which holds the entire web, the major ampullate silk possesses the biggest tensile strength among all the seven spider silks(Lewis 2006). The energy used to break the major ampullate silk is ten times more than that of man-made synthetic fibers, such as Kevlar (Table 1.1). The flageliform silk is extremely elastic. It could be elongated more than two times without breaking(Spiess, Lammel et al. 2010). In addition, it has similar energy absorption as the major ampullate silk. All these mechanical properties of flagelifrom silk are necessary for the capture spiral to stop the flying insect without breaking. Minor ampullate silk has comparable tensile strength to the major ampullate silk but less elasticity(Colgin and

Lewis 1998). Due to this low elasticity, the minor ampullate silk cannot restore its initial formation when stretched.

Table 1.1 Comparisons of mechanical properties of spider silks and other materials(Gosline, Denny et al. 1984; Sponner, Unger et al. 2004; Lewis 2006)**.**

material	strength ($N\ m^{-2}$)	elongation (%)	energy to break ($J\ kg^{-1}$)
dragline silk	4×10^9	35	4×10^5
minor ampullate silk	1×10^9	5	3×10^4
fagelliform silk	1×10^9	>200	4×10^5
tubliform	1×10^9	20	1×10^5
aciniform	0.7×10^9	80	6×10^9
Kevlar	4×10^9	5	3×10^4
Rubber	1×10^6	600	8×10^4
tendon	1×10^6	5	5×10^3

Besides all the classical mechanical properties, the supercontraction when exposed to water is another unique feature of the major ampullate silk. When exposed to an environment with humidity greater than 60%, or directly contacted with water, the major ampullate silk starts to increase its diameter and reduce its length by almost 50%(Shao and Vollrath 1999; Shao, Vollrath et al. 1999). During supercontraction, the uptake of water allows rearrangement of hydrogen bonds between spider silk protein molecules(Guinea, Elices et al. 2003), resulting in altered mechanical properties of the silk. Through this process, the spider's web could regain its rigidity in the morning dew by tightening the attachment lines and the framework of the web.

1.2 Protein sequences

The extraordinary characteristics of spider silk make it an ideal candidate for a new class of biomaterial. Understanding the molecular building blocks of spider silk (proteins) is the first step to achieve the goal of artificially synthesizing spider silk fibers. The protein sequence information will be reviewed in detail based on silk types.

1.2.1 Major ampullate silk

In 1990, the first cDNA partial sequence of the major ampullate spider silk protein (MaSp1) from *Nephila clavipes* was published(Xu and Lewis 1990). Then, based on the identification of a proline-rich peptide in the silk, the second major ampullate silk protein (MaSp2) was identified(Hinman and Lewis 1992). The MaSps are huge proteins, ranging from 250-350 kDa(Scheibel 2004; Gaines Iv and Marcotte Jr 2008). Their core sequences are usually highly repetitive and consist of certain consensus repeating motifs. A_n, $(GA)_n$, GPGXX(X often representing Q), and GGX(X representing A, L, Q, Y) are the consensus motifs repeating in the core sequence of the MaSps(Scheibel 2004). Due to the difficulty of cloning long stretches of repetitive DNA, the first full-length sequences encoding MaSp1 and MaSp2 from *Latrodectus hesperus* was not identified until 2007(Ayoub, Garb et al. 2007). There are evidences showing that core sequence conservation has been maintained over millions of years(Guerette, Ginzinger et al. 1996; Gatesy, Hayashi et al. 2001). Because of the repetitive nature of relatively short consensus motifs, MaSps possess remarkably high content of certain types of amino acids (G, Q, A, P and S). The carboxy- and amino-terminal sequences of MaSps are non-repetitive but highly conserved(Beckwitt and Arcidiacono 1994; Sponner, Unger et al. 2004).

1.2.2 Minor ampullate silk

A few years after the identification of sequences of the major ampullate silk proteins, two cDNA partial sequences representing transcripts of minor ampullate silk proteins, MiSp1 and MiSp2, from *Nephila clavipes* were published(Colgin and Lewis 1998). The sizes of the two transcripts are 9.5 kb and 7.5 kb, respectively. Similar to MaSps, the MiSps are organized into many highly repetitive units connected to a relatively conserved non-repetitive C-terminal domain. In addition, poly A and GGX motifs are presented in the repetitive units. But unlike MaSp, each MiSp repetitive unit contains an additional relatively large domain that lacks repeats of short motifs(Colgin and Lewis 1998). This domain, which is highly conserved in sequence,

was initially called “spacer”. A and S are the most abundant residues in the spacer domains. The function of the spacer domains is currently unclear.

In 2012, the first full length gene sequence of MiSp from *Araneus ventricosus* was published(Chen, Liu et al. 2012). The full length transcript of this MiSp gene is around 5 kb in size, encoding around 1800 amino acid. This full length transcript is significantly smaller than the transcripts of MiSp from *Nephila clavipes.* Moreover, the number of spacer domain in this full length MiSp sequence has been reduced to two. In addition, a large intron was identified in this genomic DNA, whereas no introns had been detected in any other MiSp genes(Colgin and Lewis 1998; Huang, Lin et al. 2006).

Huang et al. previously reported a MiSp clone, clone 145, from the total silk gland cDNA library of *Nephila antipodiana* (Huang, Lin et al. 2006). The deduced amino acid (aa) sequence comprises one repetitive domain (RP_{Mi}, 128aa, previously named as spacer), one non-repetitive C-terminal domain (CTD_{Mi}, 107aa), and one linker domain (LK_{Mi}, 89aa, previously named as repetitive sequence) that links RP_{Mi} and CTD_{Mi} or in general links two structural domains (Figure 1.4). Recently, the N-terminal domain sequence of MiSp (NTD_{Mi}, 123aa) from *Nephila antipodiana* has been determined by one of our group members. The NTD_{Mi} and RP_{Mi} are also linked by the LK_{Mi} domain (Figure 1.4). RP_{Mi} and CTD_{Mi} are highly conserved among different spider species (Figure 1.5), but LK_{Mi}s vary significantly in the number of amino acids among different repetitive units in the same MiSp(Colgin and Lewis 1998; Bittencourt, Souto et al. 2007).

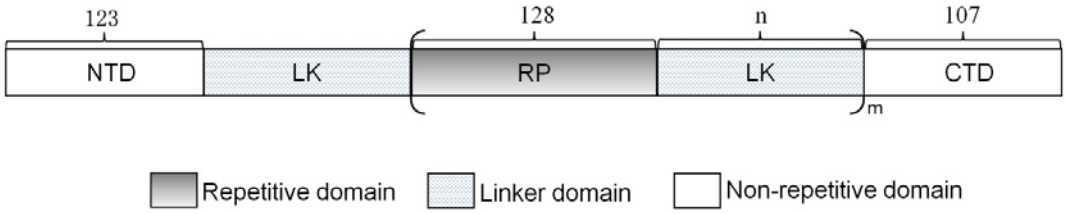


Figure 1.4 Molecular organization of partial MiSp. The amino acid (aa) number of each domain is indicated above the corresponding bar. The number of aa in LK domain (n) varies from 83-174. The total number of RP and LK in the MiSp sequence is still unknown and is denoted as “m”.

a)

b)

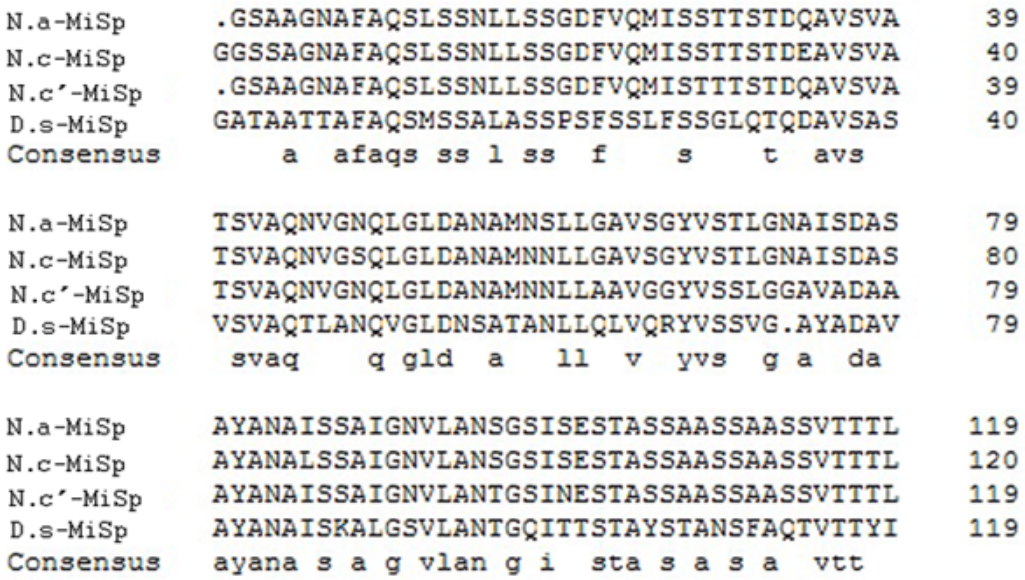


Figure 1.5 Sequence alignments of: C-terminal domains (a) of MiSps from *Nephila antipodiana (N.a), Nephila clavipes (N.c), Latrodectus Hesperus (L.h), Lephilengys cruentata (L.c) and Uloborus diversus (U.d)* and MaSp from *Araneus diadematus (ADF-3),* repetitive domains (b) from *N.a, N.c, Nephilengys cruentata (N.c') and Deinopis spinosa (D.s).*

1.2.3 Aciniform silk

The first partial sequence of aciniform silk protein was obtained from *Argiope trifasciata*(Hayashi, Blackledge et al. 2004). From the cDNA libraries of two sets of isolated aciniform glands, one long cDNA clone with 8618 bp in length was isolated and completely sequenced. The protein encoded by this transcript was called "aciniform spidroin 1" (AcSp1).

The partial AcSp1 sequence is comprised of 14 identical repetitive (RP_{Ac}, 200 aa) domains, ending with one non-repetitive C-terminal domain (99 aa). The BLAST search result showed that the C-terminal domain of AcSp1 is conserved with the published C-terminal domains of other spider silk spidroins. Although the RP domains are extremely conserved within AcSp1, homological sequence could not be found in either nucleotide or protein databases. The RP domain of AcSp1 is unique because it does not contain the repetitive motifs, like GA, GGX, GPGGX and poly A(Lewis 2006), which are abundant in other spidroins.

Another partial AcSp1-like sequence from *Latrodectus Hesperus* was later published(Vasanthavada, Hu et al. 2007). The RP domain in this partial AcSp-like sequence shares 45% sequence identity with the AcSp-RP domain. Later, the AcSp1-RP domain sequence of *Nephila antipodiana* was reported, which is 202 aa in length, containing one long structured region and one short disordered region(Wang, Huang et al. 2012). Surprisingly, this newly published RP domain sequence only shares 16-22% sequence identity with the two previously reported sequences.

Recently, the first complete gene sequence encoding AcSp1 from *Latrodectus Hesperus* was reported(Ayoub, Garb et al. 2013). In total, 16 RP domains ranging from 371-375 amino acid was identified and they shared >99% identity. Analysis of this full-length sequence further demonstrated that extreme homogeneity of intragenic repeats is a general feature of AcSp1.

1.2.4 Tubuliform Silk

Numerous tubuliform silk cDNA clones from orb-weaving spider superfamilies

have been isolated, the length of which differs from 10 to 13 kb in different species(Garb and Hayashi 2005; Huang, Lin et al. 2006). The deduced protein, tubuliform spidroin 1 (TuSp1), is the major component of tubulifrom silk, because the protein sequence could match the amino acid composition of tubuliform silk very well. In addition, TuSp1 is specifically expressed in the tubulifrom gland, evidenced by both immunoblot analyses and in situ hybridization(Huang, Lin et al. 2006).

The molecular architecture of TuSp1 is similar to other spider silk proteins. For example, TuSp1 from *Nephila antipodiana* contains 20 identical type 1 repetitive domains (RP1) and a type 2 repetitive domain (RP2) flanked by N- and C-terminal domains(Huang, Lin et al. 2006; Lin, Huang et al. 2009). However, the four common repetitive motifs found in MaSp and MiSp are rarely represented in TuSp1. Instead, the RP domains are composed of some new motifs such as S_n, $(SA)_n$, $(SQ)_n$, and GX (X represents Q, N, I, L, A, V, Y, F or D)(Lewis 2006).

1.3 Biophysical study

A great of effort has been put on biophysical study of spider silk protein, since understanding the structure of silk protein could provide important insights into spider silk protein storage, spinning process, and determination of final fiber properties. The biophysical studies could be divided into two groups based on the state of silk proteins.

The first group of studies focuses on the structure of silk protein after fiber formation (solid state). The data was mainly obtained from major ampullate silk because of its availability and the unique combination of elasticity and high tensile strength. Spider silks contain extended β-sheets oriented parallel to the fiber axis, demonstrated by nuclear magnetic resonance (NMR) studies and X-ray fiber diffraction measurements(Zhang and Rich 1997; Riekel, Branden et al. 1999). These β-sheets are formed by $(GA)_n$ and poly A motifs. Besides β-sheets, β-turns and 3_{10}-helices are represented in silk materials evidenced by solid-state NMR(Lazo and Downing 1999; van Beek, Kummerlen et al. 1999). Since similar types of crystalline

β-sheets are detected in minor, major ampullate silks, and silkworm silks, the tensile strengths of all the three silks are at the same level. The large amount of hydrogen bonds formed by β-sheets could contribute to these high tensile strengths. One possible reason for major ampullate silk possessing higher tensile strength is that the alanine residues in major ampullate could provide additional hydrophobic interactions(Colgin and Lewis 1998).

The second group of studies focuses on the structure of silk protein before fiber formation (solution state). Spider silk proteins before fiber formation mainly exist as α-helical and disordered conformations as confirmed by infrared spectroscopy and circular dichroism(Dicko, Knight et al. 2004; Rising, Hjalm et al. 2006). Until now, the mechanism of structure transition from α-helical and disordered conformations to mainly β-sheets structure during fiber spinning is still not thoroughly understood. Therefore recently, structural studies, especially three dimensional (3D) structural studies, on the structured regions of spider silk proteins in solution state have received great attention. Several 3D structures from three different types of spider silk proteins have been reported.

1.3.1 Structural studies of MaSp

The MaSp is the most intensively studied silk protein due to its unique mechanical properties. The high conservation of both N-terminal domain (NTD) and C-terminal domain (CTD) through evolution indicates the important roles of the two terminal domains in protein storage and fiber formation(Hagn 2012). Since the repetitive core sequence of MaSp is predicted to be disordered in solution, the structural study of MaSp mainly focused on the two terminal domains.

1.3.1.1 CTD

The importance of the CTD for maintaining the solubility of MaSp and directing fiber assembly has been realized for a long time. However, the 3D structure of this domain was not known until recently(Hagn, Eisoldt et al. 2010). Franz Hagn et al

solved the high resolution structure of CTD from *Araneus diadematus* fibroin 3 (ADF-3) (Figure 1.6) using three-dimensional triple resonance NMR experiments. As shown in Figure 1.6, the CTD is a homodimer which is formed by two disulfide-linked 5 helix monomers. In addition to the disulfide bond, two salt bridges in each monomer are proved to be crucial for the stability of this CTD. The interruption of salt bridges by low pH and mutations could increase the hydrophobicity of the CTD, which indicates the correctly folded CTD is important for maintaining the solubility and preventing premature aggregation(Hagn, Eisoldt et al. 2010). 8-anilinonaphthalene-1-sulphonic acid (ANS) binding experiment showed that the CTD will partially unfold and expose its hydrophobic surface under shear force, which will result in its oligomerization. Thus, CTD might act as a trigger for fiber assembly. In addition, the protein without CTD forms unordered aggregation under shear force, while the protein with CTD forms well defined fibers. This implies that CTD plays a role in directing the alignment of the β-sheets during fiber assembly.

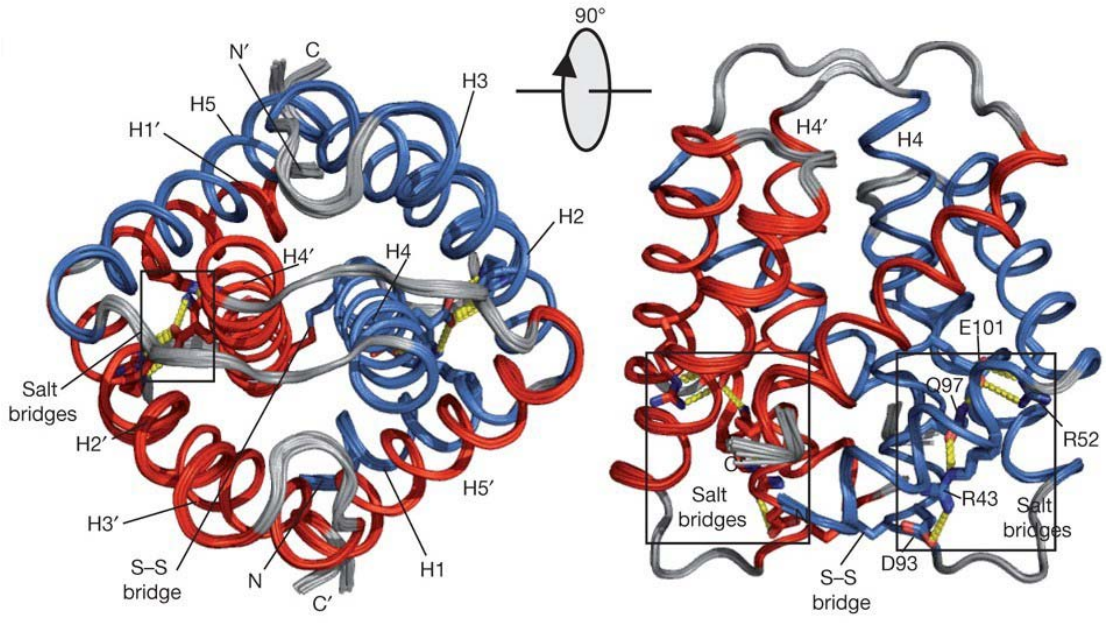


Figure 1.6 Overlay of the 20 best structures of the CTD of ADF-3(Hagn, Eisoldt et al. 2010)**.**

1.3.1.2 NTD

The function of NTD remained mysterious until Glareh Askarieh's work was published(Askarieh, Hedhammar et al. 2010). In this study, high resolution 3D structure of NTD of MaSp1 from *Euprosthenops australis* was determined by X-ray. Similar to CTD, NTD adopts a homodimeric formation (Figure 1.7). However, the dimerization of NTD is achieved mainly by electrostatic association between the two monomers instead of disulfide bond and hydrophobic interactions as found in CTD(Hagn, Eisoldt et al. 2010). Analysis of each monomer unit shows clustering of different charged residues at the opposite poles of the molecule which facilitates a antiparallel homodimer formation.(Askarieh, Hedhammar et al. 2010). The NTD existed as a dimer at pH 7.0 (as in the gland) and formed large aggregates at pH 6.3 (as in the duct), as demonstrated by dynamic light scattering, size-exclusion chromatography and so forth. However, later studies indicate NTD could form only dimers without large aggreagtes(Landreh, Askarieh et al. 2010; Hagn, Thamm et al. 2011; Jaudzems, Askarieh et al. 2012). The self-assembly speed of the construct with NTD is significantly increased when pH drops from 7 to 6.3, while the self-assembly speed of the construct without NTD is not affected by pH change. The charge distribution of NTD could be used to explain how the assembly process is accurately controlled by the intrinsic pH gradient of spider major ampullate silk gland.

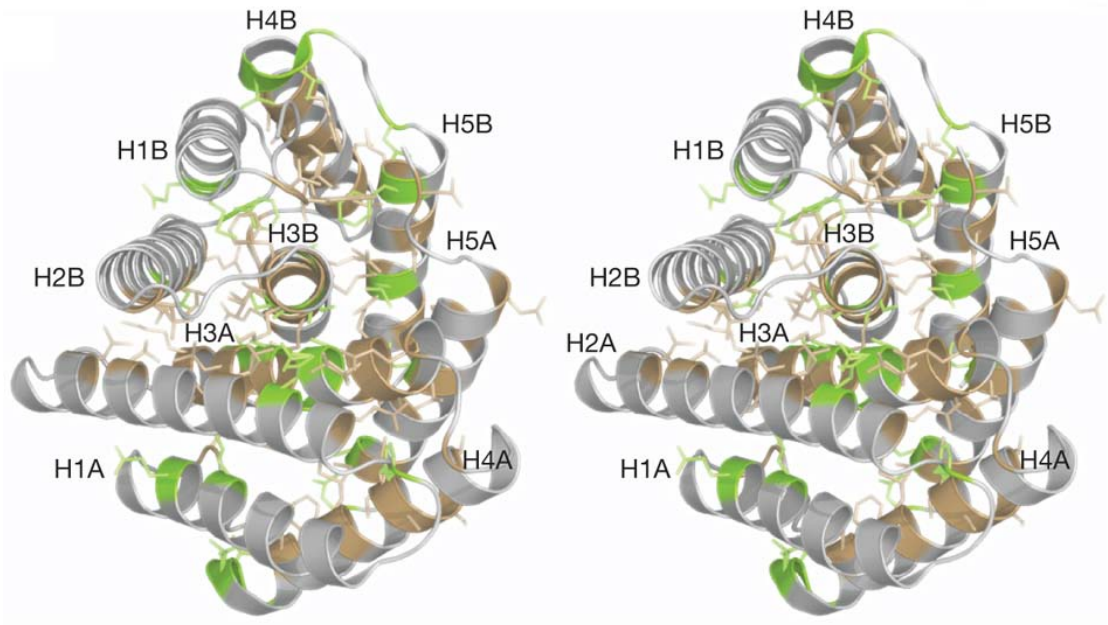


Figure 1.7 Ribbon diagram of the NTD dimer of MaSp from *Euprosthenops australis*(Askarieh, Hedhammar et al. 2010).

Later, two NMR structures of NTD of MaSp from *Latrodectus hesperus* and *Euprosthenops australis* were published(Hagn, Thamm et al. 2011; Jaudzems, Askarieh et al. 2012). Different from the X-ray dimerization structure, both of the two NMR structures of NTD are monomeric at ~pH 7 in the presence of 300 mM NaCl, with overall folding similar to that of the subunit of the dimer structure. In addition, several other biophysical studies of NTD all support the existence of monomers at pH 7 and of dimers at pH 6(Gaines, Sehorn et al. 2010; Landreh, Askarieh et al. 2010). Other than pH change, high salt concentration was also reported to affect the shifting NTD from dimer to monomer(Hagn, Thamm et al. 2011).

1.3.1.3 A proposed model for fiber assembly process

Taken together, Franz Hagn proposed a model for fiber assembly process in the gland of spider(Hagn 2012). When silk protein is stored in the gland where pH is neutral and the concentration of sodium chloride is high, it forms large micelles with the hydrophilic terminal domains exposed to the solvent and amphiphilic core sequence buried inside. Thus, this formation of micelle could protect the hydrophobic

patches of core sequence from the solvent and inhibit the formation of large aggregates. The monomeric form of NTD and correctly folded CTD at this stage are crucial for inhibiting premature aggregation and forming stable protein micelles. Moreover, the storage of silk protein could be further stabilized by the presence of sodium chloride which could enhance the stabilities of the two terminal domains.

Upon spinning, silk protein experiences drastical environment changes. The pH drops from 7 to ~6, the sodium chloride is replaced by potassium phosphate, protein concentration increases by removing water, and finally, shear force and elongational flow are presented. The dimerization of NTD at lower pH causes multivalent anchoring of single protein molecules. Under shear force, the CTD directs the assembly process by providing anchor points for the correct alignment of RP domains. Then, the elongational flow in the spinning duct of spider further aligns the RP domains and leads them to undergo a structural transition from disordered to β-sheet. By cooperating with each other, the two terminal domains could sense the environment changes and direct the fine-tuned fiber assembly process, which results in stable silk fiber with outstanding mechanical properties.

1.3.2 Structural studies of TuSp

The 3D structures of individual domains of TuSp1 (Figure 1.8) from *Nephila antipodiana* were solved by Lin et al(Lin, Huang et al. 2009). Unlike the terminal domains of MaSp, both NTD and CTD of TuSp1 form soluble aggregates in aqueous solution. Thus, the monomeric structures of two terminal domains of TuSp1 were determined in the presence of 100 mM dodecylphosphocholine (DPC) using NMR experiments. Both NTD and CTD adopt helix bundle folding with hydrophobic patches exposed on the surface (Figure 1.8). Different from MaSp, the repetitive domains (RP1 and RP2) of TuSp1 contains α-helical structure. The NMR structures of RP1 and RP2 show that they both adopt similar globular folding of six-helix bundle with hydrophilic surface (Figure 1.8). By introducing RP domains to the terminal domains, micelle-like structures were observed by transmission electron

microscopy. Dot blotting experiment was employed to investigate the organization of the micelles. Surprisingly, the result showed that the hydrophobic terminal domains formed the outer layer of the micelle and the hydrophilic RP domains were buried inside. Both the terminal domains and RP domains are needed to form micelles or macroscopic fibers, indicating that they all play important roles during protein storage and fiber assembly.

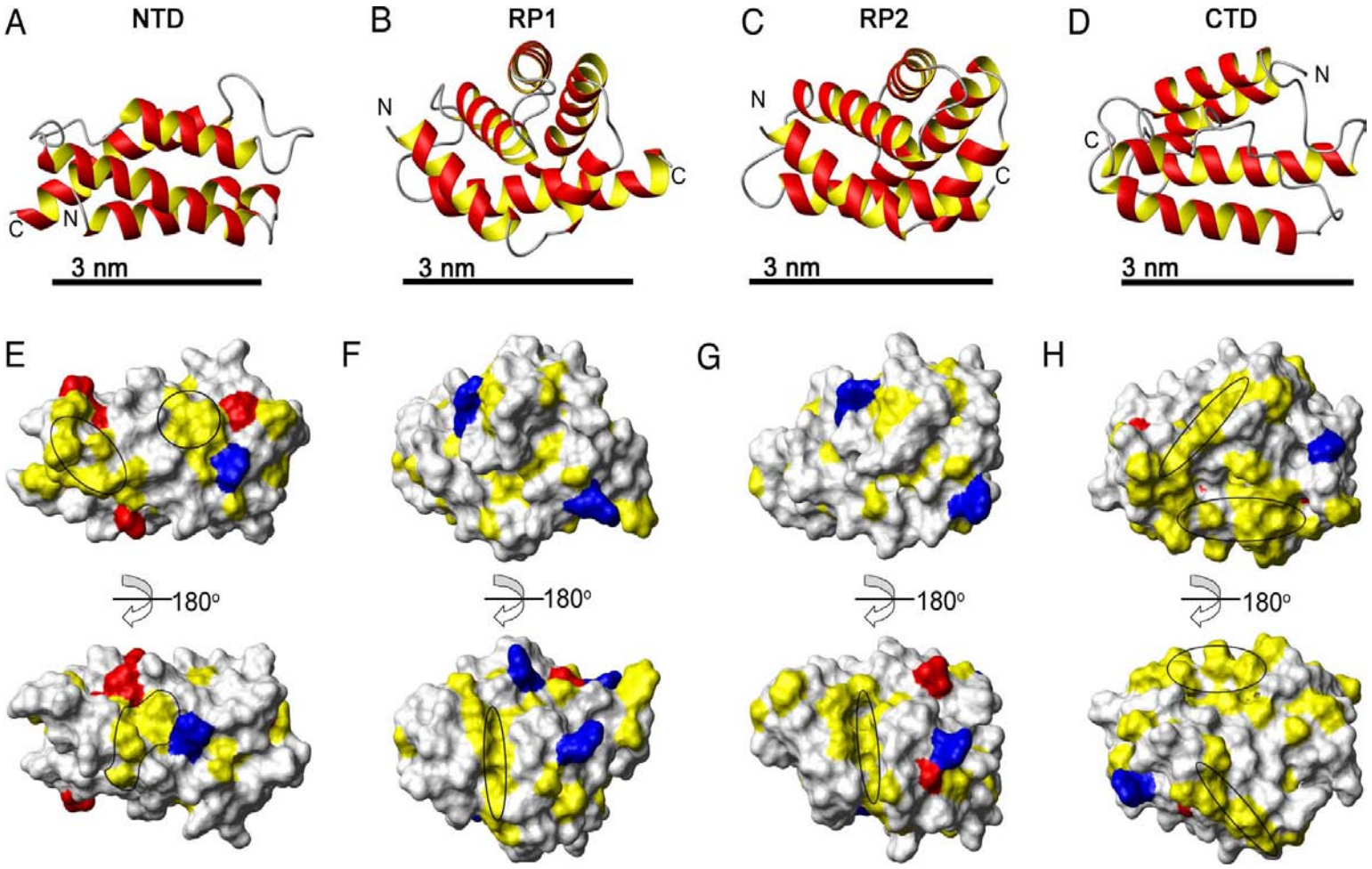


Figure 1.8 Ribbon drawing and surface plot of structures of individual domains of TuSp1. Color code of surface plot is blue for positive charges, red for negative charges, yellow for hydrophobic, white for neutral surface(Lin, Huang et al. 2009).

1.3.3 Structural studies of AcSp

The only 3D structural information of AcSp1available now is the RP domain from *Nephila antipodiana* solved by NMR experiments(Wang, Huang et al. 2012).The RP domain of AcSp1 adopts a globular folding of seven-helix bundle (Figure 1.9). The folding the first six helixes from the N-terminal are quite similar to that of the RP domains of TuSp1(Lin, Huang et al. 2009). The surface of this RP domain is extremely hydrophilic, and the hydrophobic interactions between the helixes are exceedingly strong(Wang, Huang et al. 2012) (Figure 1.9). This could be used to explain why the RP domain is highly soluble in aqueous solution and possesses high thermal stability. Interestingly, the single RP domain of AcSp1 could form well defined fiber under shear force, whereas the terminal domains are necessary for fiber formation in other studied silk proteins(Lin, Huang et al. 2009; Askarieh, Hedhammar et al. 2010; Hagn, Eisoldt et al. 2010).

Another NMR analysis showed that the RP domain of AcSp1 from *Argiope trifasciata* contains six major α-helixes(Xu, Tremblay et al. 2012).

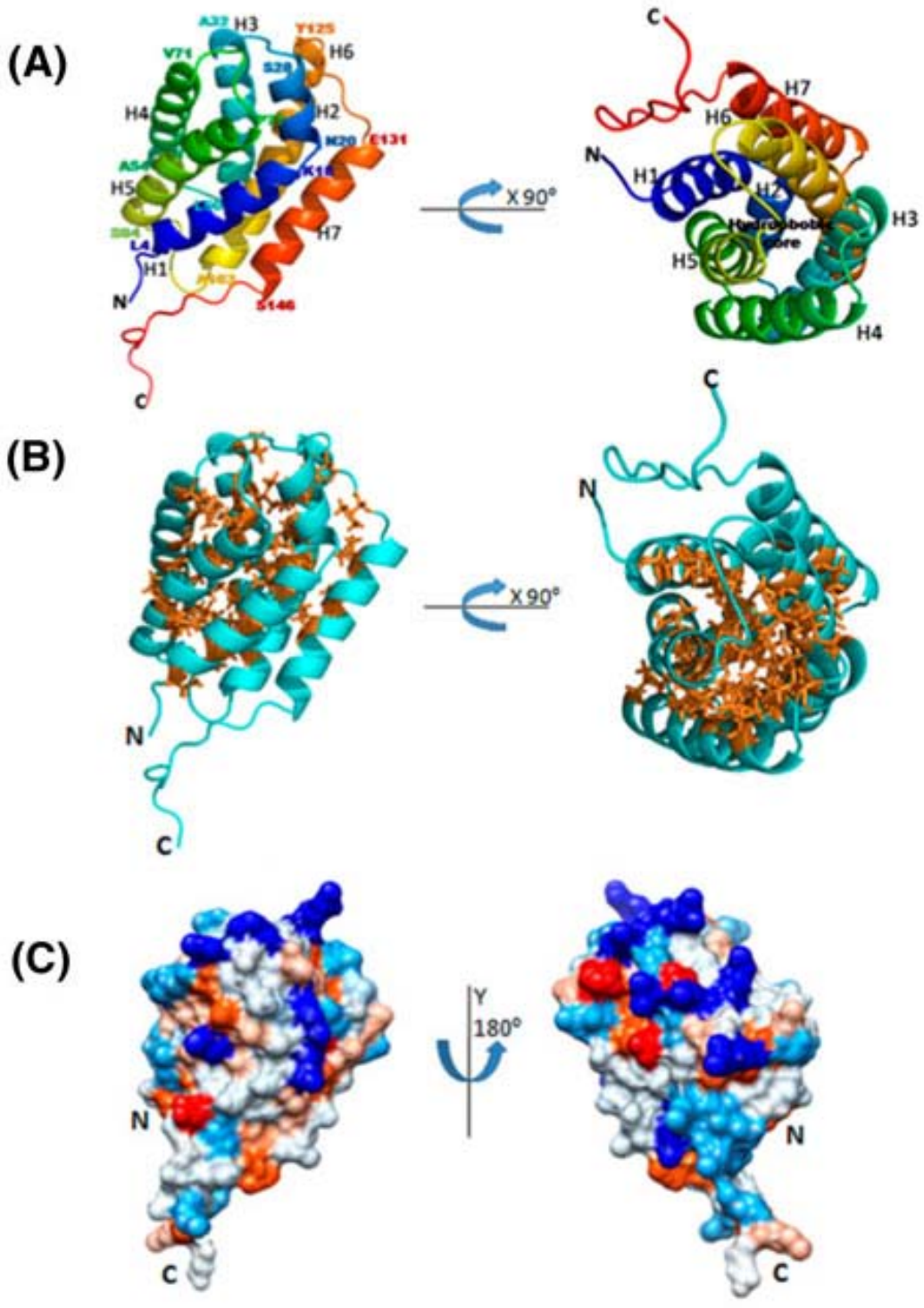


Figure 1.9 Sturcture of RP domain of AcSp1. (A) Ribbon drawing of the structure. (B) Stick view of side chains in the hydrophobic core. (C) Hydrophobicity surface plot: neutral (white), positively charged (dark blue), negatively charged (red), hydrophilic (less sky blue), hydrophobic (tan), and less hydrophobic (light tan)(Wang, Huang et al. 2012).

1.3.4 Summary

Among the seven types of spider silk protein, MaSp is the most well studied silk protein and a model of its storage and fiber assembly process has been proposed based on all the biophysical studies. However, the research on MaSp is far from enough to unravel the mystery of spider silks. Compared to MaSp, the knowledge of the other silk proteins are even less, especially MiSp which was studied in this work. The study presented here focused on characterizing the structures of MiSp domains and understanding their distinct roles in protein storage and fiber formation.

The technique used for structural characterization is NMR spectroscopy which has the advantage to study the structures of biomolecules in solution state. In the next section, a brief overview of principles of NMR and NMR structural determination is introduced.

1.4 Protein structural characterization by NMR

1.4.1 NMR phenomenon

The nuclear magnetic resonance (NMR) phenomenon was firstly reported by Bloch(Bloch 1946) and Purcell(Purcell, Torrey et al. 1946) independently. The ‘nuclear Overhauser effects’ (NOEs) observed by Overhauser could provide distance information between two protons, which is the basis for structure determination by NMR. In 1985, the first 3D structure of a protein in solution was determined by NMR spectroscopy(Williamson, Havel et al. 1985). Since then the use of NMR for structural and dynamics study of biomolecules in solution has greatly increased by improvements in NMR hardware and NMR methodology, along with the techniques for isotope labeling of protein samples. Some basic parameters of NMR are introduced in the following sections.

1.4.1.1 Chemical shift

The magnetic moment (nuclear spin) possessed by an atomic nucleus gives rise to its specific resonance frequency in a magnetic field. The resonant frequency is

affected by the electron distribution outside the nucleus. The variations in the electron distribution lead to the variations of nuclear magnetic resonance frequencies of the same type of nuclei, which are referred to as chemical shifts. With this parameter, signals for different protons in a molecule can be detected and differentiated. Some molecules with barely distorted electron distribution are selected as references, which means their chemical shift values are set as zero. The universal reference samples are tetramethylsilane for chemical compounds and 2,2-dimethy1-2-silapentanesulfonic acid for protein samples. The secondary structure of a given protein can be determined before solving its 3D structure using the difference of chemical shift values for spins in the protein from those for the same spins in a randomly coiled protein(Wishart, Sykes et al. 1991).

1.4.1.2 J-coupling

Scalar or J coupling results from the interaction between two spins connected through one or several covalent bonds. J coupling gives rise to signal splitting by altering the energy levels of the coupled spins. The connectivity information between two spins can be obtained through the values of J couplings because the coupling interaction decreases dramatically with the increase of the interconnecting bonds. In addition, J couplings are important for the development of multidimensional correlation experiments. In these experiments, the magnetization is transferred from one spin to another to identify spin systems, in which spins are connected by covalent bonds.

1.4.1.3 Relaxation

The term “relaxation” in NMR describes several processes by which a nuclear spin at a non-equilibrium state returns to its equilibrium state. The time used for the nuclear spin to restore its equilibrium state is called relaxation time. Longitudinal or spin lattice relaxation time (T_1) and transverse or spin-spin relaxation time (T_2) are the two common relaxation parameters. These relaxation times can provide useful

information regarding the geometry and dynamics of a given molecule because the relaxation is affected by the molecular weight, conformation and environment of the molecule.

1.4.1.4 NOE

Two dipolar coupled spins can transfer energy through space from one spin to another when they are cross-relaxing in space. This phenomenon is called NOE. The intensity of NOE could be used for structure determination of a molecule because it directly related to the distance between two spins (NOE ~ $1/r^6$; where r is the distance between the two spins). Normally, NOEs are too weak to be observed when the separation of two spins is bigger than 5Å. The distances converted from NOE intensities are used as restraints for structure calculation.

1.4.1.5 The advantage and disadvantage of structural studies by NMR

NMR spectroscopy is a great method for structural studies and it has its own advantages and limitations compared to X-ray crystallography, the other major biophysical technique which can provide high-resolution structures of biomolecules. The first advantage of NMR is that it can study the structures of biomolecules in solution state which is close to their physiological conditions. Since structural study of biomolecules is always related to their functions, structure determination under their physiological conditions is important. Conversely, the conditions such as temperature, pH and salt concentration can also be adjusted to some non-physiological conditions to study how these factors could affect the structure. Another advantage of NMR is that many dynamic processes like folding/unfolding, ligand binding, conformational change, protein/protein interactions can be investigated using NMR. In addition, the effects of crystal packing during crystallization do not exist in NMR since the sample is always in solution state.

Nevertheless, size limitation is the biggest disadvantage of NMR. When the size of a molecule is larger than 60 kDa, the NMR signals decay too fast to be recorded in

a multidimensional manner, which leads to the failure of structure determination. Moreover, as a complete NMR data acquisition requires time from several days to a couple of weeks, the sample needs to be stable for such long time at relatively high concentration and temperature.

1.4.2 General strategy of structure determination by NMR

Compared to the early time when NMR was only used for small molecules' structure determination, nowadays NMR has been improved to possess the ability to solve the structure (global folds) of protein up to 82 kDa with the development of technology and methodology(Tugarinov, Choy et al. 2005). A general strategy is shared by structure determination of both small molecules and larger proteins. It include several stages like sample preparation, NMR data acquisition, resonance assignments, restraint collection, structure calculation followed by refinement and evaluation of structure. The following sections provide an overall introduction of these processes.

1.4.2.1 Sample preparation

The first step for protein structure determination by NMR is to prepare the protein sample. For the isotope-labeled recombinant protein, the gene encoding the target protein is cloned into a proper expression vector, which is transformed into a host like *E. coli* and yeast. Then, the host is cultured in medium which is enriched with two NMR active nuclei ^{15}N and ^{13}C for the purpose of resonance assignments. Some types of tags are normally attached to target proteins for the purpose of purification and they will be removed after purification if necessary. A final protein concentration of 0.5-1 mM and a volume of 300-500 μl are needed for the NMR sample. Buffer condition optimization is one of the most crucial steps for sample preparation. The main purpose of buffer optimization is to find a condition under which the protein could maintain stable and soluble for at least several days. Sometimes, the protein forms soluble aggregate which could affect the quality of the

NMR spectra. In this case, certain amount of detergent could be added to the buffer to inhabit aggregation. EDTA is often used to prevent protein degradation caused by contamination of protease. Low ionic strength and no metal contamination are also important to increase the quality of the NMR spectra. Finally, 5% of D_2O in the sample solution is necessary for deuterium lock to maintain the magnetic field constant.

1.4.2.2 NMR data acquisition

The following step is to record NMR spectra. To assign all the chemical shifts (^1H, ^{15}N, ^{13}C), a set of multidimensional NMR spectra are recorded. The distance information for structure calculation is obtained by assigning interproton NOEs from the ^{13}C- and ^{15}N-edited NOESY spectra. 3D NOESY spectrum is widely used for structure determination of protein smaller than 30 kDa. If the target protein forms homodimer in solution, an extra 3D ^{13}C, ^{15}N-filtered NOESY experiment is required to differentiate the inter-molecule NOEs from intra-molecular NOEs. 4D ^{13}C, ^{15}N-edited NOESY spectrum was used in the strategy developed by Yang *et al.* which extended the size limit of high resolution structure determination by NMR to 65 kDa without deuteration or selective labeling(Xu, Zheng et al. 2006). However, only the protein samples which can maintain its initial state for a long time are suitable for 4D experiments because a single 4D spectra acquisition will take around one week even at a very high magnetic field strength (800 MHz). After recording the spectra, the raw data have to be processed using Fourier transformation to get frequency domain data.

1.4.2.3 Resonance assignment

The most time consuming process during structure determination is assigning all the relevant peaks in the NMR spectra. Several steps are involved in determining a complete 3D structure. The first step is sequence-specific backbone assignment. HNCACB and CBCA(CO)NH are the two classic 3D spectra used for backbone assignment. 4D NOESY, 3D HNCA and HNCOCA can also be used for the backbone

assignment of large proteins. With the involvement of the 4D NOESY spectra which could significantly reduce the ambiguities of assignments, the backbone assignment becomes more accurate and faster with the NOESY-based strategy. Then, 3D MQ-CCH-TOCSY spectrum together with the 4D NOESY provides the information for side chain assignments of the protein. After assigning all the chemical shift values, assignments of interproton NOEs observed in either 3D or 4D NOESY are needed to obtain the distance information. The secondary structure of a protein could be defined using all the assigned chemical shift values.

1.4.2.4 Restraint collection

The intensities of the assigned NOEs are converted to distances to obtain the distance restraints. The NOEs are normally divided into three groups: intraresidue (i-j=0), sequential (|i-j|=1), medium range ($2 \leqq |i-j| \leqq 4$), and long range ($|i-j| \geqq 5$) NOEs. The sequential and medium range NOEs are important for the calculation of the secondary structure, especially for α-helix, whereas the long range NOEs are crucial for determining the folding of the secondary structures. In addition, TALOS program package could be used to obtain the predicted angular restraints (φ and ψ) by inputting the chemical shift values of C^{α}, C^{β}, H^{α}(Dicko, Knight et al. 2004). Moreover, hydrogen bonds are useful for structure calculation. To determine the hydrogen bonds, D_2O is added to a lyophilized protein sample and $^{1}H/^{15}N$ correlation spectra are recorded. The NMR signals of amide protons which are involved in hydrogen bonding remain in the spectra because they can be protected from exchange with D_2O, whereas the rest of amide proton signals disappear rapidly.

1.4.2.5 Structure calculation

Structure calculation is a process that converts an extended structure into a folded 3D structure using experimental restraints by certain types of computer programs like CYANA(Herrmann, Guntert et al. 2002) and CNS(Brunger, Adams et al. 1998). The software calculates 3D structures by performing a gradient energy

minimization in the torsional angle space. The resultant structures should fulfill most of the conformational restraints. Therefore, the number and quality of experimental restraints directly decide the quality of the calculated structure. The final result of NMR structure calculation is an ensemble of structures, all of which should converge to the same fold. The poor quality of a NMR structure could be indicated by the poor convergence. The poor quality of a NMR structure might result from insufficient number of restraints or incorrect restrains. Thus, after the initial structure has been calculated, one should go back to the spectra to assign more NOEs or correct the wrongly assigned NOEs until the final structures show a good convergence, which is called structure refinement.

1.4.2.6 Structural quality evaluation

The root-mean-square deviation (RMSD) of the atomic coordinates between the structures of the NMR ensemble is one of the most important parameters to evaluate the quality of a given NMR structure. The RMSD value of a good NMR structure should not be larger than 1 Å. In addition, the distance and angular restraints used for structure calculation should not be violated by more than 0.5 Å and 5° respectively. If the distance between two atoms in a structure is too short, it will cause bad contact. The number of bad contact should be small. Ramachandran plot is another good measurement to evaluate the quality of a NMR structure. It is divided into several regions, including favored, additionally allowed, generally allowed and disallowed regions, based on statistical analysis of high-resolution crystal structures. All the backbone angles could be located into certain regions of the plot. Most of the angles should be within the allowed regions. If not, further refinement of the structure is needed.

1.5 Objectives

In the previous sections, we have mentioned that there is no structural and functional study on MiSp. The most intensively studied silk protein is MaSp and a

model of its mechanism of fiber assembly has been proposed(Hagn 2012). The biophysical study of TuSp1 showed some similar results with that of MaSp, such as the formation of micelle structure and the inability to form well defined fiber without the terminal domains or RP domain(Lin, Huang et al. 2009; Askarieh, Hedhammar et al. 2010; Hagn, Eisoldt et al. 2010). However, some differences were also noticed like the terminal domains of TuSp1 are existed as soluble oligomers in aqueous solution while the terminal domains of MaSp are monomers or dimers(Lin, Huang et al. 2009; Hagn 2012). Interestingly, the single RP domain of AcSp1 from *Nephila antipodiana* possesses the ability to form well defined silk-like fiber(Wang, Huang et al. 2012). All these dissimilarities indicate that fiber assembly process is different for different type of silk. Therefore, in order to better understand the self-assembly and stability of different types of silk proteins, it is necessary to have the structures of individual domains of each type of silk protein.

Although the CTD_{Mi} from *Nephila antipodiana* and CTD_{Ma} from *Araneus diadematus* share 44% sequence identity (Figure 1.5), the CTD_{Mi} contains no cysteine residues while the CTD_{Ma} has one disulfide linkage between two molecules which can enhance the stability of CTD_{Ma}. Moreover, RP_{Mi} is unique to MiSps and its functional roles in protein storage and fiber formation are unknown. Besides the difference in amino acid sequences, MaSp silk is elastic when stretched and MiSp displays irreversible deforming. Thus, MiSp may adopt a different self-assembly and fiber formation mechanism than the well characterized MaSp.

This study covered 3D and secondary structural determination of different domains of MiSp (*Nephila antipodiana*) and their functional roles in spider silk protein storage and fiber formation. The major objectives of this study are to:

1. solve the 3D structure of CTD_{Mi} and RP_{Mi} using 4D NMR experiments,
2. study the secondary structure of LK_{Mi},
3. analyze NTD_{Mi} structure,
4. test solubilities of single domains and different combinations of domains,
5. test both thermal and chemical stabilities of single domains and different combinations of domains,

6. investigate the fiber formation of different constructs of MiSp.

The structural study of MiSp domains will fill the gap of unavailability of high resolution 3D MiSp molecular structure. In addition, the functional study of different domains of MiSp in both protein storage and fiber formation will shed light into the fundamental knowledge of structural-functional relationship. Moreover, the whole study of MiSp may provide more knowledge on spider silk protein and serves as guidelines for the study of other types of spider silk proteins.

The organization of this thesis is that the first chapter provides the general introduction about the spider silk protein and the objective of this study. The second chapter describes all the methodologies and materials used in this study. The third chapter presents results and discussion of this study. Finally, chapter four gives summaries and conclusions of the whole project.

2 Chapter 2: Materials and Methods

2.1 Molecular Cloning

2.1.1 Template DNA for cloning

Two template DNAs were used in this study for all the constructs. The first one is the MiSp fragment (clone 145) previously identified by Huang(Huang, Lin et al. 2006), which includes a RP domain and a CTD linked by a LK domain. The second template is the MiSp fragment containing the entire NTD, which was identified recently from genomic DNA of *N. antipodiana*. The derived amino acid sequences from these two fragments are shown as Figure 2.1

Clone 145

GAAASGAGAGGAGGYGRGAGAGAGAGAGSAAGNAFAQSLSSNLLSSGDFVQMISSTTSTDQAVSVATS
VAQNVGNQLGLDANAMNSLLGAVSGYVSTLGNAISDASAYANAISSAIGNVLANSGSISESTASSAAS
SAASSVTTTLTSYGPAVFYAPTSSAGGYGGLVGYGAGAGAAAGAGAGAGGAGGYIGQGGYGAGAGAAA
AAGAGAGATGGYGRGAGAGATNAGGYGGQGGYGAGARAFAGAGVGVGTTVASTTSRLSTAEASSRIST
AASTLVSGGYLNTAALPSVIADLFAQVGASSPGVSDSEVLIQVLLEIVSSLIHILSSSSVGQVDFSSV
GSSAAAVGQSMQVVMG

NTD

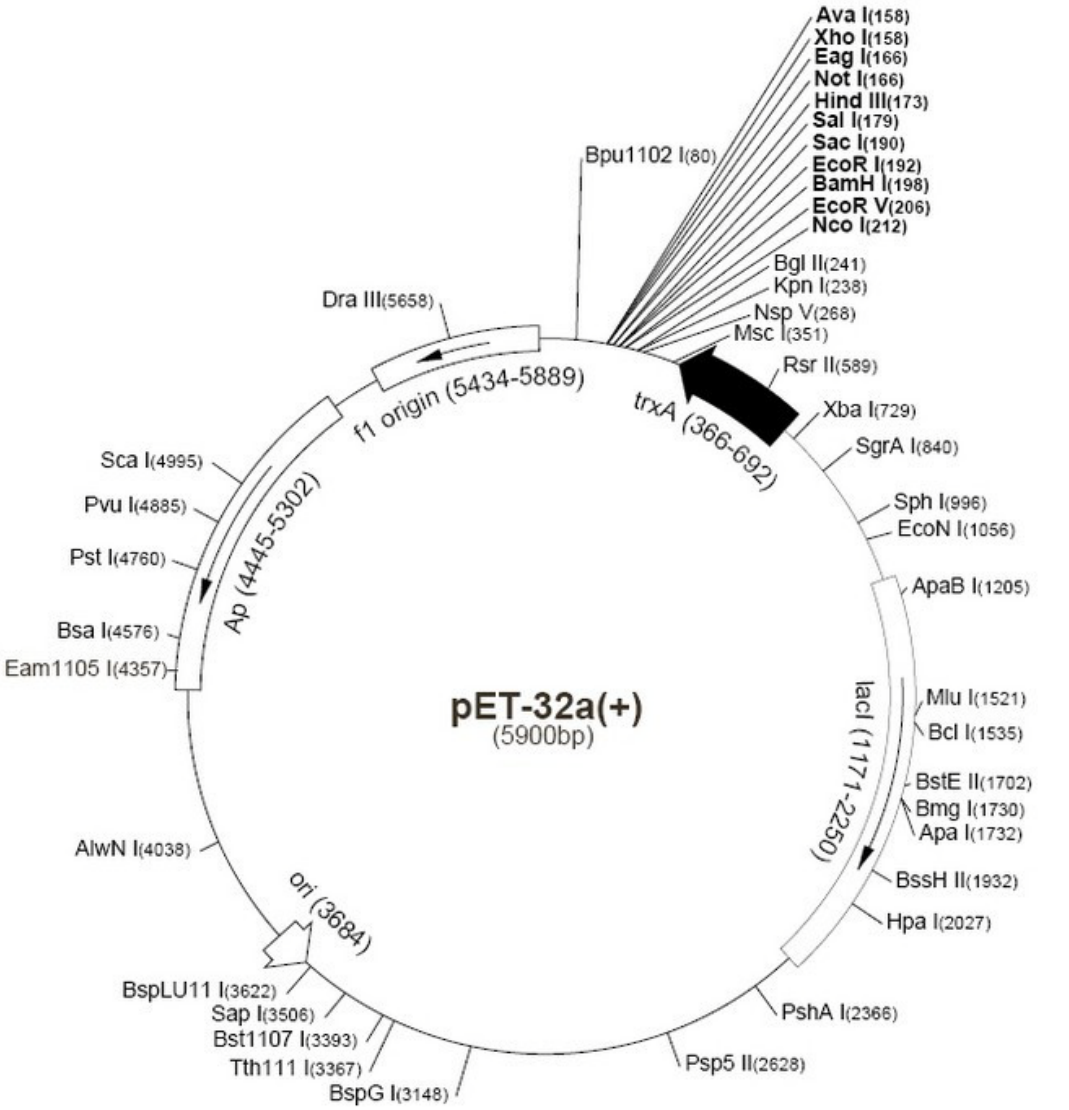
SIWSSTSMAESFMQSFTSTLGQKGVLSADQMDDIASIGDTLMGAVEKSGGKKNKLQALNMAFASSVAE
IAFADMTGLPADVKTNAILNSLSEAFLQTTGFVDNYFIQEIGGLINMFAEATANE

Figure 2.1 Amino acid sequences of the two MiSp fragments.

2.1.2 Expression vector

The expression plasmid vector used in this study was pET-M which is derived from pET-32a(LaVallie, DiBlasio et al. 1993). In this vector, Trx-tag, S-tag, Bg1II, KpnI, enterokinase, NcoI and EcoRV sites are removed, while the rest multiple cloning sites (MCS) remain the same (Figure 2.2).

a



b

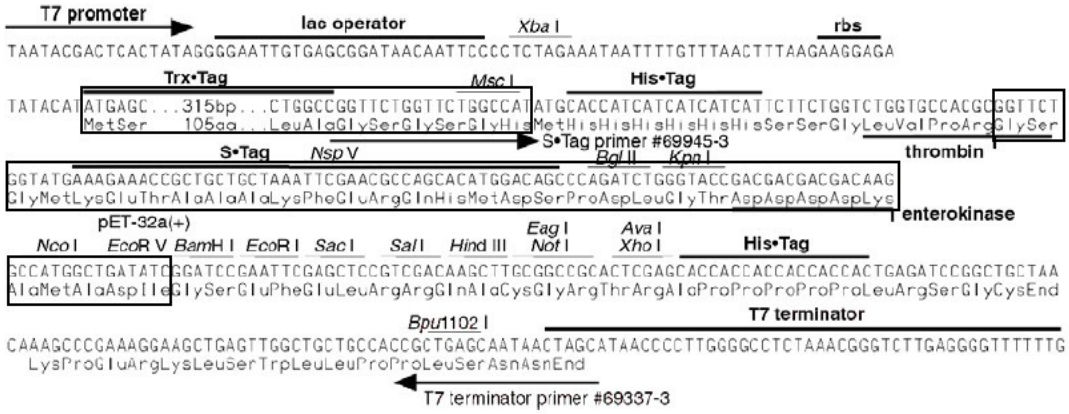


Figure 2.2 Expression vector. (a) Plasmid diagram of pET-32a(+). (b) The cloning region of pET-M vector, whereas the boxed regions has been removed.

2.1.3 Construction of expression vectors

2.1.3.1 Subcloning

The constructs of single MiSp domains and different combinations of domains were made, like CTD_{Mi}, NTD_{Mi}, RP_{Mi}, LK-CTD_{Mi}, RP-LK_{Mi}, and RP-LK-CTD_{Mi}. The target genes were firstly amplified from the two templates mentioned above using polymerase chain reaction (PCR). Then, the PCR products and pET-M vector were double digested by *BamH* I and *Xho* I (New England Biolab) at 37˚C for 1~2 hours and subsequently purified by QIAquick Gel Extraction Kit (Qiagen). The digestion products were then ligated by T4 DNA Ligase (New England Biolab) at 37˚C for 1~2 hours. Ligations products were transformed into *E. coli* DH5α for colony screening. Agarose DNA electrophoresis was used to monitor the quality of DNA products at each step. The plasmids of positive colonies were extracted by QIAprep Spin Miniprep Kit 250 (Qiagen) and the insertion sequences of the plasmids were confirmed by DNA sequencing.

2.1.3.2 Construction of mutants

Five single mutations and one double mutation of the charged residues of CTD_{Mi} were created by two-step PCR. For the single mutations, the forward primer was 5’-GCGGGATCCGTGGGAACTACAGTCGCC-3’ and the reverse primer was 5’-GCGCTCGAGTTAACCCATAACAACTTG-3’. The primers covering the mutation sites were list as below.

R27A	Primer 1: ACAACATCGGCTTTGAGCACTGC Primer 2: AGTGCTCAAAGCCGATGTTGTAG
D75N	Primer 1: CCTGGAGTGTCAAATAGCGAAGTTTTAAT Primer2: TAAAACTTCGCTATTTGACACTCCAGGAG
D61N	Primer 1: TCTGTTATTGCGAATCTTTTTGCACA Primer 2: TGCAAAAAGATTCGCAATAACAGAGG

D103N	Primer 1: GTAGGACAAGTAAATTTTAGTTCTGTTGG
	Primer 2: AACAGAACTAAAATTTACTTGTCCTACGC
E32Q	Primer 1: TTGAGCACTGCACAAGCATCTTCTAG
	Primer2: AGAAGATGCTTGTGCAGTGCTCAAAC

The double mutation, E32Q/D75N, is obtained by simply adding another round of two-step PCR at the base of one of the two single mutations using the same primers. The sequences of all the mutants were confirmed by DNA sequencing.

2.1.3.3 Construction of NTD-LK-RP-LK-CTD$_{Mi}$

The construct including all the single domains of MiSp (NTD-LK-RP-LK-CTD$_{Mi}$) was created. The normal construction method could not be applied directly to this construct because there was no such proper single DNA template to conduct the target gene PCR. Thus, linking some short constructs is necessary.

First of all, three short constructs (NTD, LK and RP-LK-CTD) were subcloned into the pET-M vector using *BamH* I and *Xho* I. Besides these two cleavage sites, *BspE* I and *Xma* I which are compatible but nonregenerable were also introduced to the three constructs for linkage as shown in figure 2.3. Double digestions were carried out on constructs of NTD and LK and the two fragments containing sequence of NTD and LK were selected to do the ligation (Figure 2.3). The ligation product was transformed into *E. coli* DH5α. The plasmids of positive colonies were extracted followed by double digestion and DNA sequencing to confirm the correct linkage. Until here, the construction of NTD-LK was finished. Next, constructs of NTD-LK and RP-LK-CTD were double digested followed by the same procedures mentioned above. Finally, the construction of NTD-LK-PR-LK-CTD$_{Mi}$ was achieved.

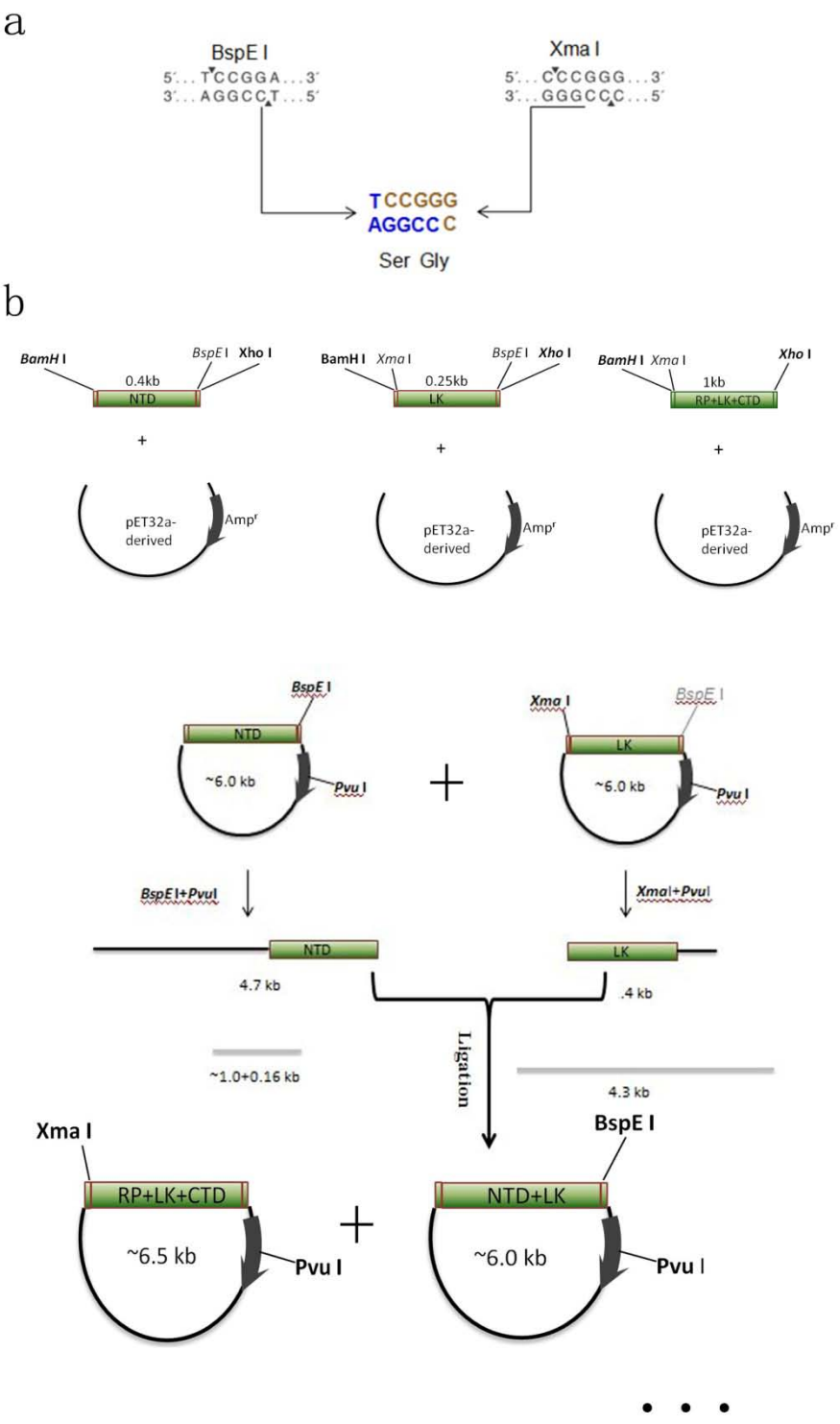


Figure 2.3 Cloning strategy of NTD-LK-RP-LK-CTD$_{Mi}$. (a) Cleavage sites of *BspE* I and *Xma* I. (b) Diagram showing the procedures of linking different constructs.

2.1.4 Transformation of *E. coli* competent cells

All the constructed expression vectors were transformed into *E. coli BL21(DE3)* to express the recombinant proteins. A tube containing 100 μl frozen competent cells for each transformation was thawed on ice. 10 μl ligation product was then added to the competent cells and mixed by gently tapping. The mixture was incubated on ice for 20 min. Then, the competent cells were heat shocked for 60 seconds at 42˚C water bath, followed by 2 min incubation on ice. 900 μl LB medium was added to the cells which were then shaken at 37˚C for 1 hour. The mixture was spread onto a LB agar plate containing the appropriate antibiotic selective for the plasmid. The plate was incubated at 37˚C for 12~16 hours.

2.2 Media

The media used in this study include LB broth, LB agar and M9 medium. 1 L M9 minimal medium contained 1~4 g Glucose, 1 g NH_4Cl, 0.24 g $MgSO_4$, 0.0111 g $CaCl_2$, 6.78 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 100 mg ampicillin and distilled deionized water. The LB broth and LB agar were prepared following the instructions from manufacturers.

2.3 Protein expression

2.3.1 Expression of unlabeled proteins

After transforming all the constructed expression vectors into *E. coli BL21(DE3)*, one single colony of each construct was inoculated into 10 mL LB medium containing 100 μg/ml ampicilin (LB+Amp) and grown overnight at 37˚C with shaking. The overnight culture was further inoculated into 1 L LB+Amp and cultured at 37˚C until the OD_{600} reached ~0.6. Right after induction by 0.2 mM IPTG (isopropyl β-D-thiogalactoside), cells were shifted to 20˚C and further cultured for 16 hours. Protein expression was checked by comparing the SDS-PAGE results of before-induced and after-induced cell culture samples.

2.3.2 Expression of ^{15}N, ^{13}C-labeled (^{15}N-labeled) proteins

For ^{13}C, ^{15}N-labeled (^{15}N-labeled) proteins, the cells were cultured in M9 medium which contained only ^{15}N-labeled NH_4Cl and ^{13}C-labeled (non-labeled) glucose as the sole nitrogen and carbon source. The rest procedures were the same as those of the unlabeled protein expression.

2.4 Protein purification

The cell pellet harvested from 1 L culture was re-suspended in 50 ml lysis buffer (10 mM Tris, 100 mM NaCl, pH 8.0) and sonicated thoroughly. The lysate was then centrifuged for 1 hour at 12000 g. The over expressed proteins could be found in both supernatant and pellet, and the protein purification procedures from supernatant and pellet were different.

2.4.1 Purification of proteins expressed in supernatant

All the proteins used in this study were fused with a 6xHis-tag and a thrombin cleavage sequence at the N-terminus. For the proteins expressed in supernatant (CTD_{Mi}, NTD_{Mi}, LK-CTD_{Mi} and all the mutants of CTD_{Mi}), the supernatant of lysate was loaded into a 10 ml Ni-NTA resin column. After 1 hour binding, the resin was washed with 100 ml washing buffer (10 mM Tris, 100 mM NaCl, 20mM imidazole, pH 8.0). The target proteins were then eluted out by a 3x15 ml elution buffer (10 mM Tris, 100 mM NaCl, 250 mM imidazole, pH 8.0). The imidazole in the protein samples was then removed by dialysis. Finally, the protein samples were concentrated using centrifugal filter units with 3 kDa cutoff membrane at centrifugal force of 3000 g. The concentrated samples were further purified by gel filtration column or/and ion exchange column.

2.4.2 Protein purification by refolding

For those proteins expressed in inclusion body of bacteria (RP_{Mi}, RP-LK_{Mi} and RP-LK-CTD_{Mi}), the refolding process was needed. After sonication, the pellet of

lysate was dissolved by 8 M urea which could denature most of protein and make them soluble. The denatured protein solution was then centrifuged at 12000 g for 1 hour. The supernatant was loaded to Ni-NTA resin column. Except that the binding time was extended to overnight and the washing and elution buffers were prepared in 8 M urea, the rest of Ni-NTA resin column purification procedures were the same as mentioned above. The protein solution was then refolded by dialysis against 10 mM Tris buffer at pH 7.0 with 100 mM NaCl. After refolding, the protein was further purified by gel filtration or/and ion exchange column. Although NTD-LK-RP-LK-CTD_{Mi} was expressed in supernatant, the refolding method was employed because the native state protein formed fiber or aggregation easily which would stack in the Ni-NTA column and compromise the effect of purification.

2.5 NMR samples preparation

The sample used for structure determination of CTD_{Mi} contained 1 mM ^{13}C, ^{15}N-labeled protein, 10 mM phosphate buffer (pH 6.8), 5 mM EDTA, 50 mM NaCl and 0.01% sodium azide. One additional sample of CTD_{Mi} containing 50% ^{13}C, ^{15}N-labeled and 50% unlabeled proteins(Zwahlen, Legault et al. 1997) was prepared for the 3D ^{13}C, ^{15}N-filtered NOESY experiment. This sample was prepared by mixing equal amount of labeled and unlabeled proteins in 8 M urea for 2 hrs and then removing the urea by dialysis against 10mM phosphate buffer.

The sample used for structure determination of RP_{Mi} contained 0.4 mM ^{13}C, ^{15}N-labeled protein, 10 mM Tris buffer (pH 7.0), 5 mM EDTA, 100 mM NaCl and 0.01% sodium azide.

The sample used for the backbone assignment of NTD_{Mi} contained 0.5 mM ^{13}C, ^{15}N-labeled protein, 10 mM phosphate buffer (pH 7.0), 5 mM EDTA, 300 mM NaCl and 0.01% sodium azide.

The rest of unlabeled and ^{15}N-labeled NMR samples were prepared in different protein concentrations and buffer conditions according the purposes of the experiments.

2.6 NMR spectrascopy

All NMR experiments were performed on a Bruker 800 MHz NMR spectrometer at 25°C. The 1D ^{1}H NMR spectra were recorded using the water gate W5 pulse scheme with 64 scans and an interscan delay of 2 s. To obtain sequence-specific assignments and NOEs of CTD_{Mi} and RP_{Mi}, the following spectra were recorded: 2D ^{1}H-^{15}N HSQC, 2D ^{1}H-^{13}C HSQC, 3D HNCA, 3D HN(CO)CA, 3D MQ-CCH-TOCSY(Yang, Zheng et al. 2004), 4D time-shared ^{13}C, ^{15}N-edited NOESY(Xu, Long et al. 2007). Inter-molecule NOEs of CTD_{Mi} were identified from a 3D ^{13}C, ^{15}N-filtered NOESY experiment. To obtain hydrogen bond restraints, a series of ^{1}H-^{15}N HSQC spectra were recorded at a range of time intervals after dissolve the lyophilized protein sample in D_2O.

To obtain sequence-specific assignments of NTD_{Mi}, 2D ^{1}H-^{15}N HSQC, 3D HNCACB and 3D CBCA(CO)NH were recorded.

All the spectra were processed with NMRpipe software(Delaglio, Grzesiek et al. 1995).

2.7 Resonances assignment

2.7.1 Backbone and side chain assignment

For the backbone assignment of NTD_{Mi}, All the cross peaks from HNCACB, CBCA(CO)NH were clustered based on ^{1}H, ^{15}N chemical shifts using an NMRspy (http://yangdw.science.nus.edu.sg/Software&Scripts/NMRspy/contacts.htm) extension written by one of our group member. The cross peaks in the ^{1}H-^{15}N HSQC were used as a reference. The ^{1}H, ^{15}N pair can be correlated with the corresponding C^{α}, C^{β} for i, and i-1 residues. The clusters can be linked to small fragments containing several residues based on the redundant chemical shift information of C^{α}, C^{β}. The spin systems of Ala, Thr, Ser and Gly residues could be assigned to residue types because they all have characteristic C^{α} and/or C^{β}. Finally, according to the positions of characteristic residues in the fragments and the protein sequence of NTD_{Mi}, the

fragments could be mapped into the corresponding position in the sequence.

The backbone and side chain assignment of CTD_{Mi} and RP_{Mi} were achieved following the strategy described previously(Xu, Zheng et al. 2006) and using NMRspy and XYZ4D (http://yangdw.science.nus.edu.sg/Software&Scripts/XYZ4D/index.htm).

2.7.2 NOE assignment, structural calculation and refinement

Crosspeaks from 3D and/or 4D ^{15}N, ^{13}C-edited NOESY spectra were picked and unambiguously assigned using NMRspy. The upper limits of the distance restraints were calibrated to 3 categories (2.8, 3.4 and 5 Å) according to their peak intensities. All the lower limits of the distance restraints were set as 1.8 Å. Together with the hydrogen bond and/or angular restraints, the unambiguous distance restraints were input to the software CYANA(Herrmann, Güntert et al. 2002) to calculate a initial structure. According to the initial structure, some of the ambiguous NOEs could be unambiguously assigned. The assignment of NOEs and structure calculation were iterative processes. Final structure calculation was initiated from 100 conformers with random torsion angle values. 10,000 simulated annealing steps per conformer with torsion angle dynamics algorithm were performed. The qualities of the calculated structures were evaluated using Procheck-NMR(Laskowski, Rullmannn et al. 1996).

2.8 Structure display and analysis

The software UCSF Chimera(Pettersen, Goddard et al. 2004) was used to display and analysis the structures. The color codes for all the hydrophobic and charged surface plots were: dark blue for positively charged, red for negatively charged, light blue for hydrophilic, yellow for hydrophobic. Since the hydrophobicity of each hydrophobic residue are different, different degrees of yellow color were employed to indicate the hydrophobicity. The normalized hydrophobicity scores of the hydrophobic residues are listed as follow(Roseman 1988):

Phe	1
Trp	0.94
Ile	0.80
Leu	0.80
Tyr	0.65
Val	0.57
Pro	0.44
Met	0.42
Ala	0.17
Gly	0

The colors of these hydrophobic residues on the surface plots were a mixture of yellow and white. The percentage of yellow in the color corresponded to the normalized hydrophobicity scores of these residues. For example, the color of Phe with a score of 1 was 100% yellow, whereas the color of Gly with a score of 0 was 100% white.

2.9 Dynamic light scattering

Dynamic light scattering (DLS) experiments were performed on a DynaPro instrument (Protein Solutions, Lakewood, NJ) with a He-Ne laser. All the protein samples were centrifuged for 30 min at 15000 g to spin down the large artificial particles which could affect the experiment results. 15 μl protein solution was taken and loaded into a 1.5 mm quartz cuvette. For each experiment, 10-20 data points were collected at a speed of 1 point/10s and analyzed using Dynamics 5.0 software.

2.10 Circular dichroism and protein unfolding

All the circular dichroism **(**CD) spectra were recorded on a Jasco J-810 spectropolarimeter equipped with a thermal controller. A 0.1 cm path length cuvette and 200 μl protein samples were used for all the CD experiments. The far-UV spectra

were recorded at a 0.1 nm spectral resolution with 3 scans. In general, all samples were prepared in a 10 mM phosphate buffer at ~pH 6.8. Both urea- and thermal-induced unfolding processes were monitored at 222 nm using samples with 10 μM protein, 10 mM sodium phosphate at pH 6.8.

For the thermal-induced unfolding experiments, the temperature changed from 20 ℃ to 90 ℃ with a speed of 1 ℃/min. For urea-induced unfolding experiments, a series of samples were prepared by adding 1 mM protein stock solution to pre-mixed solutions of urea and water. The transition curves of all the unfolding experiments were obtained by plotting the change of ellipticity at 222 nm against temperature or urea concentration.

Except for RP-LK-CTD_{Mi}, urea denaturation curves for other MiSp constructs were analyzed with a two-state model (Yang, Noble et al. 2009). For urea-denaturation of RP-LK-CTD_{Mi}, a three-state model was used to fit the experimental data by a linear combination of two two-state equations.

2.11 Size exclusion chromatography

A Superdex^{TM} 75 PG (GE Healthcare) column with a total volume of 120 ml was used to run all the protein samples except NTD-LK-RP-LK-CTD_{Mi} which was purified by a Superdex^{TM} 200 PG (GE Healthcare) column. In general, the running buffer contained 10 mM sodium phosphate or Tris (~pH 7.0) with different concentrations of NaCl. The flow rate used was 1 ml/min, and fractions were collected every 2 ml. The fractions were analyzed by SDS-PAGE to confirm which peak in the UV absorbing profile was the target protein. A molecular mass standard set consisting of Ribomuclease A (13.7 kDa), Chymotrypsinogen A (25 kDa), Ovalbumin (43 kDa), and BSA (67 kDa) was chromatographed to estimate the apparent molecular weights of target proteins.

In addition, A Superdex 75 10/300 GL (GE Healthcare) analytical column was used to investigate the buffer effect on NTD_{Mi}.

2.12 Protein solubility

The purified protein samples in respective 10 mM sodium phosphate and 10 mM Tris buffers (pH 7.0) were concentrated using centrifugal filter units with 3 kDa cutoff membrane at centrifugal force of 3000 g. When the protein concentration was >5 mg/ml, 2 μl samples were regularly taken out from the solution until precipitate or gel was observed. Otherwise, larger volumes of samples were taken for concentration measurements. To determine protein concentrations, the taken samples were diluted in the same buffers as those used for the protein samples. The concentrations were measured using the absorbance at 280 nm and also estimated using SDS PAGE.

2.13 Shear force-induced aggregation

To study protein aggregation induced by shear force, samples of 2 ml with 0.05 mg/ml proteins and 10 mM phosphate buffer (pH 6.8) were placed into a UV/Vis cuvette with a small magnetic star bar stirring at 500 rpm, 25 ℃. The turbidity of the samples was monitored by measuring OD_{350} on a BIO-RAD Smart Spec^{TM} Plus Spectrophotometer at a series of time intervals.

To determine the effect of sodium chloride and sodium phosphate on the aggregation of RP-LK-CTD_{Mi}, the shear force-induced aggregation experiments were performed under two salt concentrations: 0 and 200 mM. The samples were shaken at 150 rpm, 25 ℃. At different time points, the samples were taken out. After removing the precipitate by centrifuge, the concentration of the soluble portion was measured and then the total amount of precipitated protein was calculated.

2.14 Scanning electron microscopy

1 ml purified protein sample containing 5 mg/ml RP-LK-CTD_{Mi} in 10 mM sodium phosphate buffer (pH 6.8) was placed into a 2 ml eppendorf tube and the sample was shaken at 200 rpm, 25℃ for 5 minutes. Then, silk fibers formed in the

tube were picked out by a needle. SEM micrographs of the fibers were observed on a JEOL JSM-6510 and photographed at a voltage of 15 kV and room temperature.

2.15 Prediction of disorder, hydrophobicity and aggregation propensity

The disordered residues in LK_{Mi} were predicted using PONDR-FIT (http://www.disprot.org/pondr-fit.php). If the disordered score of a residue is >0.5, this residues is considered as disordered (Xue, Dunbrack et al.). The aggregation-prone regions in LK_{Mi} were predicted using Zyggregator (http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php). When a region of several consecutive residues each have aggregation scores larger than 1, this region is considered to be prone to aggregate (Pawar, DuBay et al. 2005). The hydrophobicity plot of LK_{Mi} was obtained using Protscale (http://web.expasy.org/cgi-bin/protscale/protscale.pl) with the scale option of Hphob./Roseman (Roseman 1988).

3 Chapter 3: Results and discussion

In this chapter, all the results of this work are presented and the implications of these results are discussed.

3.1 Molecular architecture of MiSp

The spider species used in this study is *N. antipodiana* (golden web spider) which is popular in Singapore. Until now, the full length sequence of MiSp from this species is not available. However, with the newly identified N-terminal domain sequence, the molecular architecture of MiSp is more clear (Figure 1.4). Between the two terminal domains (NTD and CTD), RP domain and LK domain forms a repetitive unit which repeats several times.

PCR results from the genomic DNA showed that all the repetitive domains in the MiSp from *N. antipodiana* are identical although the exact number of repeats has not been determined yet. 5 types of linker domains were identified with different size ranging from 83 to 174 amino acids in genomic DNA (Figure 3.1a). Glycine (45 – 48%) and alanine (33 – 39%) are dominant in linker domains, which are consistent with previous reports(Colgin and Lewis 1998; Bittencourt, Souto et al. 2007). Interestingly, the linker domain between the CTD and RP domains of *N. antipodiana* (LK_{Mi}) obtained here is much shorter than that of *N. clavipes*(Colgin and Lewis 1998). RP_{Mi} is highly conserved among different species (Figure 1.5b). NTD_{Mi} from *N. antipodiana* shares ~30% identity with NTD_{Mi} from *A. ventricosus* (Figure 3.1b). Surprisingly, the identities shared by NTD_{Mi} from *N. antipodiana* and NTDs of MaSps are even higher (>40%) (Figure 3.1b).

The conservation of MiSp domains (NTD, CTD, RP) indicates their importance in both spider silk protein storage and fiber formation.

a

```
1   AGGYGGLVGYGAGAAAAAAAGAGSGGAGGYGRGAGAGAGAAAGAGAGGAGGNGGQGGYGS 60
2   -GDYGGLVGY-----------------------GAGAGAAAGAGAG------------- 22
3   -GDYGGLVGYG---------------------------------------------------- 10
4   -GDYGGLVGYGAGAGAAAGAGAGAGRAGGYIGQGGYGAGAGVAAAAAAGAGAGATGGYGR 59
5   AGGYGGLVGYG---------------------------------------------------- 11
     *.*******

1   GAGAAAAAGAGAGAGGAGGYGGQGGYGARAGAGAAAAAGAGAG-AGSY----GRGGGAGA 115
2   ---------------AAGGYGGQGGYGARAGAGAAAAAGAGAG-AGSY----GRGGGAGA 62
3   ----------------------------AGAGAAAAAGAGAG-AGSY----GRGAGAGA 36
4   GAGAGAAAAAGAGAGAGGAGGYGGQGGYGAGAGAGAAAAAGAGAG-AGSY----GRGGGAG- 113
5   ----------------------------AGAGAAAGAGAGAGGAGGYIGQGGYGAGAGA 42
                                 *******.****** **.*    * *.***

1   GAAAGAAAGGAGGY-RGQGGYGAGAGVAGAATSGAGAGGAGGYGRGAGAVAGAGAGSAAG 174
2   GAAAGAAAGGAGGY-RGQGGYGAGAGVAGAATSGAGAGGAGGYGRGAGAVAGAGAGSAAG 121
3   AAGAGAGAGGAGGY-RGQGGY------------GAGAGGAGGFGRGAGAGAGAGGGSAAG 83
4   ---AGAAAGGAGGY-RGQGGYGAGAGAAGAAAAGAGAGGAGGYGRGAGAGAGAGAGSAAG 169
5   AAAAGAGAGATGGYGRGAGAG---------ATNAGGYGGQGGYGAGARAFAGAGVG---- 89
      ***.**.:*** ** *.             ..* ** **:* ** * **** *
```

b

```
MiSp(N.a)       ------------SIWSSTSMAESFMQSFTSTLGQKGVLSADQMDDIASIGDTLMGAVEKS
MiSp(A.v)       ------------PIWTNPNAAMTMTNNLVQCASRSGVLTADQMDDMGMMADSVNSQMQKM
MaSp(N.c)       GPLGSPGIPGQNTPWSSTELADAFINAFMNEAGRTGAFTADQLDDMSTIGDTIKTAMDKM
MaSp(E.a)       -----GSGNSHTTPWTNPGLAENFMNSFMQGLSSMPGFTASQLDDMSTIAQSMVQSIQSL
MaSp(L.h)       -------MGQANTPWSSKANADAFINSFISAASNTGSFSQDQMEDMSLIGNTLMAAMDNM
                             *:.   *  : : : .  .    :: .*::*:. :.:::   ::.

MiSp(N.a)       GGK----KNKLQALNMAFASSVAEIAFADMTGLPADVKTNAILNSLSEAFLQTTGFVDNY
MiSp(A.v)       GPN--PPQHRLRAMNTAMAAEVAEVVATSPP-QSYSAVLNTIGACLRESMMQATGSVDNA
MaSp(N.c)       ARSNKSSKGKLQALNMAFASSMAEIAAVEQGGLSVDAKTNAIADSLNSAFYQTTGAANPQ
MaSp(E.a)       AAQGRTSPNKLQALNMAFASSMAEIAASEEGGGSLSTKTSSIASAMSNAFLQTTGVVNQP
MaSp(L.h)       GGR--ITPSKLQALDMAFASSVAEIAASEGGDLGV--TTNAIADALTSAFYQTTGVVNSR
                .        :*:*:: *:*:.:**:.  .          .:*  .: .:: *:** .:

MiSp(N.a)       FIQEIGGLINMFAEATANE-----
MiSp(A.v)       FTNEVMQLVKMLSADSANEVSTAS
MaSp(N.c)       FVNEIRSLINMFAQSSANEVSYGG
MaSp(E.a)       FINEITQLVSMFAQAGMNDVSA--
MaSp(L.h)       FISEIRSLIGMFAQASANDVYAS-
                * .*:  *: *::    *:
```

Figure 3.1 Sequence alignments of LKs and NTDs. (a) Sequence alignment of 5 types of linker domains from *Nephila antipodiana*. Type 5 is the linker domain between the RP and CTD domains. (b) Sequence alignment of N-terminal domains of MiSps from *Nephila antipodiana* (N.a) and *Aranecus ventricosus* (A.v) and MaSps from *Nephila clavipes* (N.c), *Euprosthenops australis* (E.a) and *Latrodectus hesperus* (L.h).

3.2 CTD_{Mi}

3.2.1 Expression and purification of CTD_{Mi}

The CTD_{Mi} was expressed at 20℃ for 16 hours and most of the protein was released to supernatant after sonication. Because CTD_{Mi} was fused with His-tag, Ni-NTA affinity chromatography was employed to purify the protein. Most impurities could be removed in this step and the purity of CTD_{Mi} in the elution was higher than 80% (Figure 3.2). To further purify the protein, size-exclusion chromatography (SEC) was used. The result is shown in Figure 3.3. The protein was eluted out at the elution volume of 68 ml which corresponds to protein size of 25 kDa. The fractions corresponding to the target protein were collected and concentrated. The purity of the protein after SEC was higher than 90% (Figure 3.3b). The final purity of the protein was improved by passing through the ion exchange column (Figure 3.3c).

1D ^{1}H NMR and CD experiments of unlabeled CTD_{Mi} show that the CTD_{Mi} is well folded and its secondary structure is mainly α-helix (Figure 3.4). Thus, ^{15}N, ^{13}C labeled sample was expressed in M9 minimal medium. The double labeled sample was purified using the same method as the unlabeled sample.

The molecular weight of CTD_{Mi} is 12.2 kDa. The SEC result indicates that the CTD_{Mi} exists as dimers in aqueous solution (Figure 3.3a). In addition, the DLS result further supports the dimerization of CTD_{Mi} (Figure 3.5). Therefore, an additional half labeled sample was needed to differentiate the intra molecular NOEs from inter molecular NOEs. Unfolding-refolding process was involved in this sample preparation. Figure 3.4b shows that the CD spectra of native and refolded CTD_{Mi} are the same, which indicates that the structure of CTD_{Mi} is not affected by the refolding process.

The secondary structure of CTD_{Mi} does not change within a wide range of pH (6-12) as evidenced by CD results (Figure 3.6). Thus, pH 6.8 which is suitable for NMR experiments was chosen for the NMR samples.

Figure 3.2 Purification of CTD_{Mi} by Ni-NTA column. Line 1: flow-through of the supernatant after binding to Ni-NTA column; Line 2: first 50 ml wash; Line 3: second 25 ml wash; Line 4: third 25 ml wash; Line 5-7: first, second and third 15 ml eluents; Line 8: protein marker. His-CTD_{Mi} is indicated by arrow.

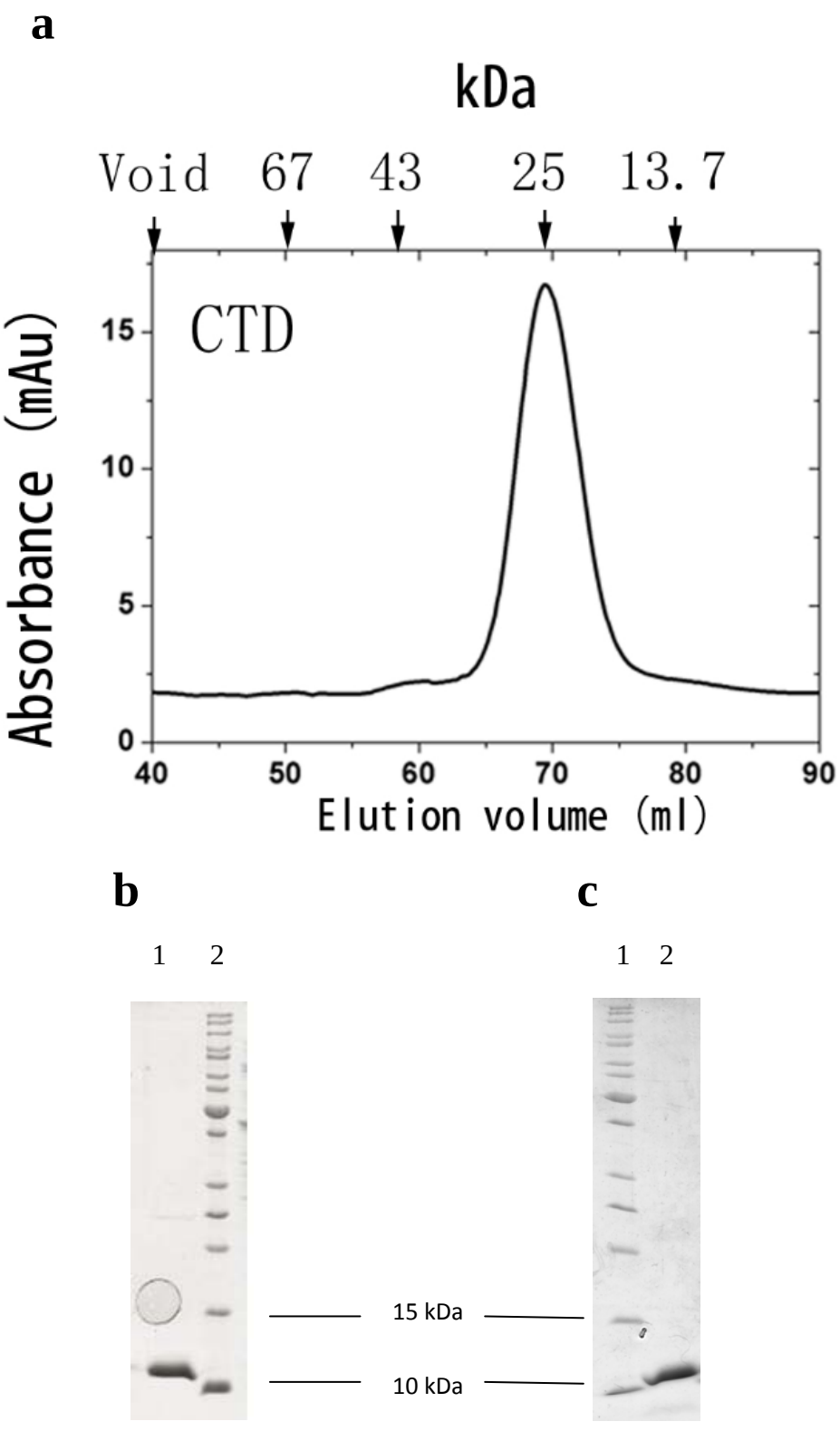


Figure 3.3 Purification of CTD_{Mi}. (a) Size exclusion chromatography profile of CTD_{Mi}. Molecular weight makers are indicated on the top. (b) SDS PAGE of CTD_{Mi} after size exclusion chromatography. Line 1: CTD_{Mi}; Line 2: protein marker. (c) SDS PAGE of CTD_{Mi} after passing through ion exchange column. Line 1: protein marker; Line 2: CTD_{Mi}.

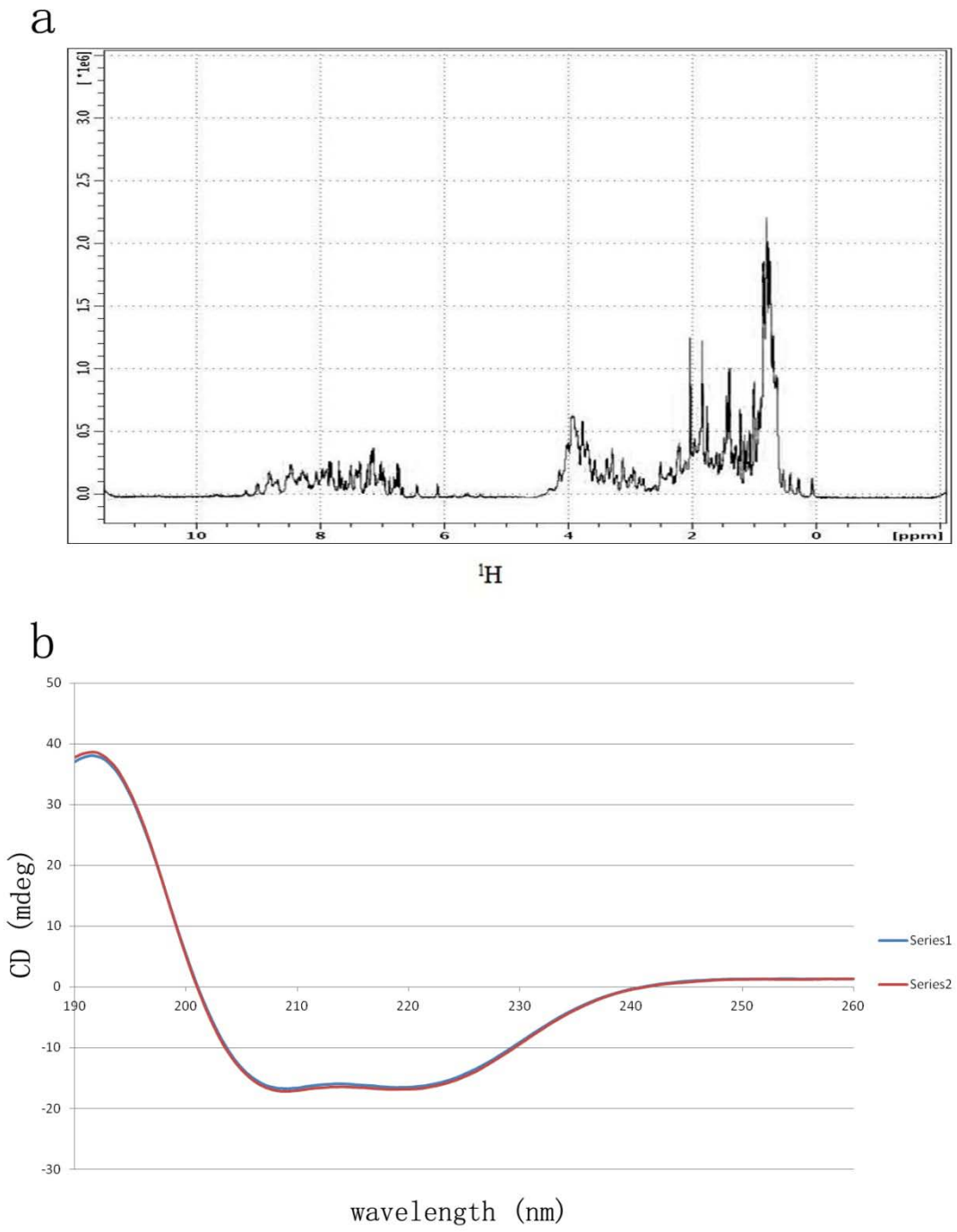


Figure 3.4 NMR and CD results of CTD_{Mi}. (a) 1D 1H NMR experiment result of CTD_{Mi}. Sample buffer is 10 mM phosphate at pH 6.8. (b) CD spectra of CTD_{Mi}. Sample buffer is 10 mM phosphate at pH 6.8. Series 1: CTD_{Mi} at native state; Series 2: CTD_{Mi} after urea denaturing and then refolding.

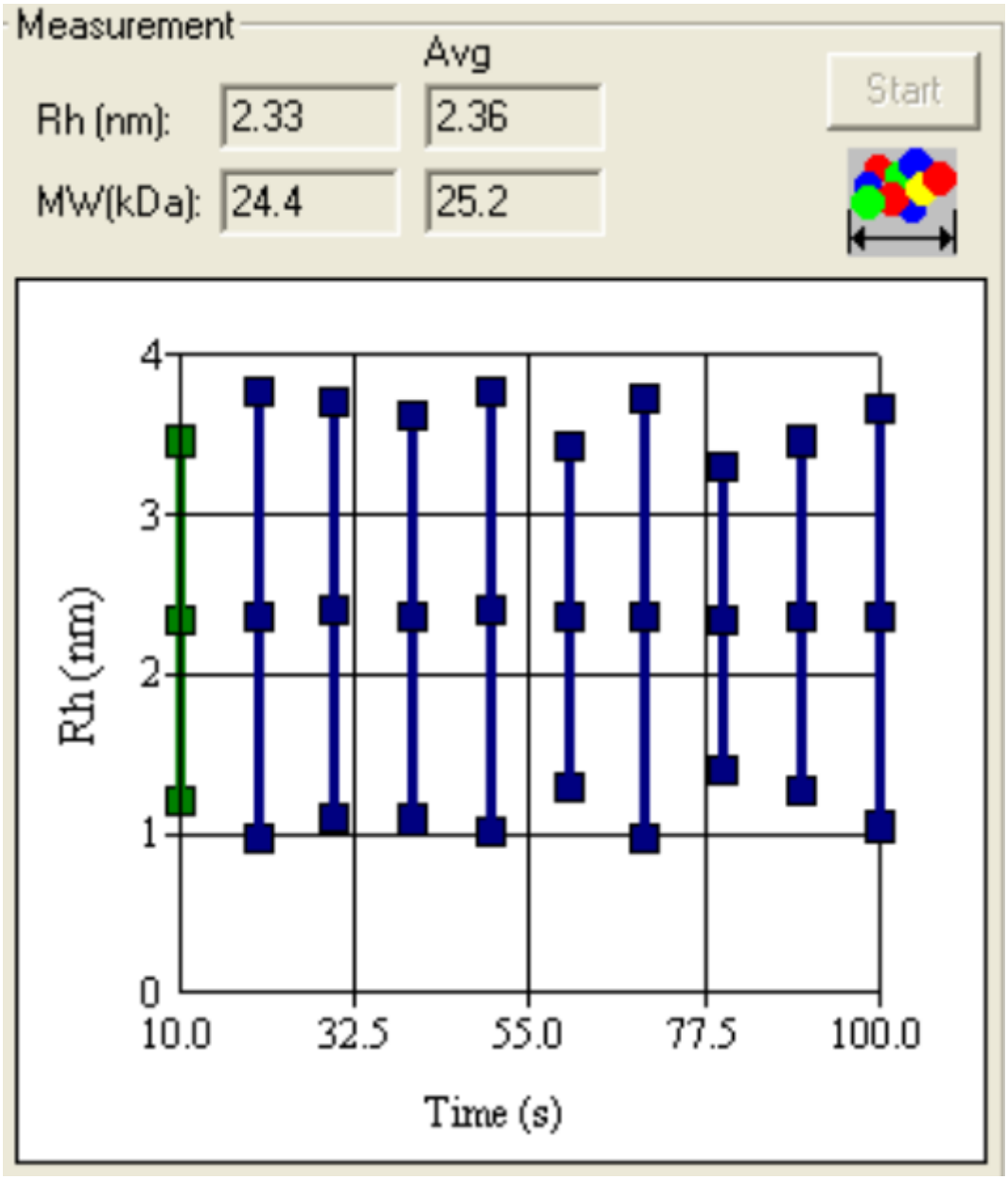


Figure 3.5 Dynamic light scattering result of CTD_{Mi}. 1 mM protein was used in 10 mM phosphate buffer at pH 6.8.

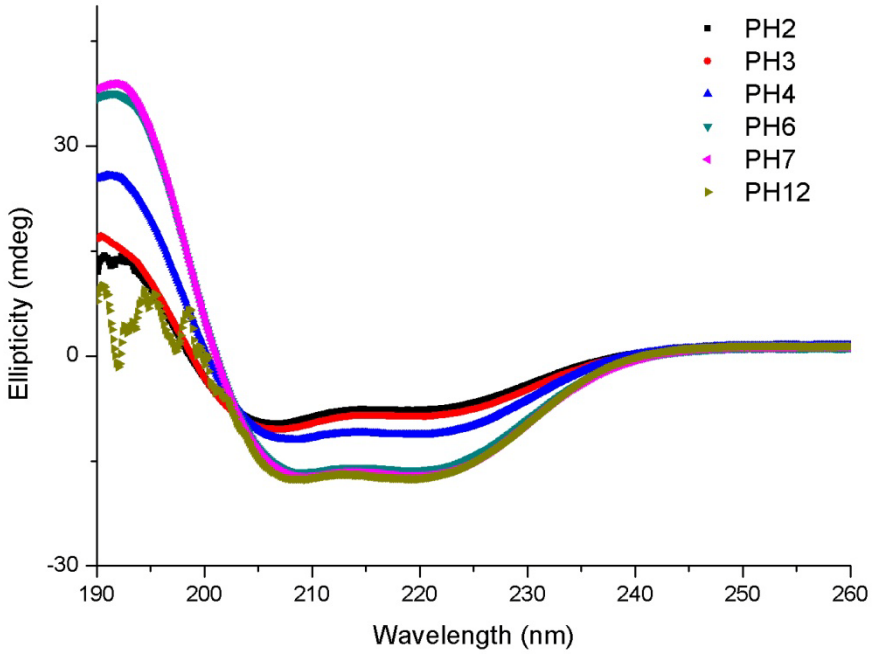


Figure 3.6 pH titration of CTD_{Mi} monitored by CD. Different pH conditions are indicated by different colors.

3.2.2 NMR resonance assignment of CTD_{Mi}

3.2.2.1 Backbone and side chain resonance assignments of CTD_{Mi}

100% backbone resonance assignments of CTD_{Mi} (except Pro and the region Met^{1}-Ser^{16}, corresponding to the His-tag and the thrombin cleavage site) was achieved by using 3D HNCA, 3D HNCOCA, and 4D NOESY experiments (Figure 3.7). With the backbone assignments, most of the aliphatic side chain resonance assignments were obtained from the 3D CCH-TOCSY and the 4D NOESY experiments. The aromatic side chain resonances were assigned from 4D NOESY experiment whenever possible.

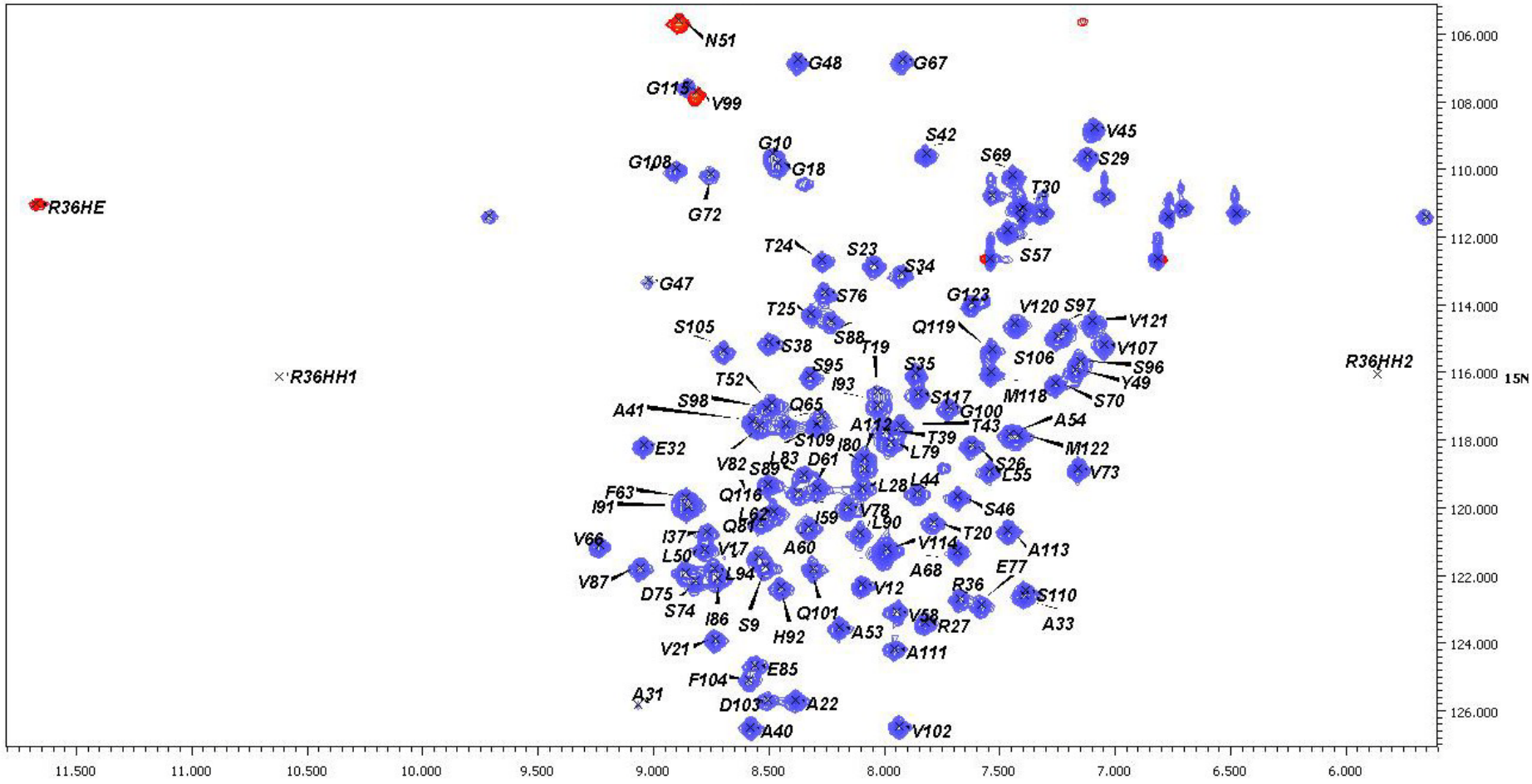


Figure 3.7 Assigned ^{1}H, ^{15}N HSQC spectrum of CTD_{Mi}. Positive peaks are blue and negative peaks are red.

3.2.2.2 Secondary structure prediction by chemical shift index (CSI)

Chemical shift values of C^{α} and C^{β} of a residue are related to the secondary structure of a protein. Thus, the difference between ^{13}C chemical shift values for each residue obtained from experiments and those reported for the same residue in a random coil conformation (δC^{α} and δC^{β}) are used to predict the secondary structure of a protein before solving its 3D structure. Normally, the chemical shift index value, (δC^{α}-δC^{β}), is positive for residues in α-helix and negative in β-sheet. The (δC^{α}-δC^{β}) of CTD_{Mi} was calculated and the result is shown in Figure 3.8. From the result, five α-helical regions were predicted.

3.2.2.3 NOE assignment

The H^{N}-H^{C}, H^{N}-H^{N} and H^{C}-H^{C} NOE resonances were firstly assigned from 4D ^{15}N, ^{13}C-edited NOESY spectrum manually using NMRspy. After assigning all the unambiguous NOEs, the 3D ^{13}C, ^{15}N-filtered NOESY spectrum was used to differentiate the intra-molecular NOEs from inter-molecule NOEs. Those crosspeaks showed on both 4D NOESY and 3D ^{13}C, ^{15}N-filtered NOESY spectra corresponded to the inter-molecule NOEs. The ambiguous NOEs were later assigned according to the preliminary structure obtained by iterated structure calculation. In total, 2910 NOEs were assigned including 171 inter-molecular NOEs (Table 3.1).

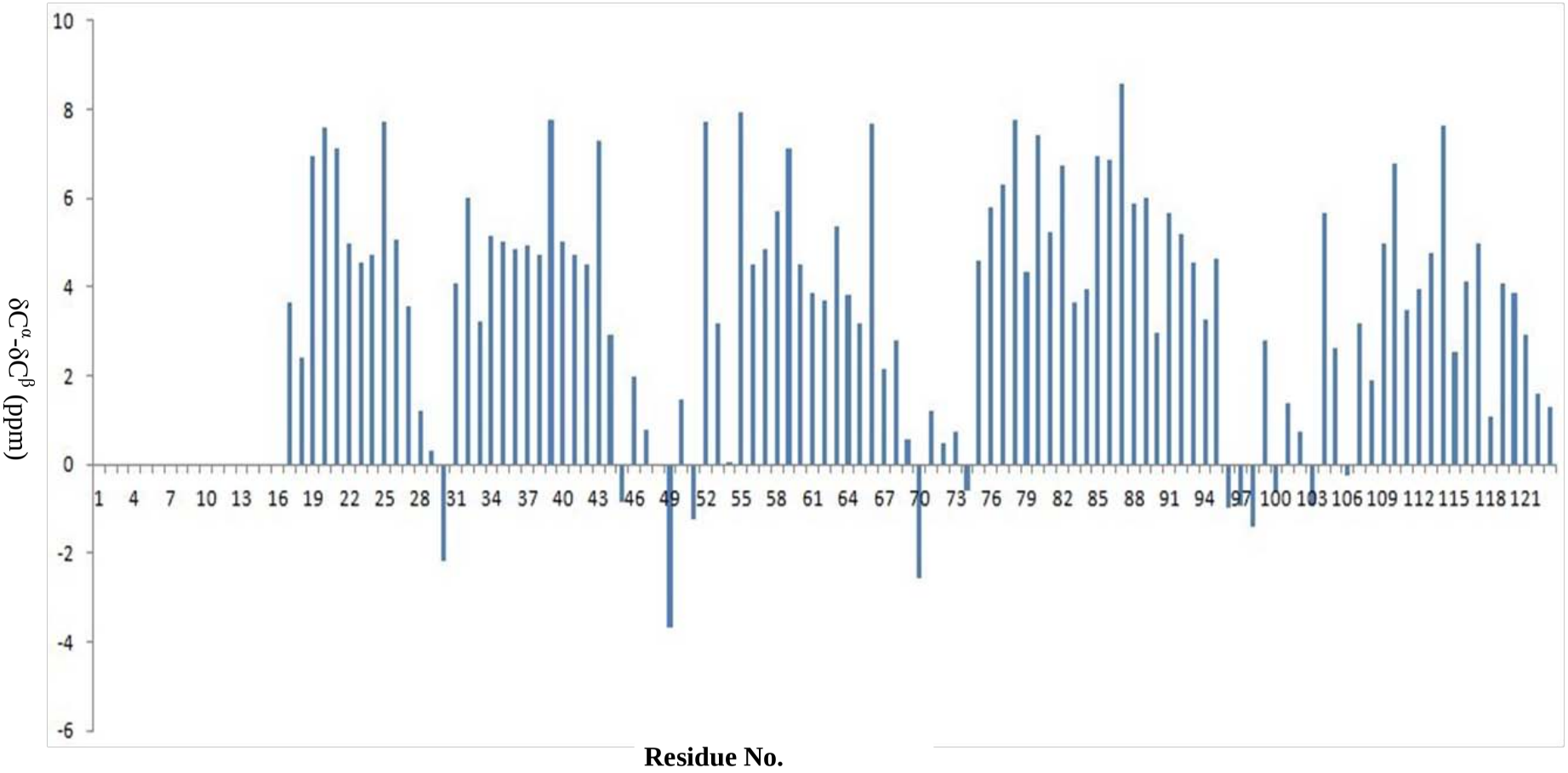


Figure 3.8 Chemical shift index of CTD $_{Mi}$. Residue 1-16 are the His-tag and thrombin cleavage site.

3.2.3 Solution structure of CTD_{Mi}

2910 distance constrains obtained from multidimensional NMR spectroscopy and 116 backbone dihedral angle constrains derived by TALOS were used to determine the structure of CTD_{Mi} (Table 3.1 and Figure 3.9a). The detailed structural statistics are listed in Table 3.1. The final ten dimers calculated by CYANA have backbone and heavy atoms r.m.s. deviations of 0.5 Å and 0.82 Å, respectively. No dihedral angle violation greater than 5° or distance violation greater than 0.5 Å was observed during the structure calculation, indicating that the NMR structure of CTD_{Mi} was well defined and in great agreement with experimental restraints. Ramachandran plot of the final ten structures shows that 99.7 % of the residues have backbone dihedral angles in allowed regions.

Overall, the structure of CTD_{Mi} adopts a globular fold of two twisted five-helix bundles (α1 [Gly^{18}-Leu^{28}], α2 [Ala^{31}-Val^{45}], α3 [Leu^{55}-Ser^{69}], α4 [Asp^{75}-Ser^{97}], α5 [Val^{107}-Met^{122}]) which pack in parallel to form a homodimer (Figure 3.9a). α5 is swapped to stabilize the dimeric structure. The major dimer interface involves helices α1/ α5', α4/α4' and α5/ α1'. Given that most of hydrophobic residues are buried inside, the surface of the dimer mainly display hydrophilic and charged residues (Figure 3.9b). Many hydrophobic residues are located in the interface and are in close contact (Figure 3.10a), suggesting that hydrophobic interactions are the dominant factor for holding the two monomers together. Similarly, hydrophobic interactions (26 hydrophobic residues are involved) are critical for the stability of each monomeric unit (Figure 3.10b). In addition, α4 is connected with α1 and α2 through two salt bridges R27-E77 and R36-E85 in each monomer, respectively (Figure 3.11a). The formation of the R36-E85 salt bridge is evident from the extremely large chemical shift of $^{1}H_{\varepsilon}$ of R36 (11.7 ppm) and the observation of $^{1}H_{\eta 1}$ (10.5 ppm) and $^{2}H_{\eta 2}$ (5.8ppm) of R36 (Figure 3.7). Although $^{1}H_{\eta 1}$ and $^{2}H_{\eta 2}$ of R27 were not detected, the side-chains of R27 and E77 are in close proximity to be able to form a salt-bridge (Figure 3.11a). In addition, mutation of R27 into A27 reduced the transition

temperature of thermal denaturation (T_m) by ~20 °C (Figure 3.10b), confirming the presence of the R27-E77 salt bridge.

Table 3.1 Experimental restraints and structural statistics for ten lowest-energy NMR structures of CTD_{Mi} out of 100 calculated structures.

NMR distance and dihedral constrains	
Distance constraints	
Total NOE	2910
Intra-molecule	2697
Intra-residue	899
Sequential (|i-j| = 1)	736
Medium-range($2 \leqq \|i-j\| \leqq 4$)	760
Long-range($\|i-j\| \geqq 5$)	302
Inter-molecule	171
Total dihedral angle restraints*	116
Structure statistics	
Violations, mean ± SD	
Distance constraints, Å	0.42 ± 0.06
Dihedral angle violation, °	3.46 ± 0.58
Max. distance constraint violation, Å	0.49
Max. dihedral angle violation, °	4.64
Ramachandran plot	
Most favoured regions	90.1%
Additional allowed regions	9.3%
Generously allowed regions	0.3%
Disallowed regions	0.3%
Average rms deviation, Å†	
Heavy atoms	0.82 ± 0.11
Backbone atoms	0.50 ± 0.14

*Dihedral angle constraints were generated by TALOS based on C^{α} and C^{β} chemical shifts.

†Average rms deviation in the structural region (residue 17-123) was calculated among 10 refined structures.

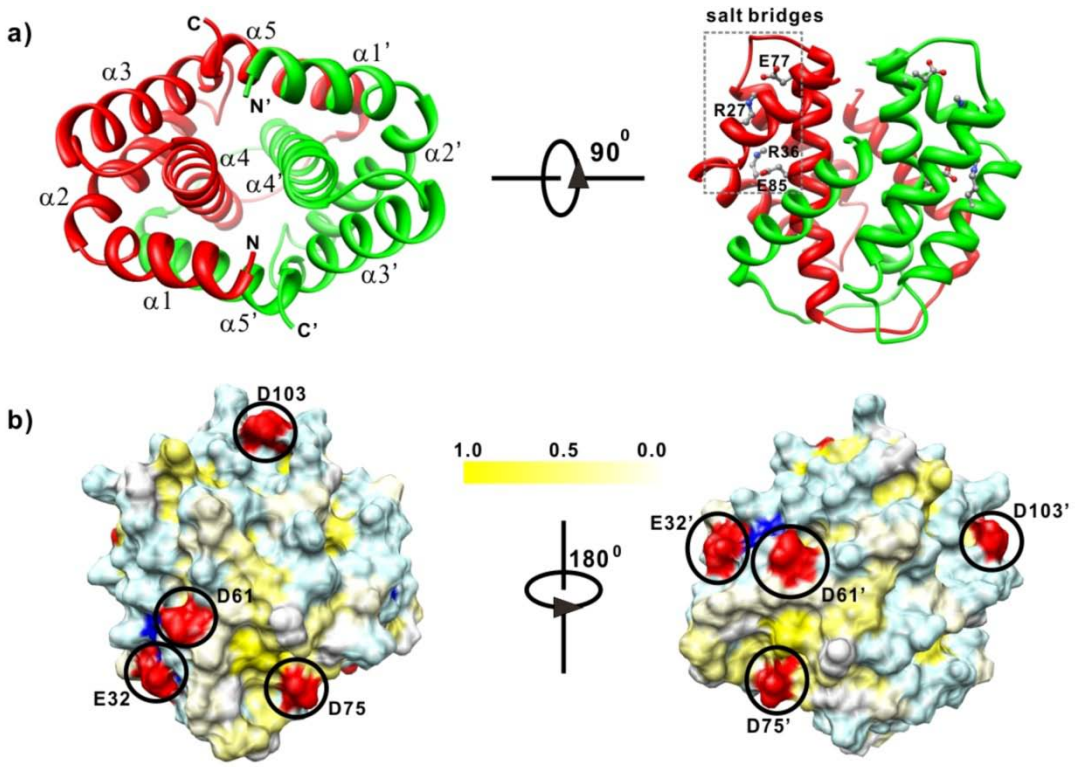


Figure 3.9 Structure and surface plots of CTD_{Mi}. (a) Cartoon drawing of the lowest energy conformer of CTD_{Mi}. The two monomers are indicated by two different colors. (b) Hydrophobic and charged surface of CTD_{Mi}. Hydrophobic residues are colored by a scale based on normalized hydrophobicity values: Phe (1.0) for yellow, Val (0.57) for light yellow and Gly (0.0) for white. Positively charged, negatively charged and polar residues are colored by blue, red and light blue.

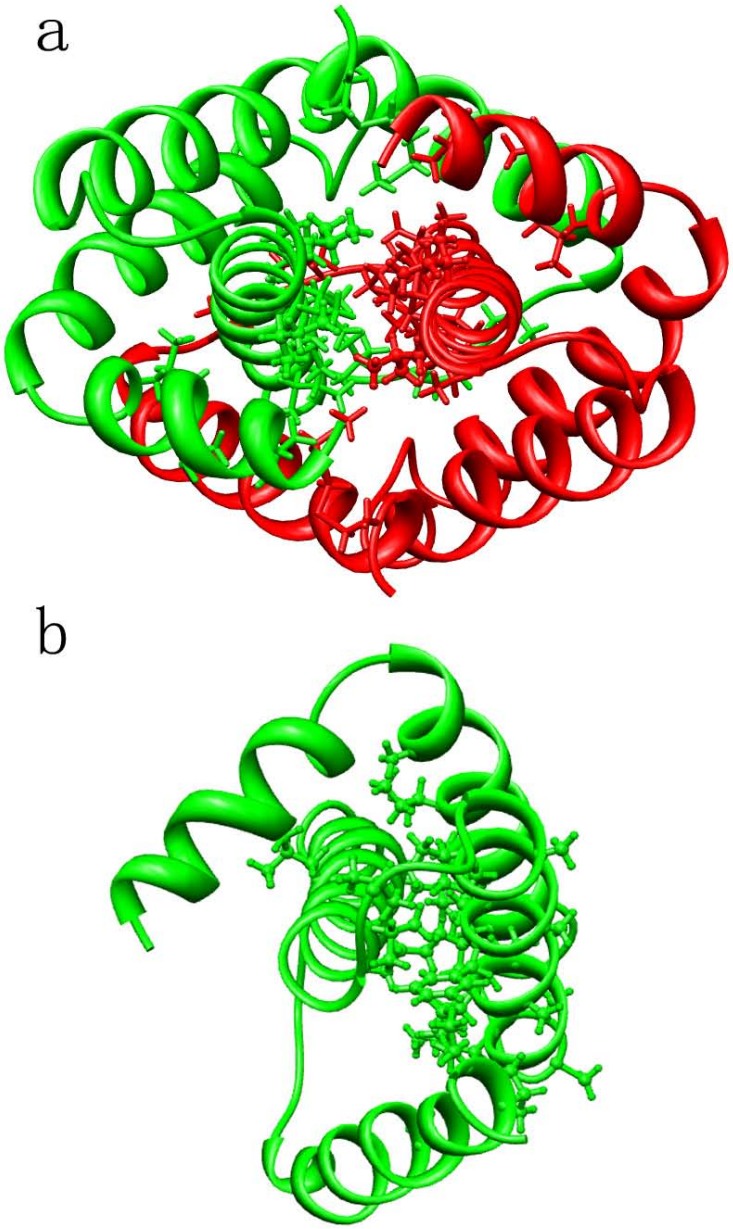


Figure 3.10 Hydrophobic core of CTD_{Mi}. (a) Cartoon drawing of a dimer of CTD_{Mi}. The two monomers are indicated by two different colors. The side chains of hydrophobic residues located at dimer interface are shown as ball and stick. (b) Cartoon drawing of a monomer of CTD_{Mi}. The side chains of hydrophobic residues involved in the hydrophobic core within one monomer are shown as ball and stick.

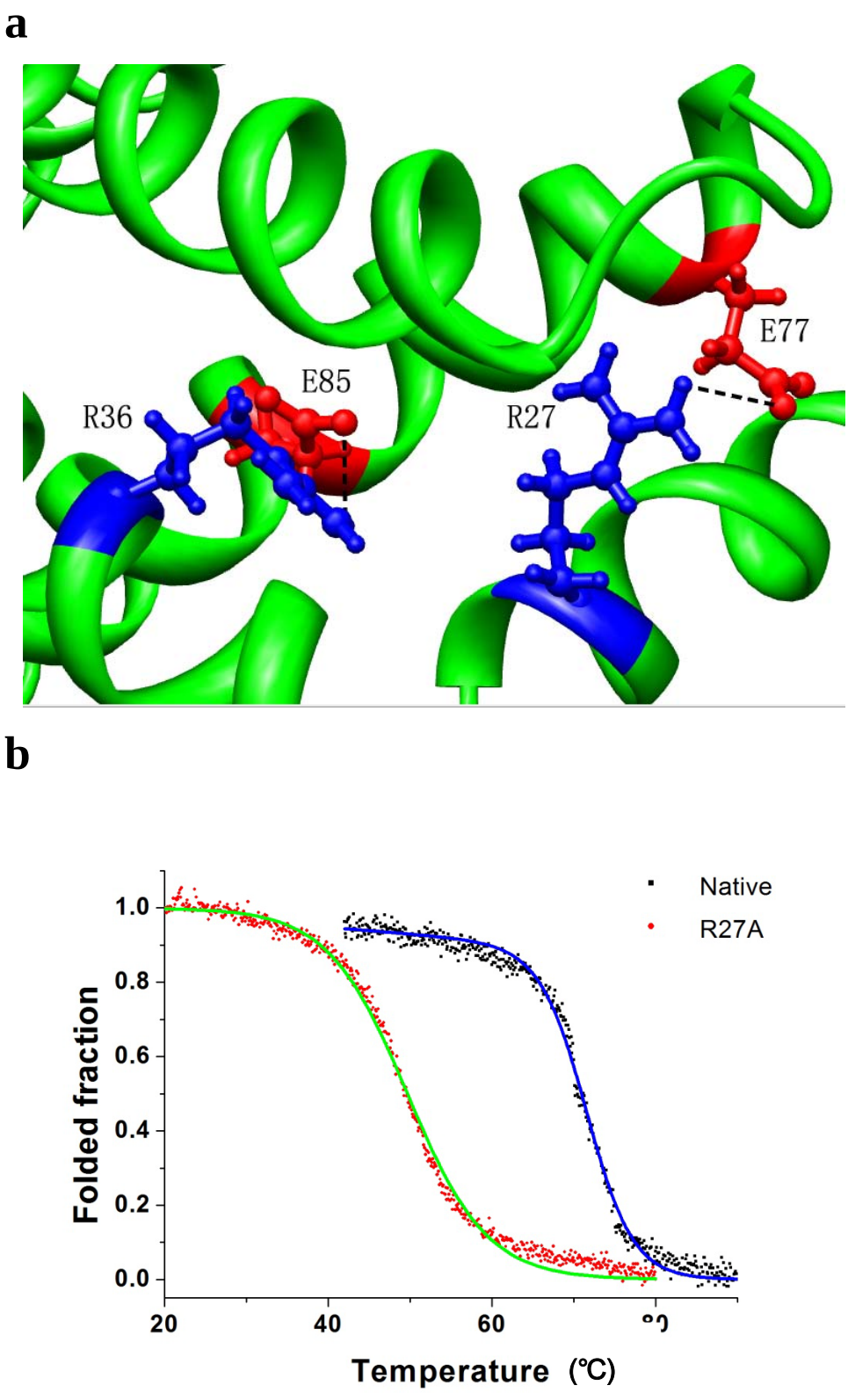


Figure 3.11 Salt bridges of CTD_{Mi}. (a) Ball and stick show of two salt bridges within one monomer of CTD_{Mi}. Positively and negatively charged residues are indicated by blue and red, respectively. (b) Temperature-induced denaturation of CTD_{Mi} and its mutant. The curves are fitted.

3.2.4 Comparison of CTD_{Mi} and CTD_{Ma}

The overall structure of CTD_{Mi} is very similar to the previously reported structure of CTD_{Ma}(Hagn, Eisoldt et al. 2010) with a Dali Z-score of 15(Holm and Rosenström 2010) (Figure 3.12). In addition, both CTD_{Mi} and CTD_{Ma} contain two intra-molecular salt bridges and have many hydrophilic residues located on the surface. Nevertheless, there are several key differences in local structures. 1) For CTD_{Mi} dimer, eight negatively charged carboxyl groups (four in each monomeric unit: E32, D61, D75, D103) are exposed on the protein surface (Figure 3.9b), but no net charges on the surface of CTD_{Ma} (Figure 3.13). Note that all positively charged residues are involved in salt-bridges for both CTD_{Mi} and CTD_{Ma}. 2) CTD_{Mi} contains no cysteine residue and there is no intermolecular disulfide bridge, but one intermolecular disulfide bond exists in CTD_{Ma} dimer (Hagn, Eisoldt et al. 2010) (Figure 3.12). 3) There are more hydrophobic residues located in between $\alpha5$ and $\alpha3$ and between $\alpha5$ and $\alpha1'$ in CTD_{Mi} than in CTD_{Ma} (Figure 3.14).

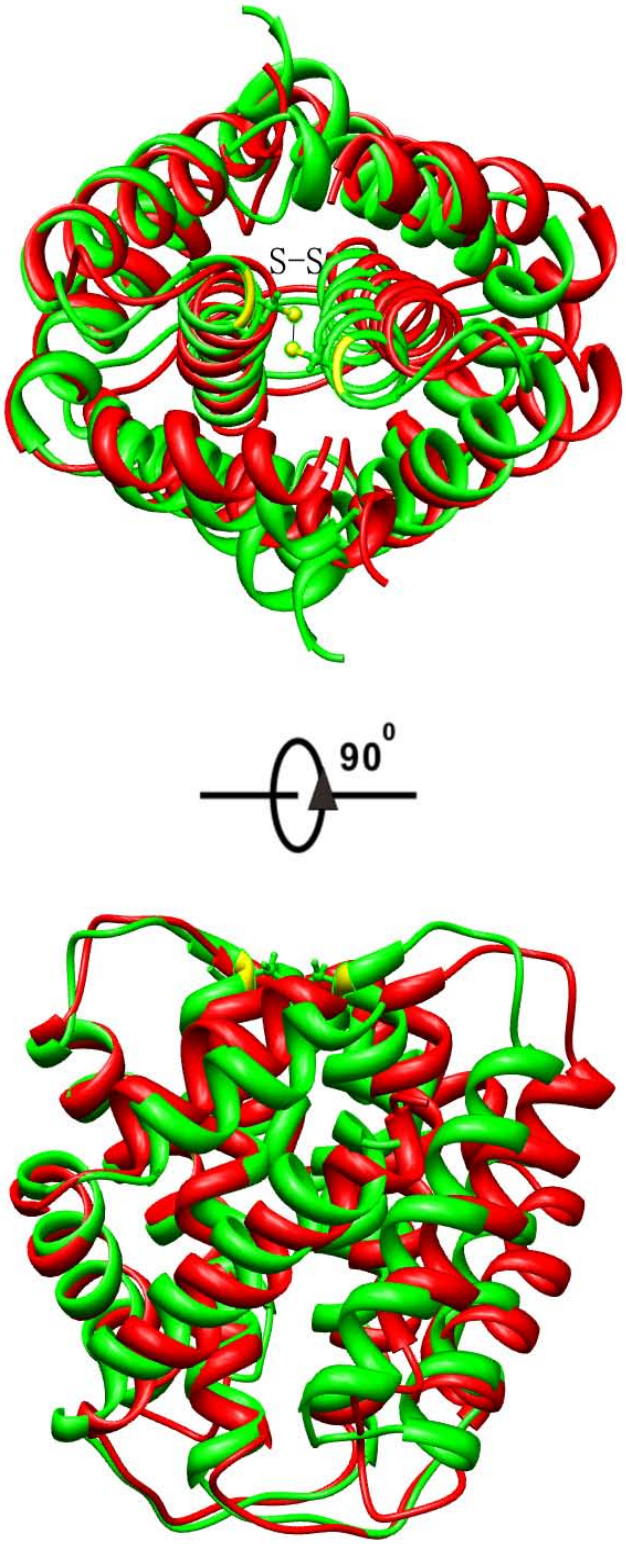


Figure 3.12 Superimposition of CTD_{Mi} and CTD_{Ma}. Cartoon drawing of CTD_{Mi} (red) and CTD_{Ma} (green), which are superimposed together. The disulfide bridge of CTD_{Ma} is labeled as yellow.

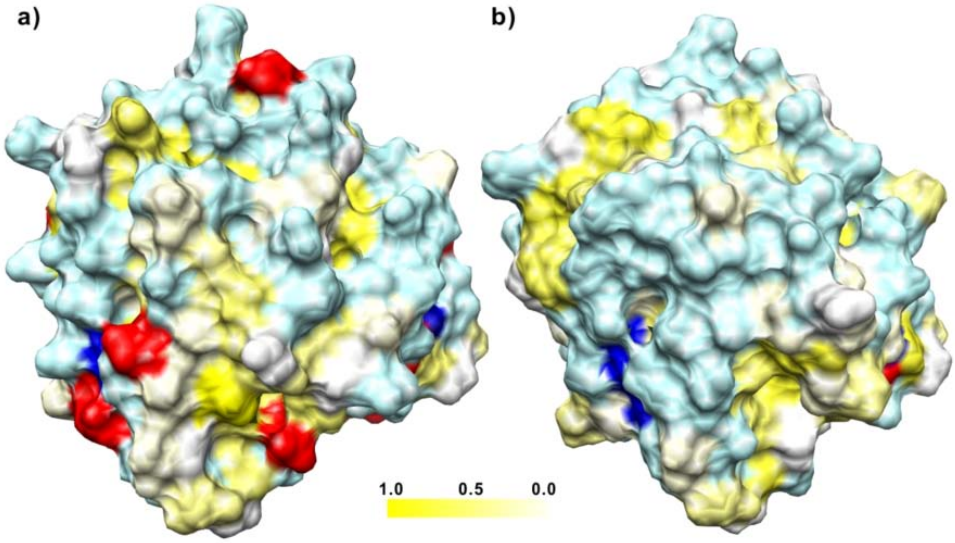


Figure 3.13 Comparison of surface plots of CTD_{Mi} (a) and CTD_{Ma} (b). Hydrophobic residues are colored by a scale based on normalized hydrophobicity values: Phe (1.0) for yellow, Val (0.57) for light yellow and Gly (0.0) for white. Positively charged, negatively charged and polar residues are colored by blue, red and light blue.

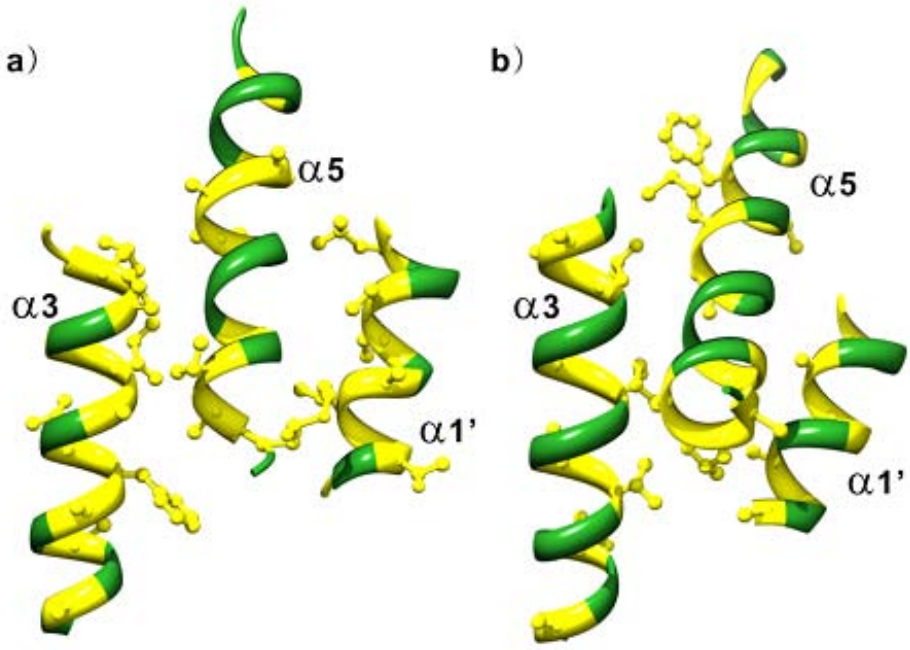


Figure 3.14 Comparison of hydrophobic interactions between α5 and α3 and between α5 and α1' for CTD_{Mi} (a) and CTD_{Ma} (b). The hydrophobic residues are labeled as yellow and their side chains are shown as ball and stick. The rest residues are labeled as green.

3.3 LK_{Mi}

3.3.1 Expression and purification of LK_{Mi} and LK-CTD_{Mi}

The LK_{Mi} (89 aa) contains 46.1% Gly and 32.3% Ala (Figure 3.15a). It was predicted to be intrinsically disordered and quite hydrophobic (Figure 3.15b and Figure 3.16a). To determine the secondary structure of LK_{Mi} experimentally, it was firstly expressed in *E. coli* using pET-M expression vector. However, no obvious expression of LK_{Mi} was noticed under any tested conditions by comparing the SDS PAGE results of whole cell samples both before and after induction. The failure of expression of LK_{Mi} might be because the intrinsically disordered characteristic of LK_{Mi} could lead to the degradation of LK_{Mi} right after expression. Then, the His-tag was replaced by GST-tag which is larger and can protect the LK_{Mi} from degradation. The expression of this construct was successful. Nevertheless, LK_{Mi} was degraded again when thrombin was used to remove the GST-tag. Thus, the attempt to obtain single LK_{Mi} could not be fulfilled.

LK-CTD_{Mi} was created to indirectly study the structure of LK_{Mi}. The expression of LK-CTD_{Mi} was successful (Figure 3.17a). The purification procedure of LK-CTD_{Mi} was the same as CTD_{Mi}. LK-CTD_{Mi} also forms dimer as evidenced by the SEC result (Figure 3.17b). ^{15}N labeled LK-CTD_{Mi} was prepared to record the ^{1}H-^{15}N HSQC. A comparison of the ^{1}H-^{15}N HSQC spectra of the LK-CTD_{Mi} and CTD_{Mi} reveals that the backbone ^{1}H-^{15}N correlation peaks for the residues from the LK domain are located in the range of 7.7 – 8.5 ppm in the ^{1}H dimension and most Gly and Ala ^{1}H-^{15}N correlations are clustered together (Figure 3.18). The result shows that the LK domain is indeed intrinsically disordered. Except for the correlation peaks from the N-terminal region of the isolated CTD_{Mi} (e.g., V17, G18 and T20), other peaks from the isolated CTD_{Mi} have the same ^{1}H and ^{15}N chemical shifts as those from the CTD in the bi-domain LK-CTD_{Mi}. Note that V17 is the N-terminal end residue of the CTD domain in the LK-CTD_{Mi} construct. The signal of G48 in the bi-domain was weak and is not observable in the current plot, but it had the same chemical shifts as the G48 in the isolated CTD. The result indicates that there are no

or only weak interactions between the two domains at a protein concentration equal to or less than 0.5 mM. To evaluate the aggregation propensity of the LK_{Mi}, the aggregation-prone regions were predicted. The prediction shows that LK_{Mi} contains three aggregation-prone regions with aggregation propensity scores > 1: Y4-A12, A27-A37 and G67-A74 (Figure 3.16b).

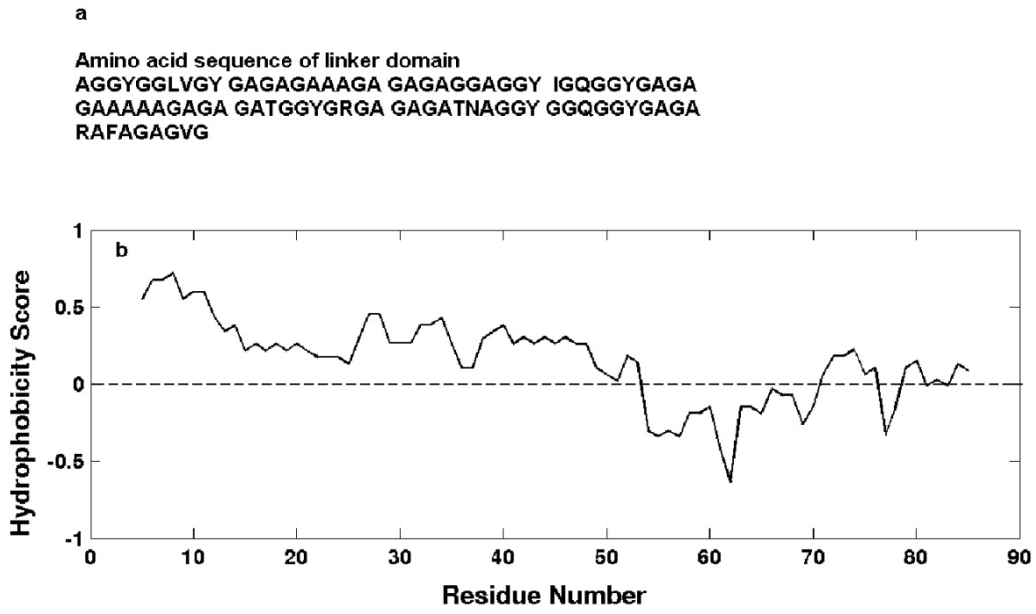


Figure 3.15 Amino acid sequence of LK_{Mi} (a) and its hydrophobicity plot (b).

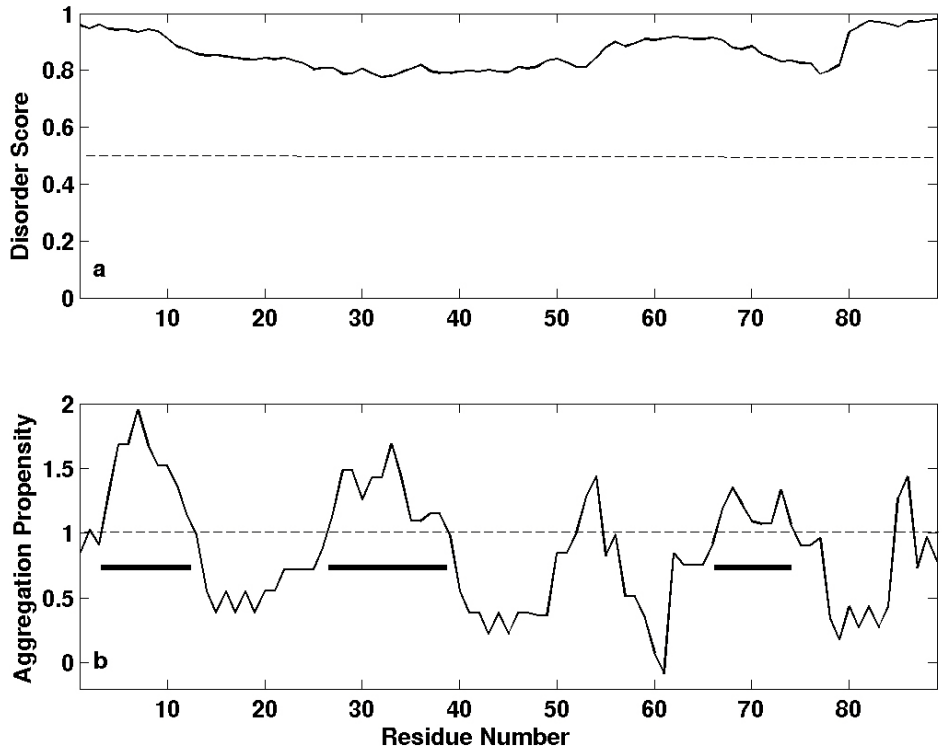


Figure 3.16 Predicted disordered residues (a) and aggregation regions (b) in LK_{Mi}. Disordered residues have scores > 0.5. The aggregation-prone regions with aggregation propensity scores > 1 are indicated by bars in panel b.

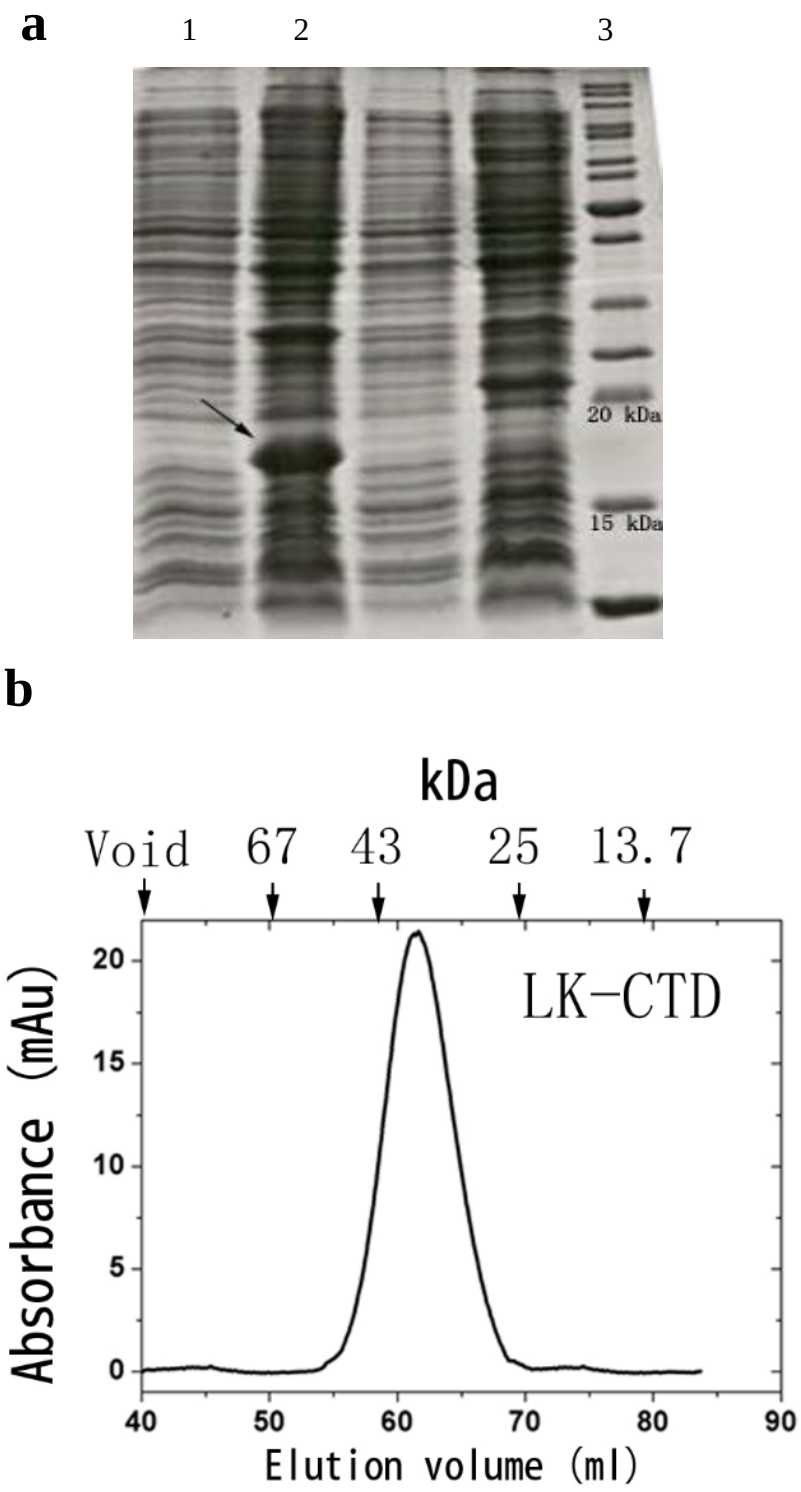


Figure 3.17 Expression and purification of LK-CTD$_{Mi}$. (a) SDS PAGE of LK-CTD$_{Mi}$. Line 1: non-induced whole cell sample; Line2: induced whole cell sample; Line 3: protein marker. The over expressed LK-CTD$_{Mi}$ is indicated by the arrow. (b) Size exclusion chromatography profile of LK-CTD$_{Mi}$ (18.5 kD). Molecular weight makers are indicated on the top. The buffer condition was 10 mM phosphate at pH 6.8.

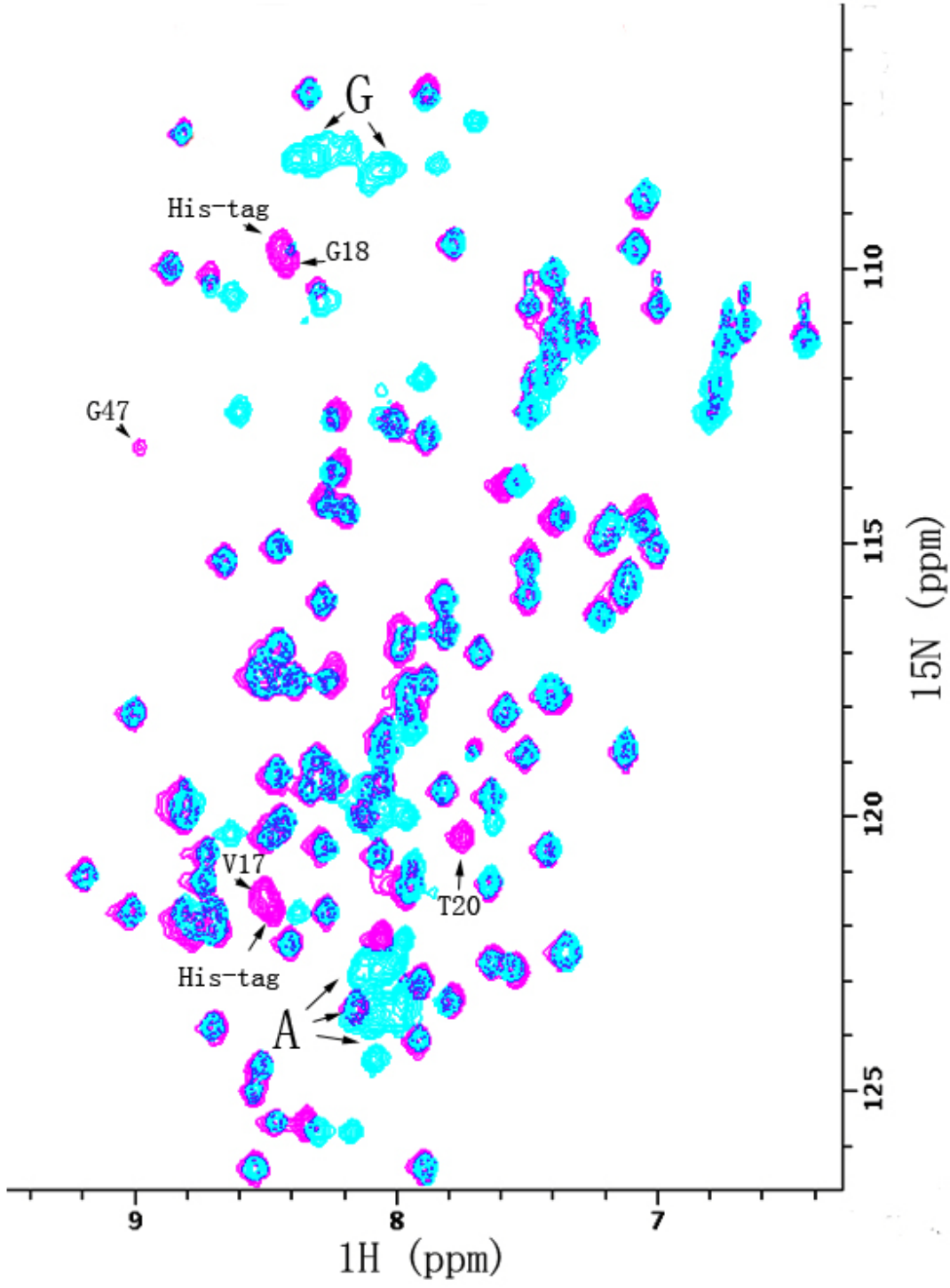


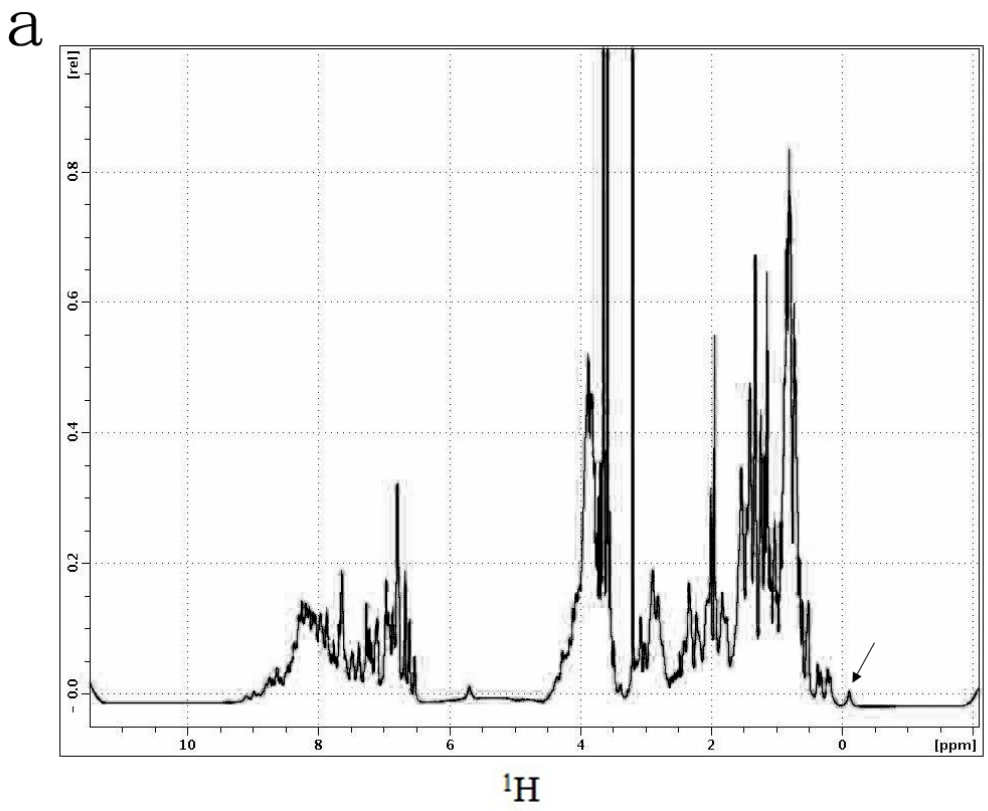
Figure 3.18 Overlay of 2D ^{1}H-^{15}N HSQC spectra of isolated CTD$_{Mi}$ (red) and di-domain LK-CTD$_{Mi}$ (cyan). The signals from the His-tag in the isolated CTD$_{Mi}$ were labeled. Correlation peaks from Ala and Gly of the LK domain in LK-CTD$_{Mi}$ are clustered in small regions and are indicated by arrows.

3.4 RP_{Mi}

3.4.1 Expression and purification of RP_{Mi}

RP_{Mi} was subcloned into pET-M expression vector and transformed into *E. coli BL21(DE3)*. RP_{Mi} was successfully expressed using similar conditions with CTD_{Mi}. However, the expressed protein was mainly in inclusion body which is not soluble in normal aqueous buffer. Thus, the protein was dissolved by 8M urea and refolded after purified by Ni-NTA column. The rest of purification procedures were the same as those of CTD_{Mi}.

The 1D ^{1}H NMR spectrum of unlabeled RP_{Mi} shows that this domain adopts a well folded 3D structure since its methyl proton signals display very good dispersion with one methyl at -0.13 ppm (Figure 3.19a). In addition, the α-helical structure is dominant in the structure of RP_{Mi} (Figure 3.19b). However, RP_{Mi} has oligomerization-prone feature, which will be discussed in detail in following sessions. The oligomerization-prone feature could compromise the NMR experiments for structure solving. After testing different buffer conditions and protein concentrations, 0.4 mM protein in 10 mM Tris with 100 mM NaCl at pH 7.0 was chosen as the final condition of the NMR sample. Under this condition, the protein could maintain the monomeric form long enough to record all the NMR experiments used for structure determination.



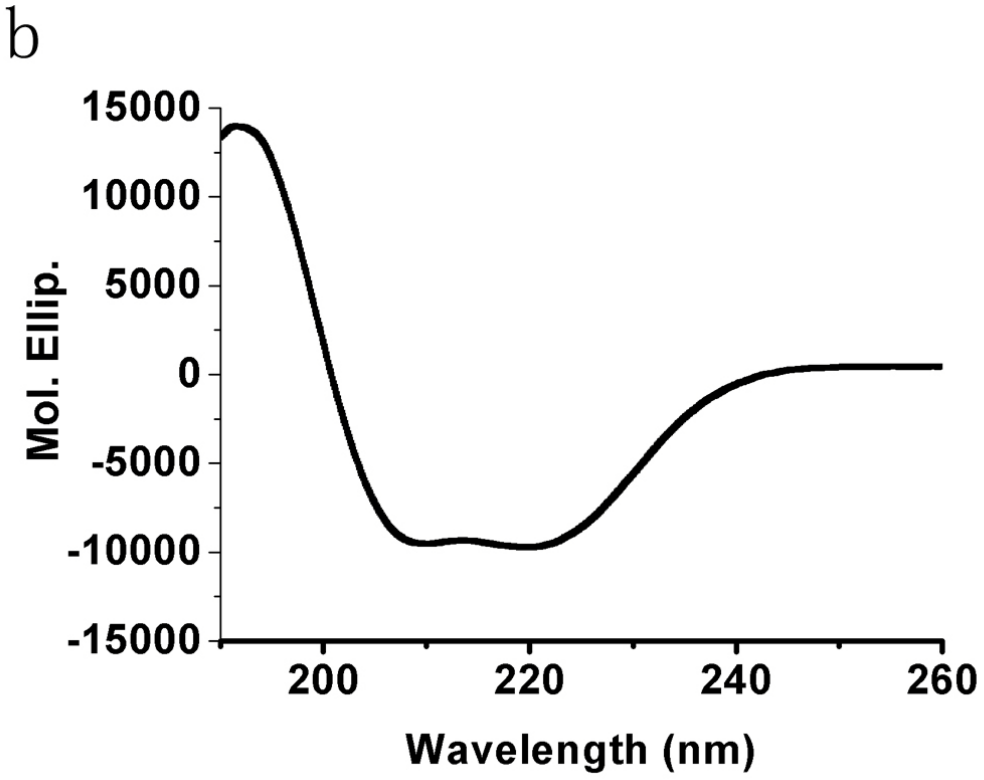


Figure 3.19 1D NMR (a) and CD (b) spectra of RP_{Mi}. The methyl proton signal at -0.13 ppm in the 1D NMR spectrum is indicated by the arrow.

3.4.2 Resonance assignments of RP_{Mi}

The backbone resonance assignments of RP_{Mi} were obtained unambiguously using the same strategy used for CTD_{Mi} (Figure 3.20). The assignments of the N-terminal region (Met^{1}-Ser^{16}) corresponding to His-tag and thrombin cleavage site and the C-terminal unstructured region (Ter^{143}-Gly^{148}) were missing. Besides these unstructured regions, only 8 other residues (Gly^{32}, Ter^{44}, Asn^{85}, Ser^{88}, Glu^{113}, Ser^{114}, Ser^{117} and Ser^{118}) could not be assigned which are all found at loop regions. With the information of backbone assignments, >90% of side chain resonance assignments of RP_{Mi} were achieved. In total, 1290 NOE resonances were assigned from 4D ^{15}N, ^{13}C-edited NOESY spectrum.

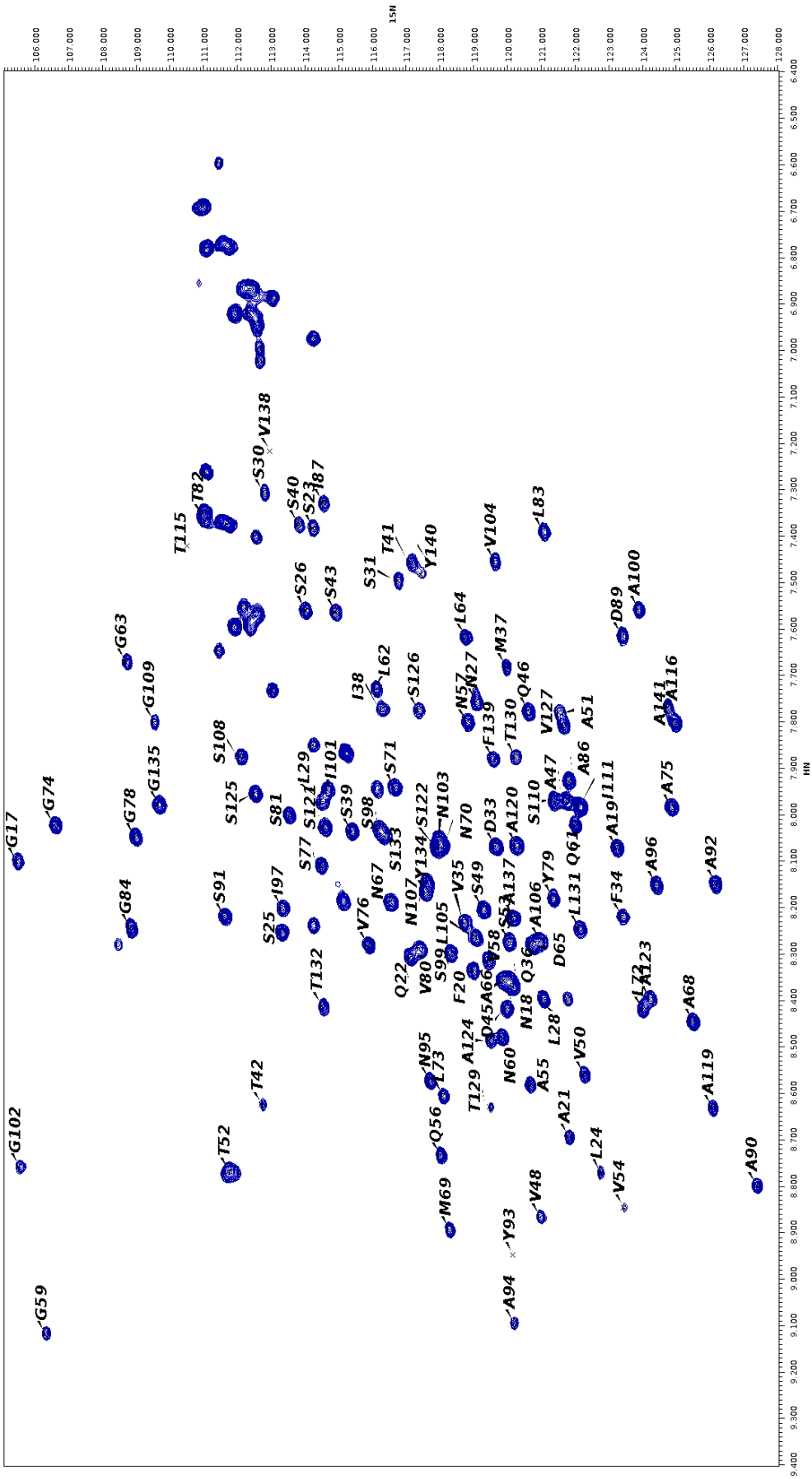


Figure 3.20 Assigned ^{1}H, ^{15}N HSQC spectrum of RP_{Mi}

3.4.3 Solution structure of RP_{Mi}

1290 distance constrains and 194 angular constrains were input to CYANA to calculate the structure of RP_{Mi} (Table 3.2). The final ten lowest energy structures have backbone and heavy atoms r.m.s. deviations of 0.99 Å and 1.28 Å, respectively (Table 3.2). Ramachandran plot of the final ten structures shows that 99.1 % of the residues have backbone dihedral angles in allowed regions (Table 3.2).

In aquous solution, the structure of RP_{Mi} adopts a globular fold which is composed of a compact seven-helix bundle (Gly^{17}-Ala^{141}) and a unstructured region (Pro^{142}-Gly^{148}) (Table 3.2 and Figure 3.21). The region corresponding to His-tag and thrombin cleavage site is unstructured and has no interaction with RP_{Mi} residues as evidenced by that the amide correlations of this region in HSQC spectrum are absent and no NOE from the interaction of this region and RP_{Mi} residues was observed. The seven helices (α1 [Asn^{18}-Leu^{29}], α2 [Asp^{33}-Ser^{40}], α3 [Thr^{44}-Leu^{62}], α4 [Ala^{66}-Thr^{82}], α5 [Ala^{90}-Ser^{108}], α6 [Ser^{118}-Thr^{132}], α7 [Ala^{137}-Tyr^{140}]) are intercepted by short turns. The packing of the seven helices is mainly mediated by hydrophobic interactions. A hydrophobic core around α3 is shielded by the rest of helices in a barrel-like manner. All the charged residues are displayed on the surface of the structure and expose their side chains to the solvent. Although most hydrophobic residues are buried inside of the structure, an exposed hydrophobic patch on the surface is found at the C-terminal region (Figure 3.21c). This hydrophobic patch is contributed mainly by the short α7 which is composed of five hydrophobic residues (Ala^{127}, Val^{128}, Phe^{129}, Tyr^{140} and Ala^{141}).

Table 3.2 Experimental restraints and structural statistics for ten lowest-energy NMR structures of RP_{Mi} out of 100 calculated structures.

NMR distance and dihedral constrains	
Distance constraints	
Total NOE	1290
Intra-residue	381
Sequential (| I-j| = 1)	286
Medium-range(2≦|i-j|≦4)	242
Long-range(|i-j|≧5)	381
Total dihedral angle restraints*	194
Structure statistics	
Violations, mean ± SD	
Distance constraints, Å	0.27 ± 0.07
Dihedral angle violation, °	4.43 ± 0.37
Max. distance constraint violation, Å	0.34
Max. dihedral angle violation, °	4.8
Ramachandran plot	
Most favoured regions	88.5%
Additional allowed regions	8%
Generously allowed regions	2.7%
Disallowed regions	0.9%
Average rms deviation, Å†	
Heavy atoms	0.99 ± 0.15
Backbone atoms	1.28 ± 0.14

*Dihedral angle constraints were generated by TALOS based on C^{α} and C^{β} chemical shifts.

†Average rms deviation in the structural region (residue 17-82, 90-107 and 118-141) was calculated among 10 refined structures.

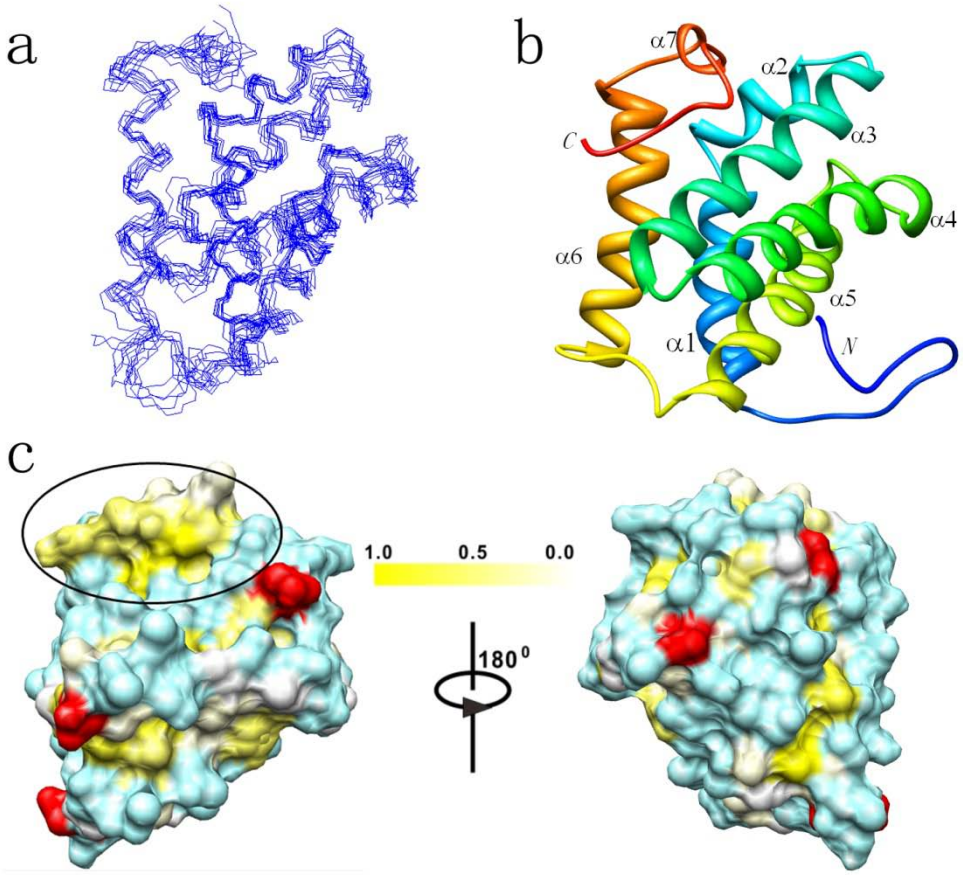


Figure 3.21 Solution structure of RP_{Mi}. (a) Backbone drawing of 10 lowest energy structures of RP_{Mi} in line mode (both C- and N-terminal unstructured regions are not included). (b) Cartoon drawing of the lowest energy conformer of RP_{Mi} with rainbow color. (c) Hydrophobic and charged surface of RP_{Mi} (both C- and N-terminal unstructured regions are not included). Hydrophobic residues are colored by a scale based on normalized hydrophobicity values: Phe (1.0) for yellow, Val (0.57) for light yellow and Gly (0.0) for white. Positively charged, negatively charged and polar residues are colored by blue, red and light blue. The exposed hydrophobic patch is indicated by a circle.

3.4.4 Comparison of the structure of RP_{Mi} with other known structures

The homologous structures of RP_{Mi} were searched online with DALI server(Holm and Rosenström 2010). Three structures were found having significant similarity with the structure of RP_{Mi} (DALI Z-socar >9). They are the repetitive domains of AcSp1 (RP_{Ac}, PDB code: 2LYI) and TuSp1 ($RP1_{Tu}$, PDB code: 2K3N; $RP2_{Tu}$, PDB code: 2K3O) from the same species, *N. antipodiana*(Lin, Huang et al. 2009; Wang, Huang et al. 2012). Although the sequence identities between RP_{Mi} and the other three RP domains are not high (20%-30%) (Figure 3.22a), the overall structures are surprisingly similar, particularly for the first five helices (Figure 3.22b). This structural conservation of RP domains indicates important functional roles of RP domains in protein storage and fiber formation. Nevertheless, significant local structure differences are noticed. First of all, the lengths of α7 of these four structures are different. The α7 of RP_{Mi} is only composed of four hydrophobic residues which have hydrophobic interactions with α6 and α3. The α7 of $RP1_{Tu}$ (8 residues) is longer than that of RP_{Mi} but no hydrophobic interactions between it and other helices are found. Thus, its orientation was not determined and it is not considered in the following structure comparison. The α7 is absence in $RP2_{Tu}$. RP_{Ac} has the longest α7 which has strong hydrophobic interactions with α1 and α6. Second of all, the hydrophobic interactions among the helices of RP_{Mi} are weaker than those of the rest three RP domains. For example, the slight different orientation of α6 of RP_{Mi} compared to the other α6s makes it has longer distances from α4 and α5 in RP_{Mi} than those in other RP domains (Figure 3.23). The overall packing of the helices in RP_{Mi} is not as tight as in the other three RP domains. Moreover, only two long side-chain hydrophobic residues (Phe, Val, Leu and Ile) are found at the C-terminus of α6 of RP_{Mi}, whereas five are found in the α6 of RP_{Ac} (Figure 3.23a, b). Although only three long side-chain hydrophobic residues are identified in the α6 of both $RP1_{Tu}$ and $RP2_{Tu}$, they all locate at the N-terminus of α6 which is closer to α4 and α5 than the C-terminus of α6 (Figure 3.23c, d). The hydrophobic interactions between α6 and α3 in RP_{Mi} are also weaker than those in the rest of RP domains because of the number

and location of the long side-chain hydrophobic residues in α6 of RP_{Mi} (Figure 3.23). Finally, an exposed hydrophobic patch is noticed on the hydrophobic and charged surface plot of RP_{Mi} (Figure 3.21c). The surface of RP_{Ac} is extremely hydrophilic since it mainly displays charged and polar residues (Figure 3.24a). Some hydrophobic residues can be found on the surface of $RP1_{Tu}$ and $RP2_{Tu}$. However, instead of forming exposed hydrophobic patch like RP_{Mi}, they either form long hydrophobic groove or scatter on the surface (Figure 3.24b, c), which could not cause aggregation in aqueous solution. The weak hydrophobic interaction between helices and the relatively hydrophobic surface contribute to the poor stability and the oligomerization-prone feature of RP_{Mi}.

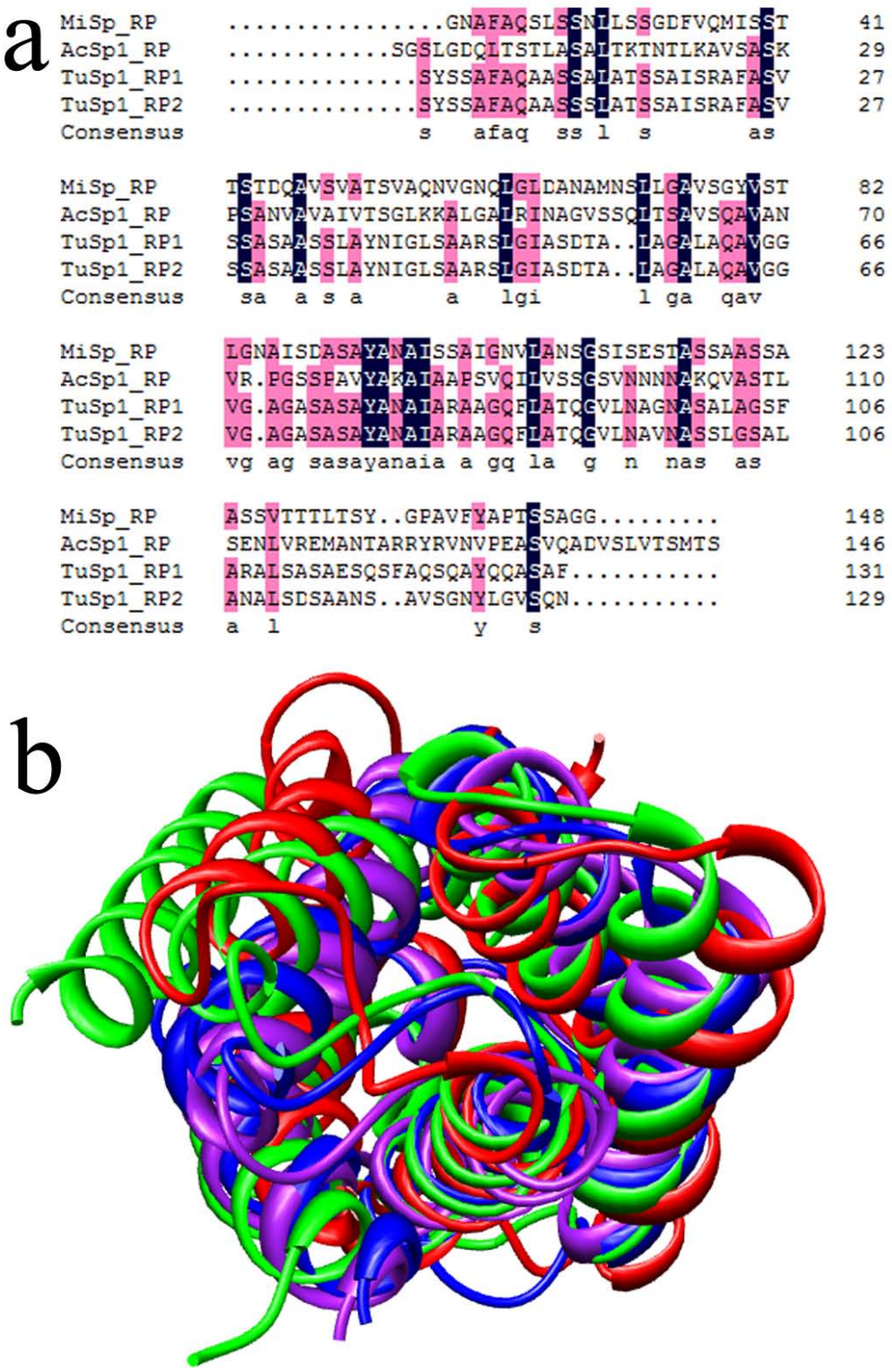


Figure 3.22 Alignments of the four RP domains. (a) Sequence alignment of RP_{Mi}, RP_{Ac}, $RP1_{Tu}$ and $RP2_{Tu}$. (b) Cartoon drawing of structure alignment of RP_{Mi} (red), RP_{Ac} (green), $RP1_{Tu}$ (blue) and $RP2_{Tu}$ (purple).

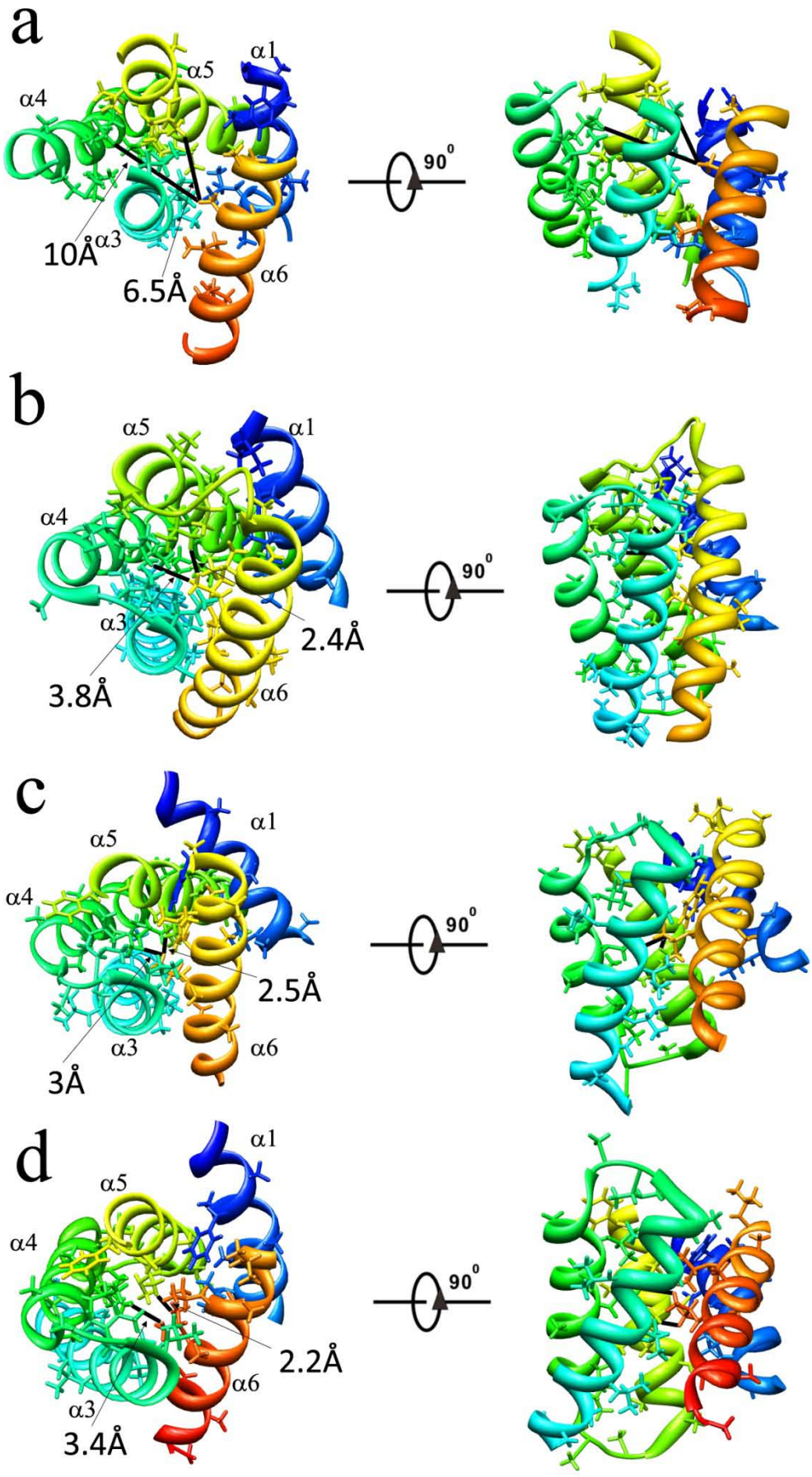

a
α1
α5
α4
10Å
α3
6.5Å
α6
90°
b
α5
α1
α4
α3
2.4Å
3.8Å
α6
90°
c
α1
α5
α4
2.5Å
α3
α6
3Å
90°
d
α1
α5
α4
2.2Å
α6
α3
3.4Å
90°

Figure 3.23 Cartoon drawing of selected helices of RPMi (a), RPAc (b), RP1Tu (c) and RP2Tu (d) in rainbow color. The side chains of the hydrophobic residues in the selected helices are shown as stick. The two shortest distances between α6 and α4 and α5 in each structure are indicated.

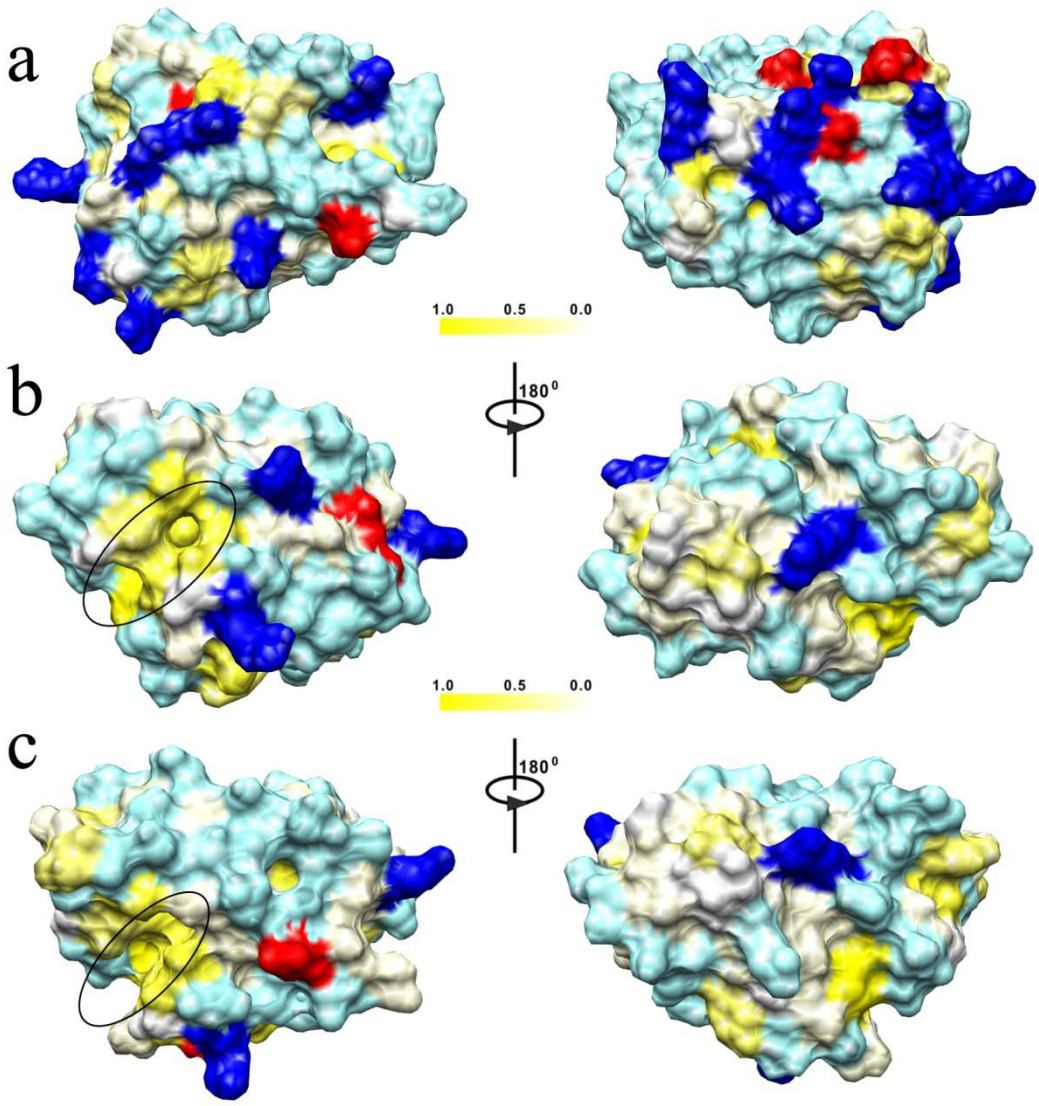


Figure 3.24 Hydrophobic and charged surface plot of RP_{Ac} (a), $RP1_{Tu}$ (b) and $RP2_{Tu}$ (c). Hydrophobic residues are colored by a scale based on normalized hydrophobicity values: Phe (1.0) for yellow, Val (0.57) for light yellow and Gly (0.0) for white. Positively charged, negatively charged and polar residues are colored by blue, red and light blue. The hydrophobic grooves of $RP1_{Tu}$ and $RP2_{Tu}$ are indicated by circles.

3.4.5 Oligomerization-prone feature of RP_{Mi}

RP_{Mi} existed mainly in monomer at low protein concentration (<1 mg/ml) in the presence of 100 mM NaCl, but mainly in dimer together with trimer in the absence of salt as indicated by SEC (Figure 3.25). Very likely, the monomer, dimer and larger oligomers are in dynamic equilibrium since a single SEC peak was observed. The 1D ^{1}H NMR spectra shows that the line width of the methyl signal of RP_{Mi} at -0.13 ppm was 40 Hz under the condition of 0.6 mM protein, 10 mM sodium phosphate and pH 6.8, which was significantly larger than that for the methyl signal of CTD_{Mi} at 0.03 ppm (26 Hz) (Figure 3.26). Because the NMR line width is proportional to the molecular size, the apparent size of RP_{Mi} must be significantly larger than the size of CTD_{Mi} under the condition of 0.6 mM protein, 10 mM sodium phosphate and pH 6.8. As mentioned earlier, RP_{Mi} (128 aa) consists of only ~20% more aa than CTD_{Mi} (107 aa) and CTD_{Mi} is dimeric in aqueous solution. Thus RP_{Mi} should exist in equilibrium between dimer and small oligomers at low salt concentration (10 mM sodium phosphate) and relatively high protein concentration (0.6 mM). This result also shows the interaction among RP_{Mi} molecules in the oligomers is relatively weak in solution, consistent with the SEC result (Figure 3.25). The concentration dependent weak association of RP_{Mi} molecules should be mediated mainly by the hydrophobic interaction among the surface exposed hydrophobic patches of RP_{Mi} molecules (Figure 3.21c).

At low protein concentrations (<1 mg/ml), RP-LK-CTD_{Mi} existed as a dimer on the basis of the SEC data (Figure 3.25). The dimerization should be mediated through the CTD. The 1D ^{1}H NMR spectra of RP_{Mi}, CTD_{Mi} and RP-LK-CTD_{Mi} (Figure 3.26) reveal that no strong interactions exist among the three different domains in RP-LK-CTD_{Mi} since the isolated methyl signals from RP-LK-CTD_{Mi} have nearly the same chemical shifts as those in the isolated domains. The line width of the methyl signals at 0.03 ppm and -0.13 ppm in the isolated CTD_{Mi} and RP_{Mi} were 25 Hz and 40 Hz respectively, very similar to those of the tri-domain fragment (27 Hz and 42 Hz) at a concentration of 0.6 mM, although the tri-domain is almost three times larger than

the individual CTD_{Mi} and RP_{Mi} in molecular weight. This can be explained by the independent motions of the CTD_{Mi} dimeric unit and RP_{Mi} domain due to the high flexibility of the disordered linker domain. When the RP-LK-CTD_{Mi} concentration was increased from ~0.6 mM to ~3 mM, the line width of the signal at 0.03 ppm (from the CTD) increased slightly from 27 Hz to 30 Hz, while the line width of the signal at -0.13 ppm (from the RP domain) increased dramatically from 42 Hz to 90 Hz. The result indicates that the tri-domain molecules assemble together to form oligomers through the weak association of RP domains from different molecules.

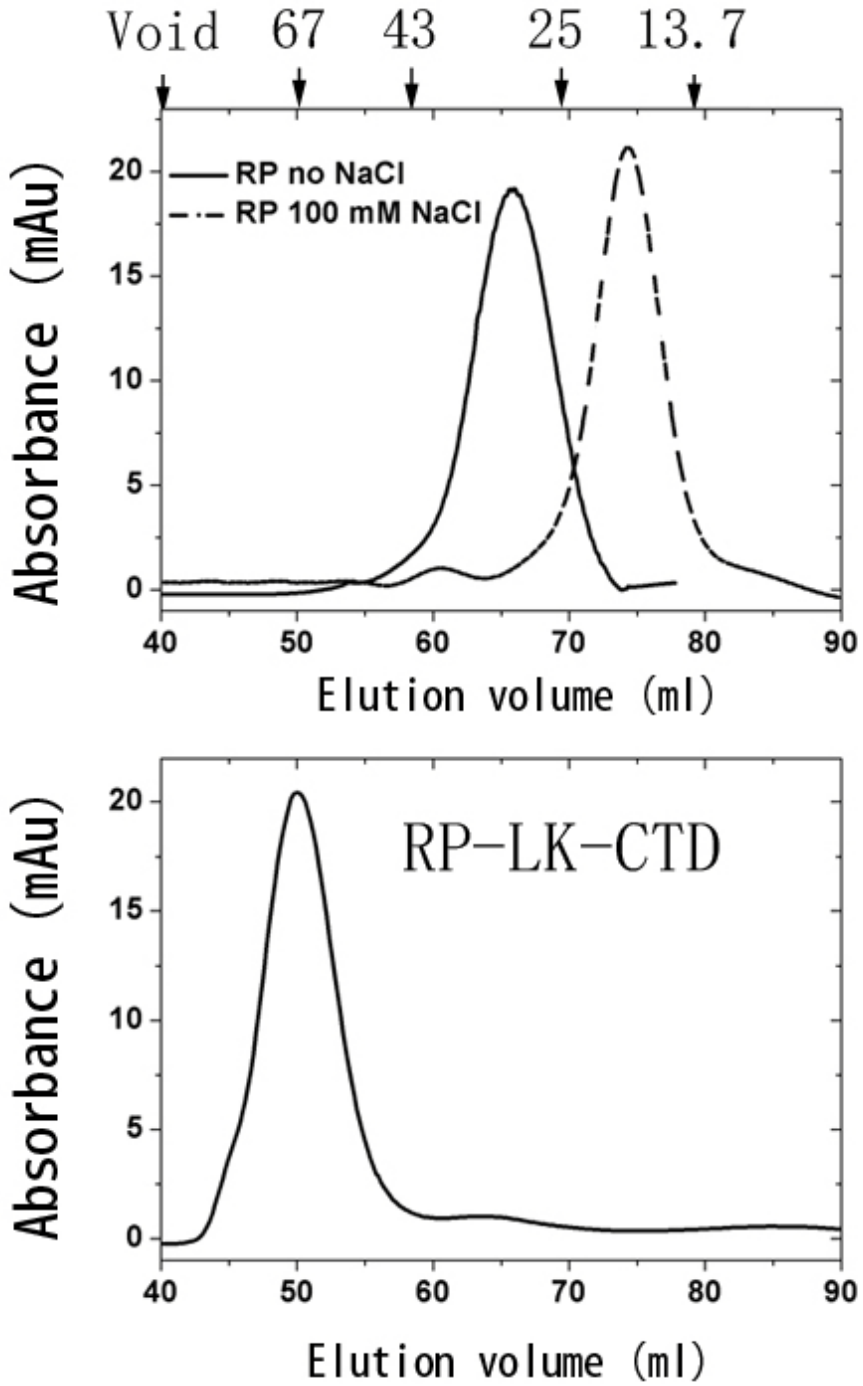


Figure 3.25 Size exclusion chromatography profile of RP_{Mi} (14.5 kD) (up panel) and RP-LK-CTDMi (31.8 kD) (bottom panel). Molecular weight makers are indicated on the top. Except for one RP_{Mi} (dashed curve) profile which was run in the presence of 100 mM NaCl, all other profiles were obtained under a buffer condition of 10 mM phosphate at pH 6.8.

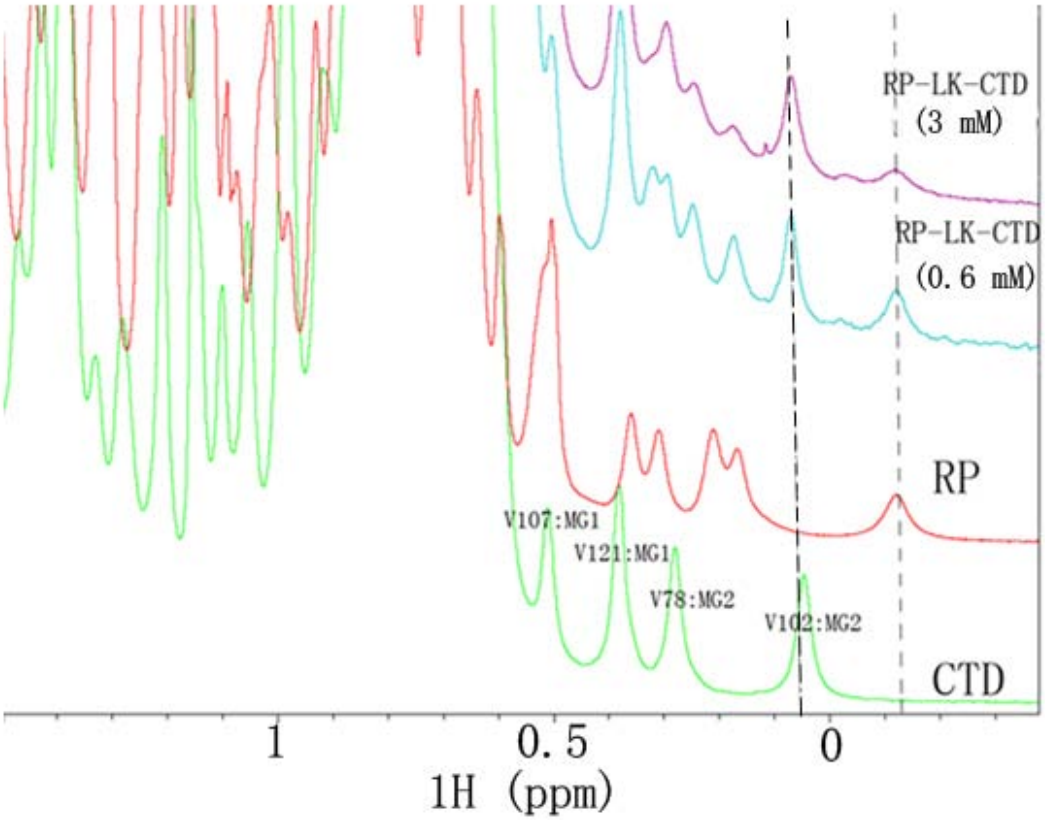


Figure 3.26 Stacked plot of 1D ^{1}H spectra (-0.5 ppm – 1.5 ppm) of CTD_{Mi} (green, 0.6 mM), RP_{Mi} (red, 0.6 mM), RP-LK-CTD_{Mi} (cyan, 0.6 mM) and RP-LK-CTD_{Mi} (pink, 3 mM).

3.5 Stability of CTD_{Mi}, RP_{Mi} , LK-CTD_{Mi} and RP-LK-CTD_{Mi}

3.5.1 Stability of CTD_{Mi}

Although CTD_{Ma} and CTD_{Mi} have similar overall structures, their chemical and thermal stabilities are significantly different. The transition midpoints in urea denaturation (C_m) and thermal denaturation (T_m) of CTD_{Mi} were ~4.8 M urea and ~71 °C, respectively (Figure 3.27a and Figure 3.28), which are significantly larger than those of CTD_{Ma} (~2 M urea at 10 mM phosphate and ~2.8 M urea at 500 mM NaCl, 64°C at 10 mM phosphate)(Hagn, Eisoldt et al. 2010). The result indicates CTD_{Mi} is much more stable than CTD_{Ma}. Interestingly, NaCl had nearly no effects on the chemical stability of CTD_{Mi} (Figure 3.27a), while NaCl could stabilize CTD_{Ma}(Hagn, Eisoldt et al. 2010). The stability of CTD_{Mi} was independent of protein concentration when the concentration was below 0.2 mM (Figure 3.29), but CTD_{Ma} was much more stable against urea denaturation at a protein concentration of 5 μM than 0.2 mM(Hagn, Eisoldt et al. 2010).

To investigate which part of CTD_{Mi} is less stable and tends to unfold first in the presence of urea, a urea titration of CTD_{Mi} monitored by ^{1}H-^{15}N HSQC spectra was carried out. Figure 3.30 shows that among the five helices, α1 has the most residues (~46%) with chemical shift change >0.1 ppm in the presence of 3.5 M urea, indicating that α1 is the easiest part to be unfolded by urea. This unstable feature of α1 can be caused by the lack of hydrogen bonding since only 15% residues of α1 are involved in hydrogen bonding as shown in figure 3.31. Similarly, α1 of CTD_{Ma} is also the most unstable region against urea(Hagn, Eisoldt et al. 2010).

To examine the importance of the solvent-exposed charges to the stability of CTD_{Mi}, four conserved single-point mutants (E32Q, D61N, D75N and D103N) and one double-point mutant (E32Q/D75N) were prepared. E32Q, D75N and D103N mutants showed significantly lower C_m values than the wild type CTD_{Mi} although the mutation of D61N had only a slight effect on the stability (Figure 3.32a). Moreover, double mutation reduced the C_m from ~4.8 M to ~3.2 M urea (Figure 3.32a). The T_ms of CTD_{Mi} mutants agree with their C_ms except that the T_m of D103N is slightly higher

than that of D75N whereas the C_m of D103N is lower than that of D75N (Figure 3.32). The results indicate that the solvent-exposed negatively charged residues are critical to the stability of CTD_{Mi}. Interestingly, these negatively charged residues are conserved or partially conserved in all MiSp, but absent in CTD_{Ma} of *A. diadematus* (Figure 1.5a). Besides the solvent-exposed negative charges, other factors such as hydrophobic interaction and hydrogen bonding which are slightly different in the two CTDs may also contribute to their significant difference in stability.

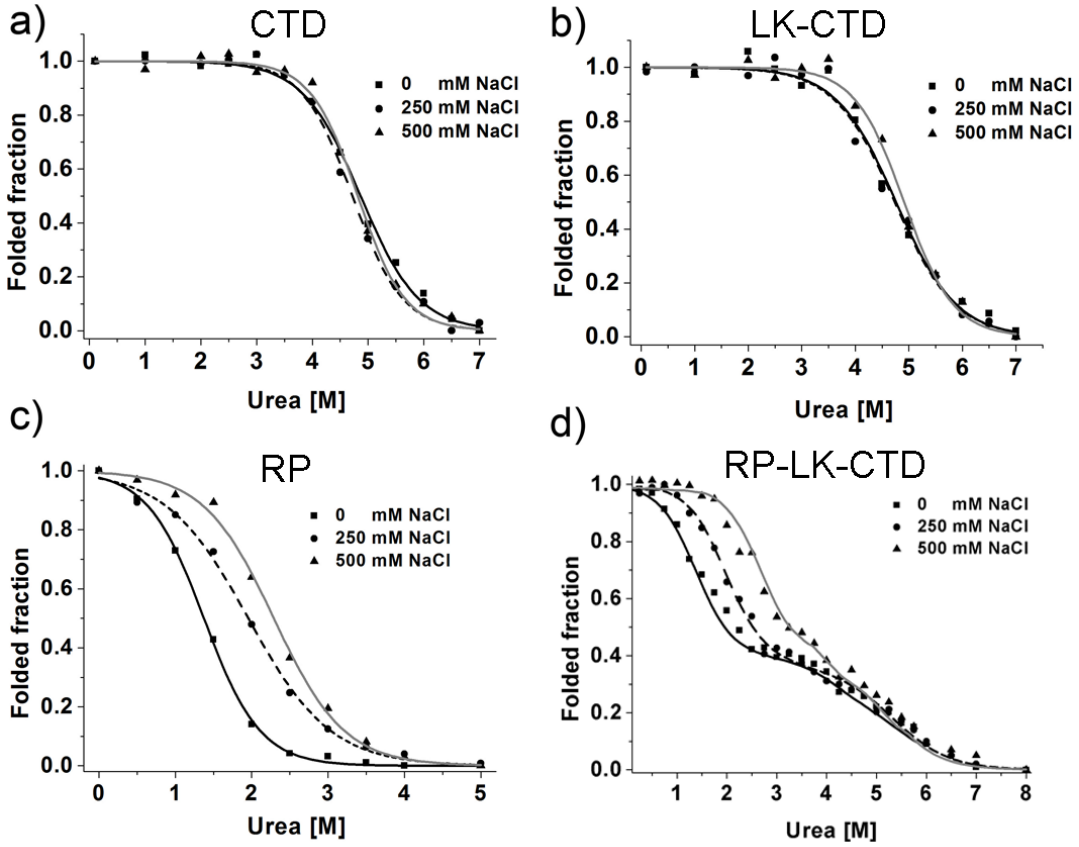


Figure 3.27 Urea-induced denaturation of MiSp fragments under different NaCl concentrations monitored by CD. (a) CTD_{Mi}; (b) LK-CTD_{Mi}; (c) RP_{Mi}; (d) RP-LK-CTD_{Mi}. The lines represent the fitting curves. All the data were normalized.

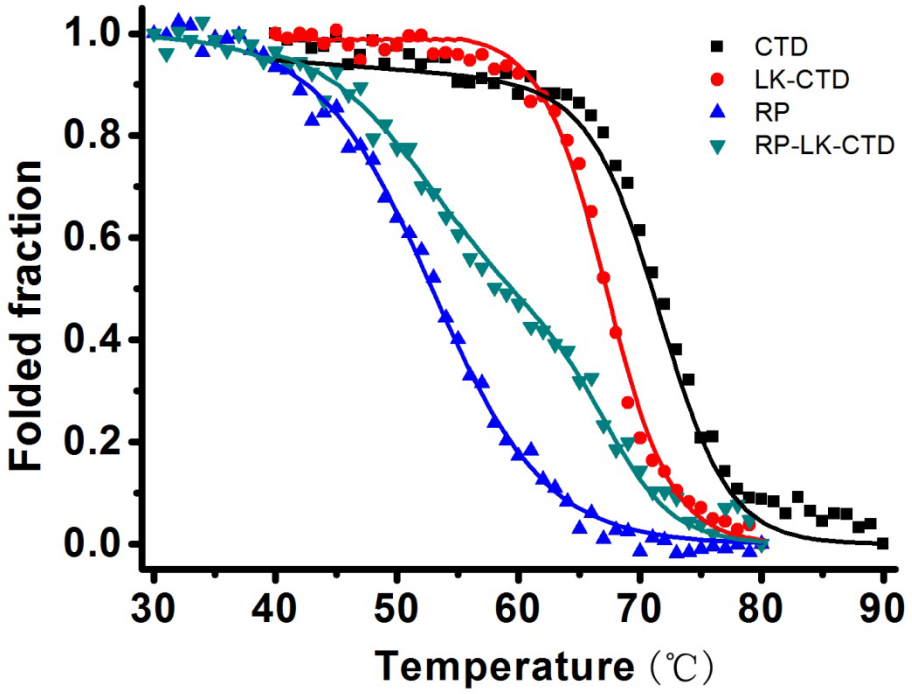


Figure 3.28 Temperature-induced unfolding of different MiSp fragments monitored by CD. The curves are fitted.

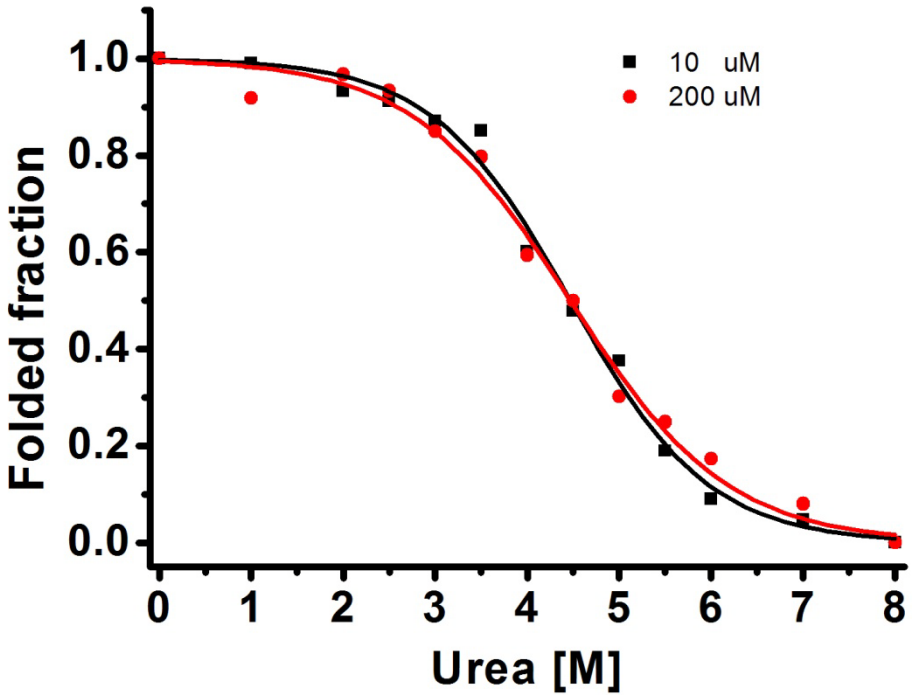


Figure 3.29 Urea-induced denaturation of CTD_{Mi} with different concentrations monitored by CD. The curves are fitted.

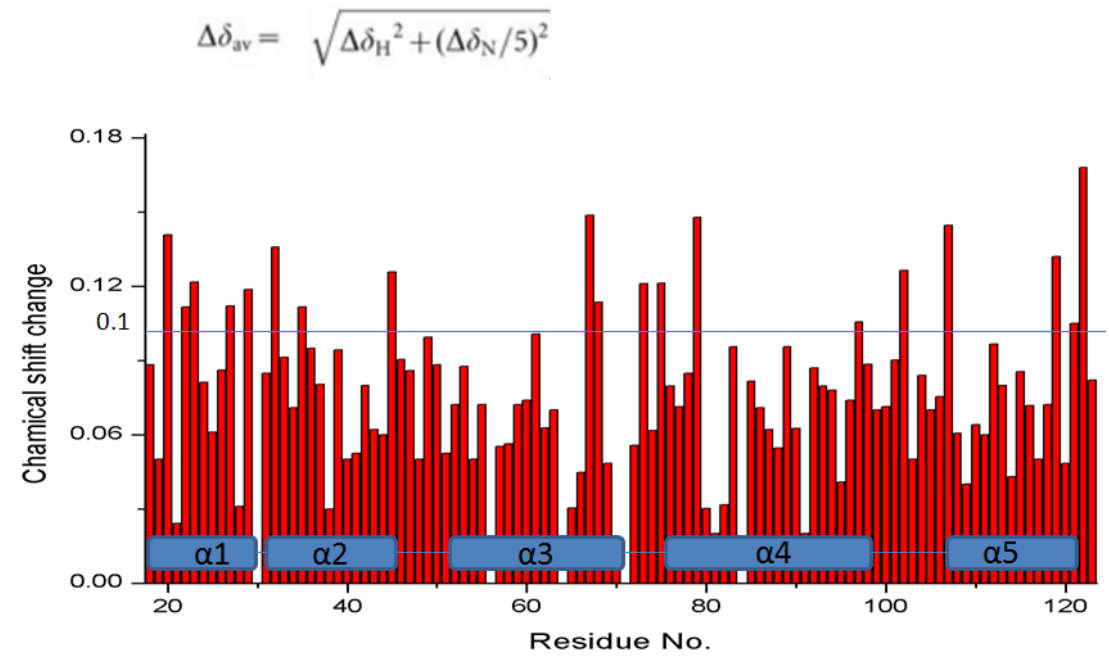


Figure 3.30 Chemical shift change of each residue of CTD_{Mi} in the presence of 3.5 M urea. The equation used to calculate the average of chemical shift change is presented on the top.

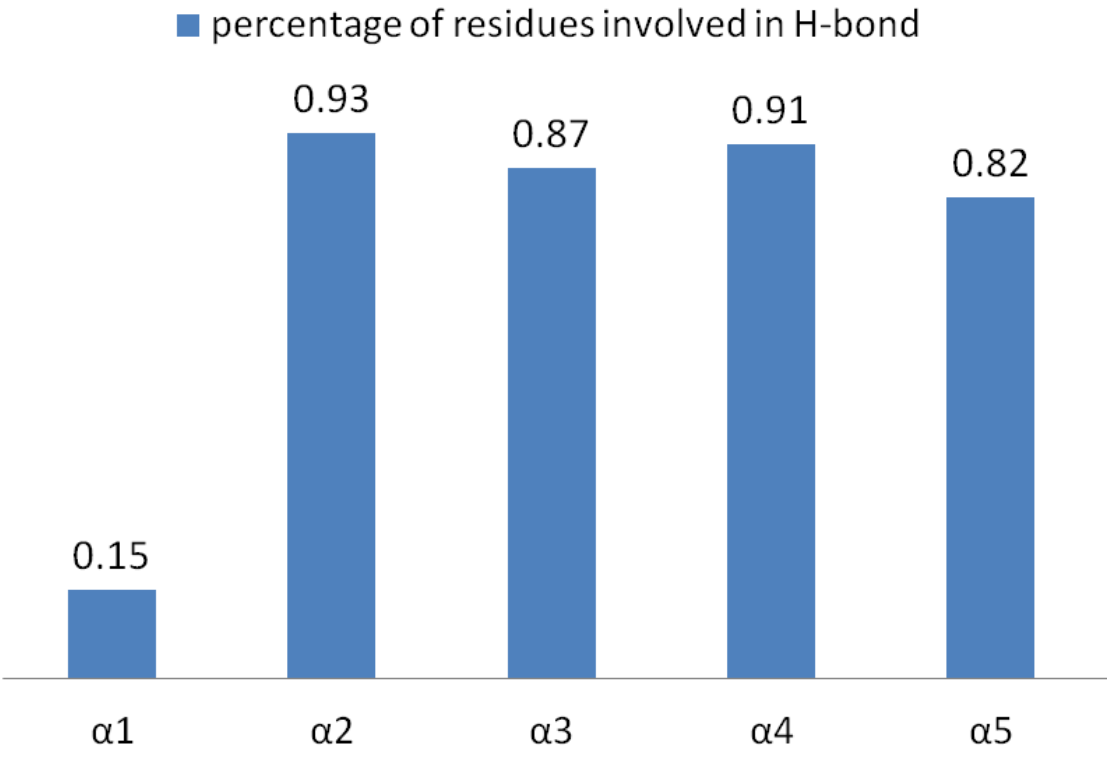


Figure 3.31 Percentages of residues involved in hydrogen bonding for each helix of CTD_{Mi}.

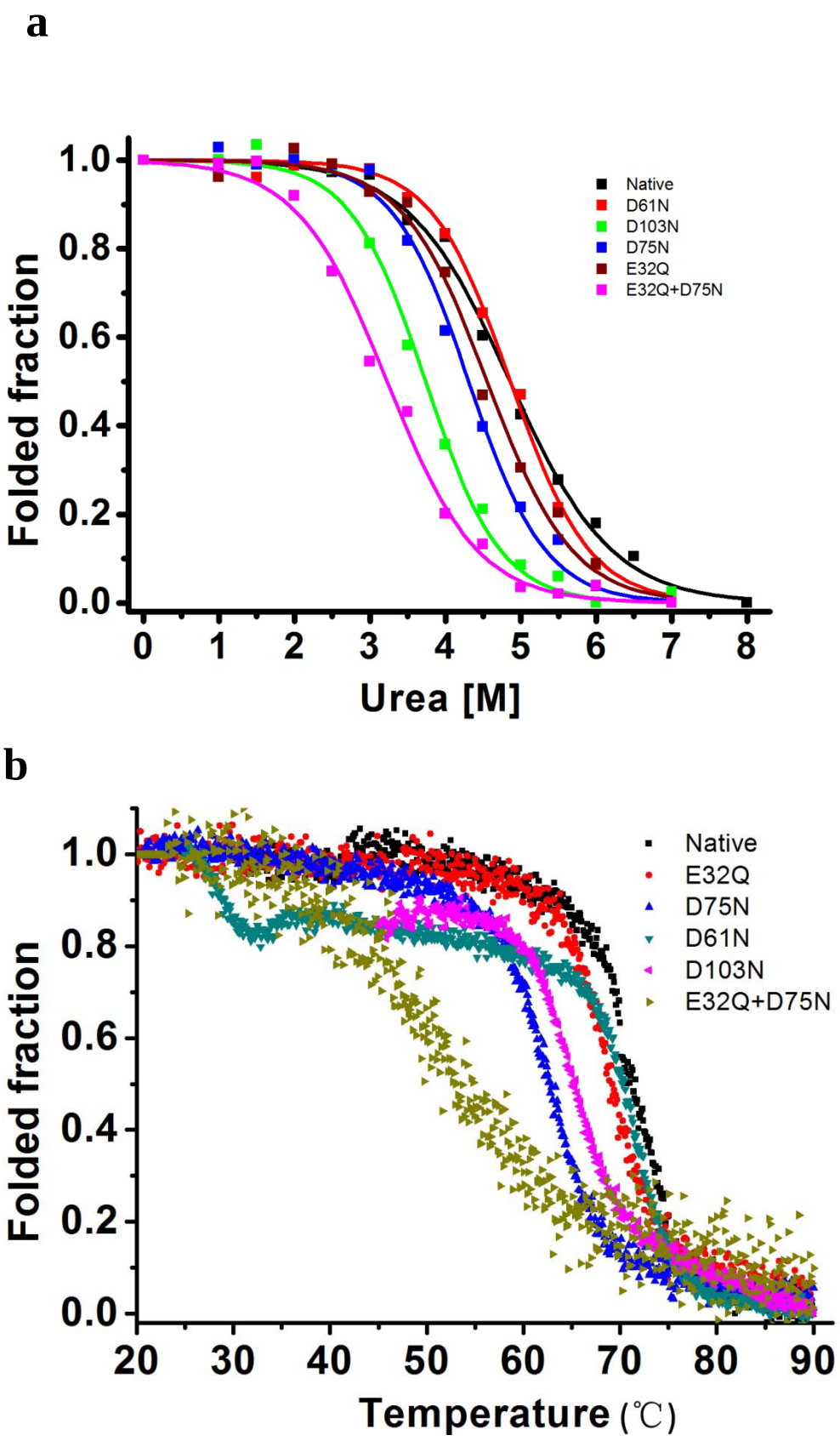


Figure 3.32 Denaturations of CTD_{Mi} mutants monitored by CD. (a) Urea-induced denaturation. The solid lines represent the fitting curves. (b) Thermal-induced denaturation.

3.5.2 Stability of LK-CTD_{Mi}

The chemical stability of LK-CTD_{Mi} and CTD_{Mi} was nearly identical (Figure 3.27a, b). Similar to CTD_{Mi}, LK-CTD_{Mi} was not influenced in stability by salt when NaCl concentration was below 500 mM. The result shows that LK_{Mi} has no obvious effects on CTD_{Mi}'s stability, implying that LK_{Mi} does not interact with CTD_{Mi} and confirming the conclusion drawn from the comparison of 2D HSQC spectra (Figure 3.18). The T_m of LK-CTD_{Mi} was about 4 °C lower than that of CTD_{Mi} (Figure 3.28). This should not have resulted from the interaction of LK and CTD, but can be caused by the gradual slight aggregation of LK-CTD_{Mi} during the temperature ramping process. It is noteworthy that a small of amount of precipitate was observed only for LK-CTD_{Mi} and RP-LK-CTD_{Mi}.

3.5.3 Stability of RP_{Mi}

RP_{Mi} also displayed a typical two-state unfolding profile, but showed a much lower stability than CTD_{Mi} (Figure 3.27c). Its C_m went up from 1.4 M to 2.3 M urea when NaCl concentration was increased from 0 to 500 mM. If the presence of high NaCl concentration can also be found during storage of MiSp in minor ampullate gland like in major ampullate gland(Hagn, Eisoldt et al. 2010), the high salt concentration could prevent the unfolding of RP_{Mi} which might initiate the undesired aggregation. This salt-dependent stability is similar to CTD_{Ma} but different from CTD_{Mi}. Thermal denaturation (Figure 3.28) also shows RP_{Mi} is much less stable than CTD_{Mi} (T_m: 53 vs 71 °C). In addition, RP_{Ac} possesses much higher thermal stability (82 °C)(Wang, Huang et al. 2012) than RP_{Mi}. The low chemical and thermal stabilities of RP_{Mi} are most likely due to the relatively weak hydrophobic interaction among helices.

3.5.4 Stability of RP-LK-CTD_{Mi}

Unlike individual domains, the tri-domain fragment, RP-LK-CTD_{Mi}, unfolded in two steps with the increase of urea concentrations. The denaturation curves were

fitted (Figure 3.27d), and the extracted first and second C_m values were the same as the C_m values of the isolated RP_{Mi} and CTD_{Mi}, respectively. This suggests RP_{Mi} tends to unfold first, followed by the unfolding of CTD_{Mi}. The result also shows that the three domains of RP-LK-CTD_{Mi} have no or very weak interactions in solution at low protein concentrations (< 10 μM). This conclusion is consistent with that drawn from the NMR data analysis, and is further supported by the fact that only the first-step unfolding of RP-LK-CTD_{Mi} is obviously dependent on salt (Figure 3.27d). The thermal denaturation data (Figure 3.28) also indicates RP-LK-CTD_{Mi} unfolds in two steps in a non-cooperative way and the CTD in the tri-domain protein cannot stabilize the RP.

3.6 Solubility of CTD_{Mi}, RP_{Mi}, LK-CTD_{Mi}, RP-LK_{Mi} and RP-LK-CTD_{Mi}

Since the extremely high water solubility of spider silk protein during storage is one of the most fantastic features of spider silk protein, the solubility of different fragments of MiSp were examined.

In 10 mM Tris buffer (pH 7.0), CTD_{Mi}, LK-CTD_{Mi}, RP-LK-CTD_{Mi}, RP_{Mi} and RP-LK_{Mi} could be concentrated to about 300, 200, 150, 60 and 5 mg/ml before the observation of precipitate or gel. In 10 mM sodium phosphate (pH 7.0), the solubility of each protein was nearly the same as that in 10 mM Tris, indicating that the solubility is not affected by buffer. RP-LK_{Mi} had the lowest solubility and was prone to precipitate. Other domains or fragments did not precipitate during the concentration process, but they formed gel when their concentrations were above their corresponding maxima. As shown in Figure 3.9b, CTD_{Mi} is purely negatively charged and very polar on its surface. The electrostatic repulsion among negatively charged dimeric CTD_{Mi} can prevent the self-assembly for the formation of random aggregates. Therefore, the high hydrophilicity and unique charge of CTD_{Mi} surface explains its extremely high solubility.

Although RP_{Mi} was easy to form small oligomers, its water solubility was still quite high. This is achieved by burying the solvent-exposed hydrophobic patch

(Figure 3.21b) in oligomeric structures. However, these oligomeric structures cannot maintain the relatively high water solubility of RP_{Mi} for long time because the concentrated RP_{Mi} (>10 mg/ml) formed gel after left at room temperature for more than one week. Given that the stability of RP_{Mi} is poor, this gel formation may be caused by the partial unfolding of RP_{Mi} which in turn leads to expose more hydrophobic surface. RP_{Mi} is the only construct which was found to have this property of time-dependent gel formation. LK-RP_{Mi} (5 mg/ml) was much less soluble than RP_{Mi} (60 mg/ml), demonstrating that the solubility of LK_{Mi} is significantly lower than 5 mg/ml. LK_{Mi}'s low solubility agrees with its high hydrophobicity (Figure 3.15b). Interestingly, the protein fragments containing the highly water soluble CTD_{Mi} (LK-CTD_{Mi} and RP-LK-CTD_{Mi}) were still very soluble. The solubility of the poorly soluble domain (LK_{Mi}) or fragment (RP-LK_{Mi}) can be enhanced by the highly soluble domain through mutual compensation in solubility. Alternatively, the poorly soluble domain or fragment may assemble to form oligomers through the aggregation-prone regions in LK_{Mi} or/and RP_{Mi}, leading to partial burial of solvent-exposed hydrophobic regions and in turn to greatly increase the solubility of the poorly soluble domain or fragment. The presence of such oligomers in the sample of 3 mM RP-LK-CTD_{Mi} is evidenced by the observation of the significant increase of the line width of methyl proton signals from the RP domain rather than from the CTD domain (Figure 3.26). In addition, the DLS result (Figure 3.33) of 0.7 mM RP-LK-CTD_{Mi} further supports the existence of the oligomeric structures. Similar to RP-LK-CTD_{Mi}, the full length MiSp may also exist in oligomers in the silk gland where the protein concentrations can reach up to ~50% w/w (Hijirida, Do et al. 1996). To understand why silk fibroins are highly soluble in silk glands, studies on the full length fibroins or large fragments including N- and C-terminal domains and several repetitive units are necessary.

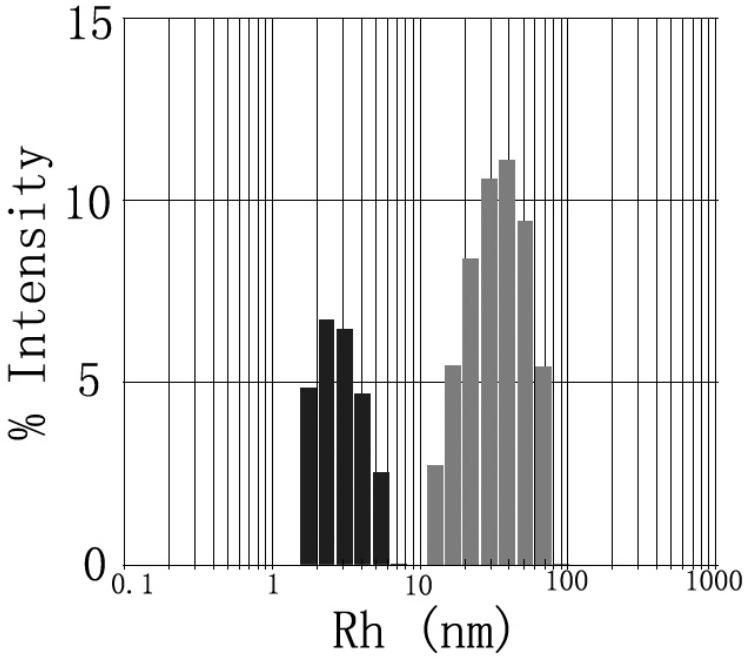


Figure 3.33 DLS result of RP-LK-CTD$_{Mi}$. The sample concentration was 20 mg/ml. The average size of black bars is corresponding to dimer, while the gray bars are corresponding to oligormers bigger than dimer.

3.7 Stability against Shear Force

The shear force is known to play important role at the last stage of fiber spinning as mentioned in chapter 1. Therefore, the stabilities of different fragments of MiSp against shear force were investigated. The maltose binding protein (MBP), whose stability is well known, was employed as a non-silk protein control.

In the absence of stirring (mechanical shear force), RP$_{Mi}$, RP-LK$_{Mi}$, RP-LK-CTD$_{Mi}$, LK-CTD$_{Mi}$, CTD$_{Mi}$ and MBP could maintain a soluble state under the condition of 0.05 mg/ml protein, 10 mM phosphate, pH 6.8 and 25 °C without detectable precipitate within two days. In the presence of stirring, however, all of them tended to aggregate to form visible precipitate that is detectable at 350 nm. The changes in the amount of aggregated proteins with time are shown in Figure 3.34. The C_m and T_m values of MBP in the absence of its ligand are 3.3 M urea and 63 °C respectively(Beena, Udgaonkar et al. 2004), indicating MBP is more stable than RP$_{Mi}$ but less stable than CTD$_{Mi}$. Thus the aggregation rates of RP$_{Mi}$, CTD$_{Mi}$ and MBP are inversely proportional to their thermal or chemical stability. This can also be

applied to CTD_{Mi} and its mutants (Figure 3.35). The mutants with lower stabilities gave more aggregation under shaken, especially E32Q+D75N. The aggregation should occur through partial protein unfolding and then assembly of the partially unfolded molecules. Note that the partially unfolded proteins have more solvent-exposed hydrophobic residues than the folded ones. Therefore the more stable a protein is, the slower the protein unfolding is, and the slower the shear-force-induced aggregation is.

The aggregation rates of CTD_{Mi} and RP_{Mi} were greatly accelerated by adding LK_{Mi} domain to them (Figure 3.34). This result can be explained by the high aggregation propensity and low water solubility of LK_{Mi} domain (Figure 3.15b and Figure 3.16b). Although LK_{Mi} is intrinsically disordered, stirring still greatly accelerated the aggregation of LK-CTD_{Mi} and RP-LK_{Mi}, implying that the aggregation-prone regions are partially protected in the bi-domain protein fragments and shear force can reduce the protection. The protection may be achieved by the partial local folding of the aggregation regions of LK_{Mi} or by the weak interaction between the disordered domain and folded domain. In any cases, LK_{Mi} plays a predominant role in the aggregation process of the protein fragments with an LK_{Mi} domain. Although RP-LK_{Mi}, LK-CTD_{Mi} and RP-LK-CTD_{Mi} all contain an LK_{Mi} domain, RP-LK_{Mi} that is lacking a CTD_{Mi} displayed the highest aggregation rate. This result shows that CTD_{Mi} can slow down the aggregation rate and may play a role in regulating the assembly of silk protein molecules to form ordered structures. Since the RP domain is prone to form oligomers and is unstable against shear force, chemical and thermal denaturation, it may assist the LK domain to assemble silk protein molecules.

In the case of MaSp, ion strength change during fiber spinning was proven to have effect on fiber assembly. To test if this ion strength change has similar effect on the fiber assembly of MiSp, the effects of two different types of salt on the shear-induced aggregation of RP-LK-CTD_{Mi} were tested. Both NaCl and Na_3PO_4 were able to enhance the aggregation of RP-LK-CTD_{Mi} in the presence of shear force in similar rates (Figure 3.36). The effect of NaCl on the aggregation of CTD_{Ma} and

RP-CTD$_{Ma}$ was much less pronounced(Hagn, Eisoldt et al. 2010). This result suggests that the fibroin storage and/or assembly conditions in MiSp and MaSp spider glands may be different.

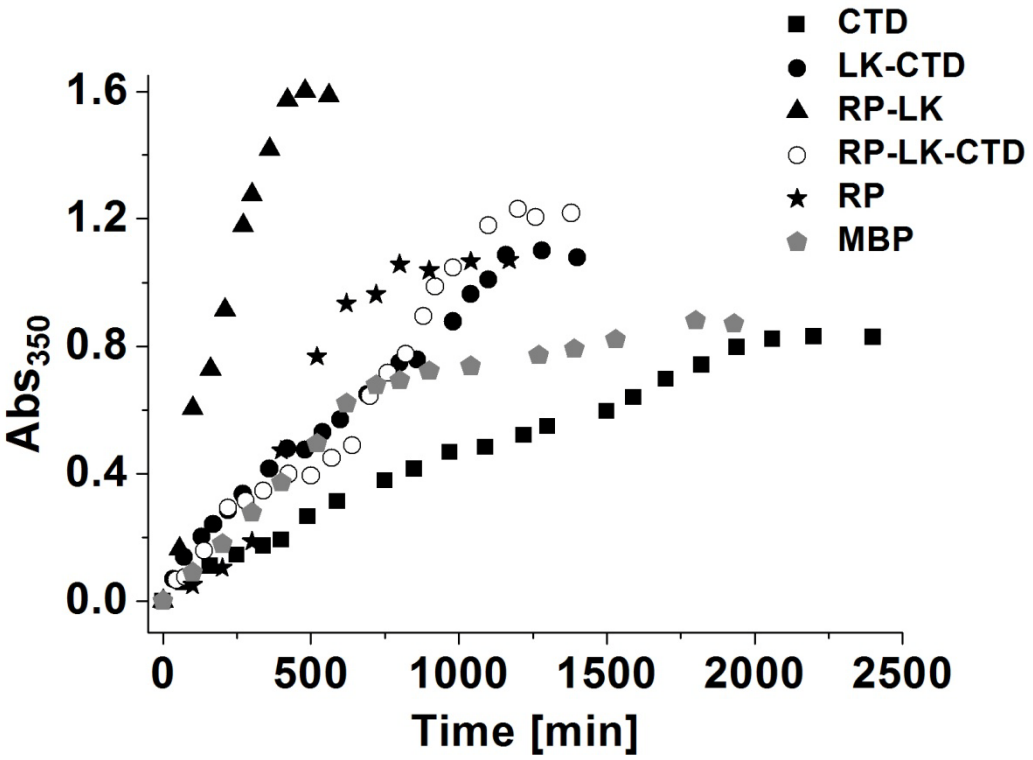


Figure 3.34 Shear-force-induced aggregation of MiSp fragments and MBP monitored by OD$_{350}$.

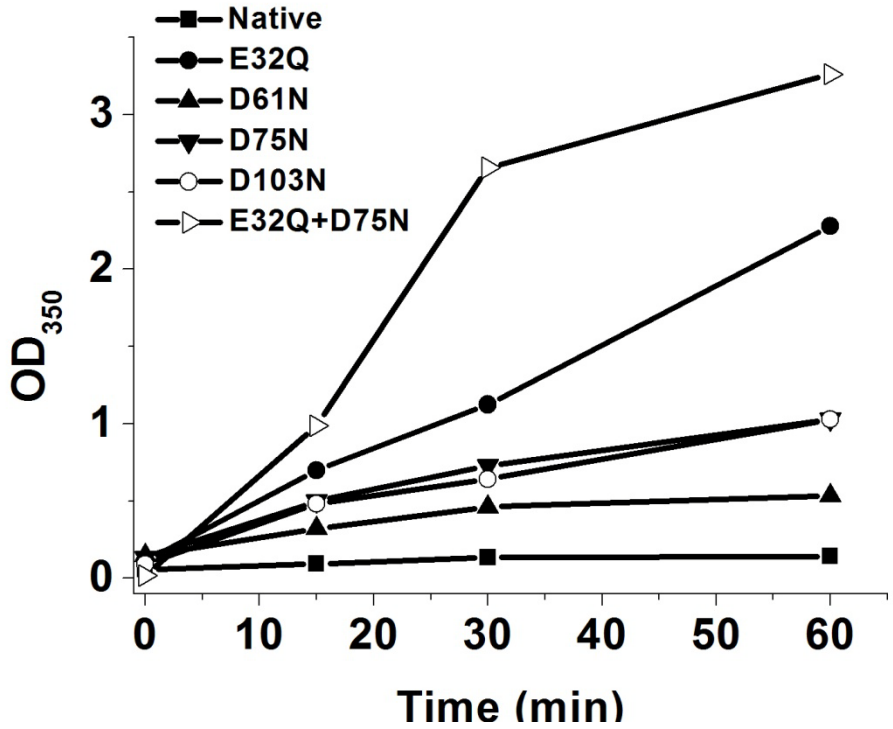


Figure 3.35 Shear-force-induced aggregation of CTD_{Mi} and its mutants monitored by OD_{350}.

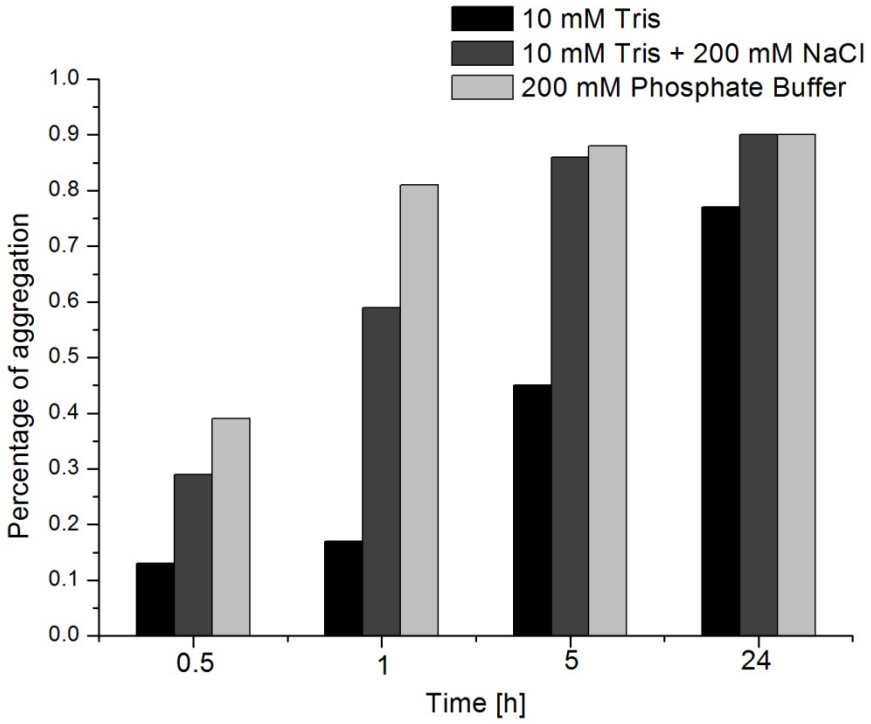


Figure 3.36 Effects of salts on the shear-force-induced aggregation of RP-LK-CTD_{Mi}. The pHs of the buffers were 7.

3.8 Fiber formation

All the single and bi-domain constructs always underwent nonspecific aggregation or precipitation in water upon gentle shaking. Under the same condition, however, RP-LK-CTD_{Mi} could form small fibers with well-aligned structure and smooth surface even at a low protein concentration of ~0.3 mg/ml (Figure 37). The diameters of the formed fibers ranged from ~2-10 μm, similar to that of the native MiSp silk (La Mattina, Reza et al. 2008). Our result reveals that all the three domains should participate in the fine-tuned process of fiber formation. LK_{Mi} domain may act as a nucleation site to initiate the assembly of RP-LK-CTD_{Mi} molecules. RP_{Mi} may assist LK_{Mi} to assemble RP-LK-CTD_{Mi} molecules together. The folded CTD_{Mi} domains may play an important role in regulating the alignment of the assembled molecules, which leads to controlled formation of well-defined fibers. Mechanical force can enhance the initial assembly and facilitate the alignment.

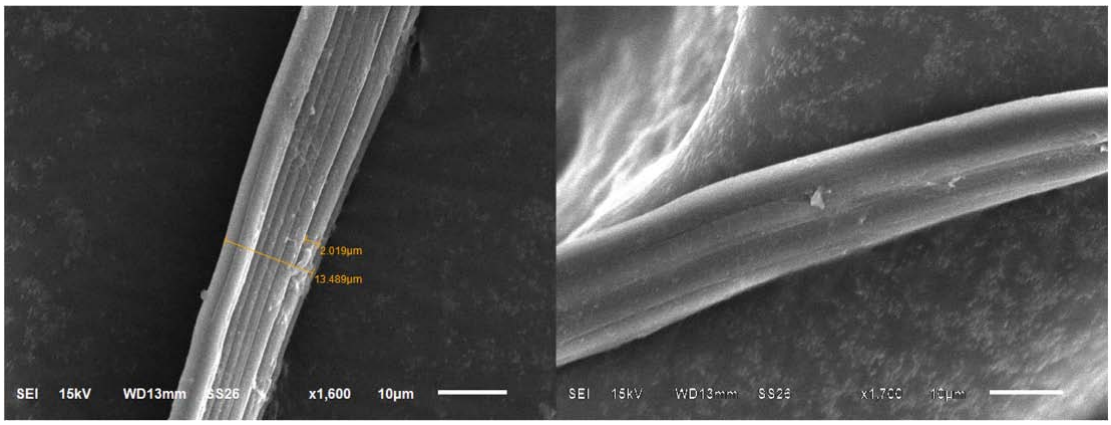


Figure 3.37 SEM pictures of silk fibers formed by RP-LK-CTD_{Mi}.

3.9 NTD_{Mi}

The sequence of NTD_{Mi} from *N. antipodiana* was identified from genomic DNA by Lai Chongcheong at the relatively late stage of my PhD period. Therefore, the work on NTD_{Mi} was not as thorough as that on the other domains. The sequence was firstly subcloned into pET-M vector and transformed into *BL21(DE3)*. The protein could be easily expressed and purified from supernatant of the cell lysate.

3.9.1 Backbone assignment of NTD_{Mi}

The sequence identities shared by NTD_{Mi} and the NTDs of MaSp are quite high (Figure 3.1b). Since the known structures of NTD_{Ma} are almost identical(Askarieh, Hedhammar et al. 2010; Hagn, Thamm et al. 2011; Jaudzems, Askarieh et al. 2012), the overall structure of NTD_{Mi} is most likely similar to those of NTD_{Ma}s. Therefore, determining the 3D structure of NTD_{Mi} is not the priority. However, the backbone assignment of NTD_{Mi} is necessary for other studies by NMR.

The ^{15}N, ^{13}C labeled sample of NTD_{Mi} was prepared. Since the NTD_{Mi} has salt dependent oligomerizition-prone characteristic which will be discussed, 300 mM NaCl was added to the sample which contained 0.5 mM protein, 10 mM phosphate buffer at pH 7.0 to improve the quality of NMR spectra. The backbone of 97.6% residues (except the region corresponding to His-tag and thrombin cleavage site, Met^{1}-Ser^{16}) were unambiguously assigned (Figure 3.38) by analyzing ^{1}H-^{15}N HSQC, 3D HNCACB and 3D CBCA(CO)NH.

The CSI of NTD_{Mi} was calculated with the assigned chemical shift values (Figure 3.39). The result reveals that NTD_{Mi} is α-helical structure and contains five helices. The boundaries of helices of NTD_{Mi} are nearly the same as the known NTD_{Ma} structures, further supporting that the overall structure of NTD_{Mi} is similar to the solved structures of NTD_{Ma}.

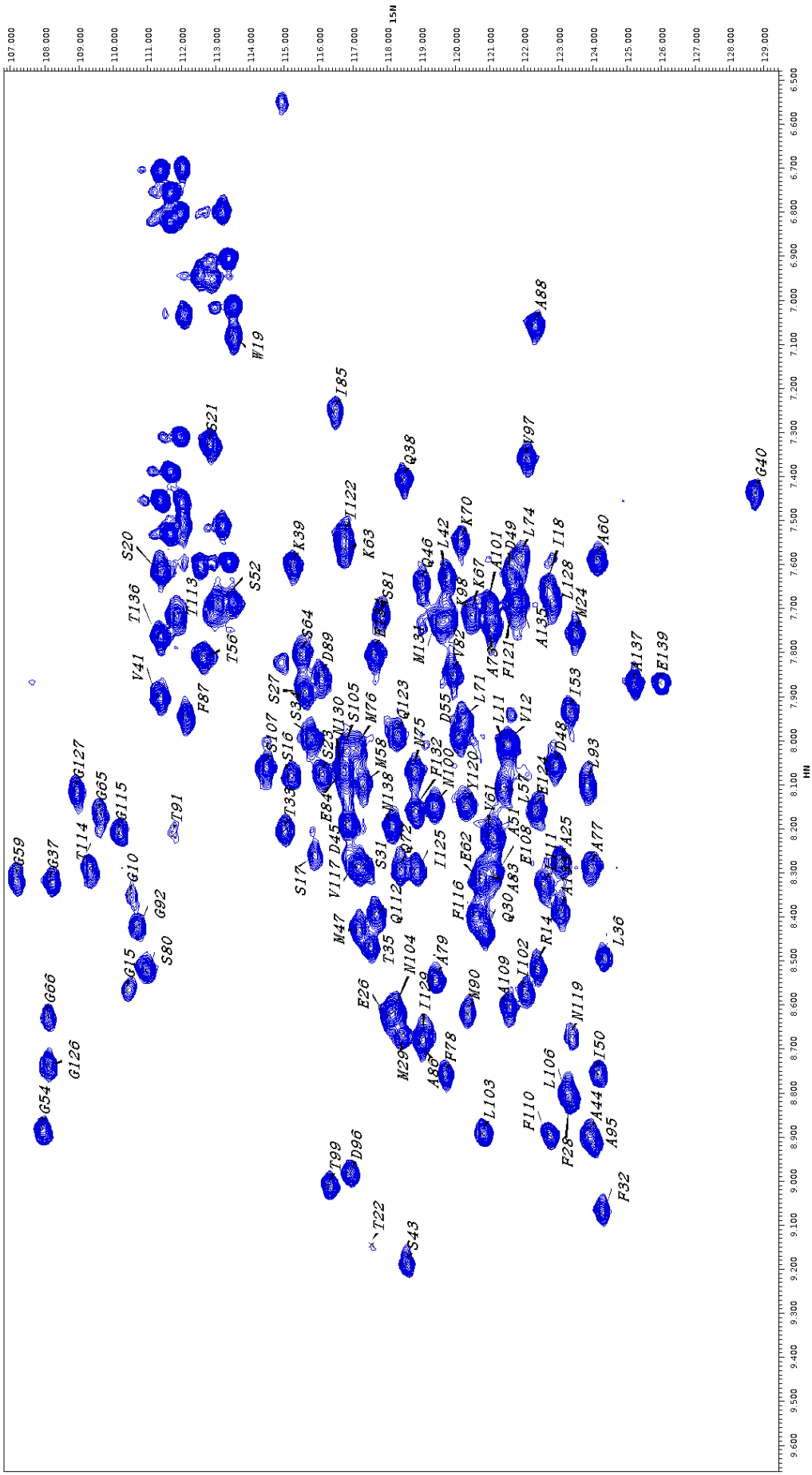


Figure 3.38 Assigned ^{1}H, ^{15}N HSQC spectrum of NTD_{Mi}.

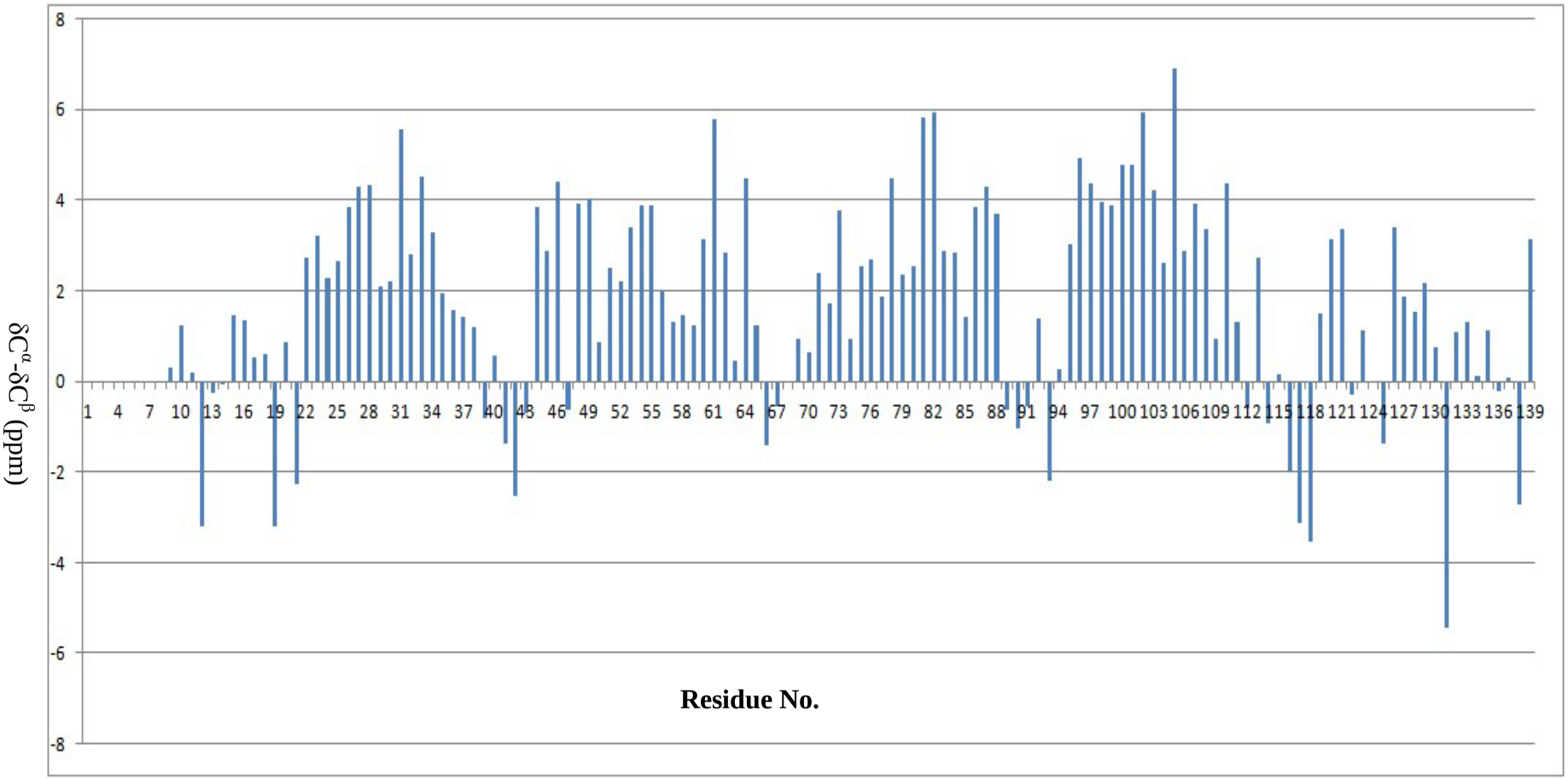


Figure 3.39 Chemical shift index of NTD $_{Mi}$. Residue 1-16 are the His-tag and thrombin cleavage site.

3.9.2 Salt-dependent oligomerization of NTD_{Mi}

Although NTD_{Mi} is highly soluble, it forms soluble oligomers in the absence of sodium chloride or at low concentration of sodium chloride as evidenced by the SEC results of NTD_{Mi} (Figure 3.40a). Intestinal fatty acid bind protein (iFABP, 15 kD), MBP (43 kD) and CTD_{Mi} (dimer, 25 kD) were used as internal reference to estimate the apparent molecular weight of NTD_{Mi}. Figure 3.40a shows that when the concentration of sodium chloride is reduced from 300 mM to 0 mM, the NTD_{Mi} peak gives rise to a significant left shift, indicating oligomerization of NTD_{Mi} happens after removing sodium chloride from the buffer. The 1D ^{1}H spectra of NTD_{Mi} (Figure 3.40b) also suggest that NTD_{Mi} undergoes a salt-dependent oligomerization. The isolated methyl signals at -0.72 ppm, -0.42 ppm and 0.3ppm show obvious decrease in line widths with the increase of sodium chloride concentration. Because the NMR line width is proportional to the molecular size but complicated by conformational exchanges, the oligomerization of NTD_{Mi} can be inhibited by high concentration of sodium chloride. Since NTD_{Mi} contains 14 negatively and 7 positively charged residues, its oligomerization can be affected by sodium chloride and the oligomerization of NTD_{Mi} is most likely mediated by charge-charge interaction.

The salt-induced conformational change of NTD_{Mi} was further monitored by ^{1}H-^{15}N HSQC (Figure 3.41). With the increase of sodium chloride in the NMR samples of NTD_{Mi}, many peaks in HSQC spectrum have significant chemical shift change or/and increase of peak intensity, implying the conformational change of NTD_{Mi} occurs during changing sodium chloride concentration. Given that the backbone assignment of NTD_{Mi} in the presence of 300 mM sodium chloride has been achieved, those peaks with significant chemical shift change or/and increase of peak intensity could be identified in the protein sequence. This information is useful for identifying the interface(s) of oligomerization of NTD_{Mi} after solving its 3D structure.

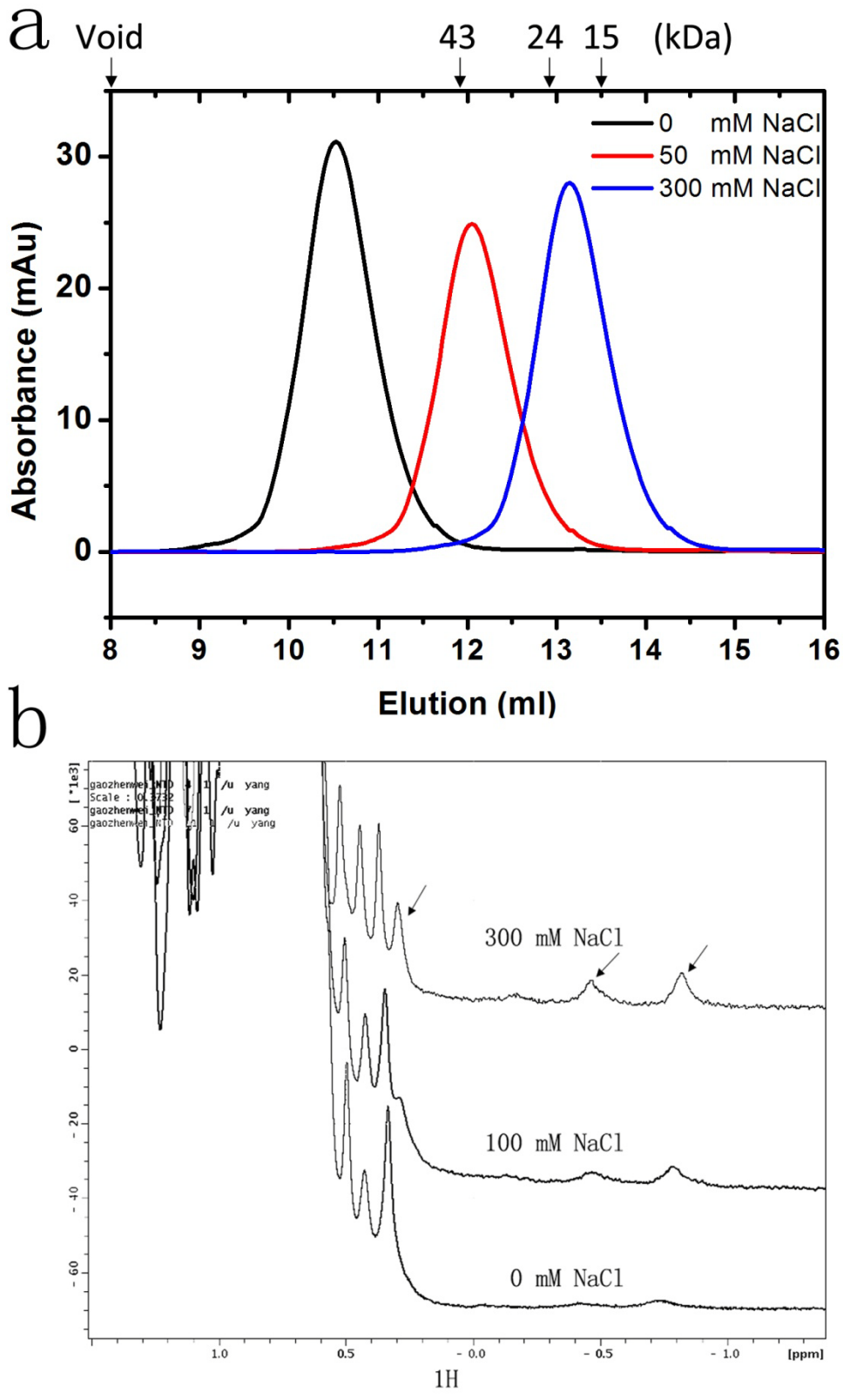


Figure 3.40 Salt-dependent oligomerization of NTD_{Mi}. (a) SEC results of NTDMi at different sodium chloride concentrations. The results of samples at different salt concentrations are indicated by different colors. (b) Stacked plot of 1D spectra of NTDMi at different sodium concentrations. The isolated methyl signals are indicated by arrows.

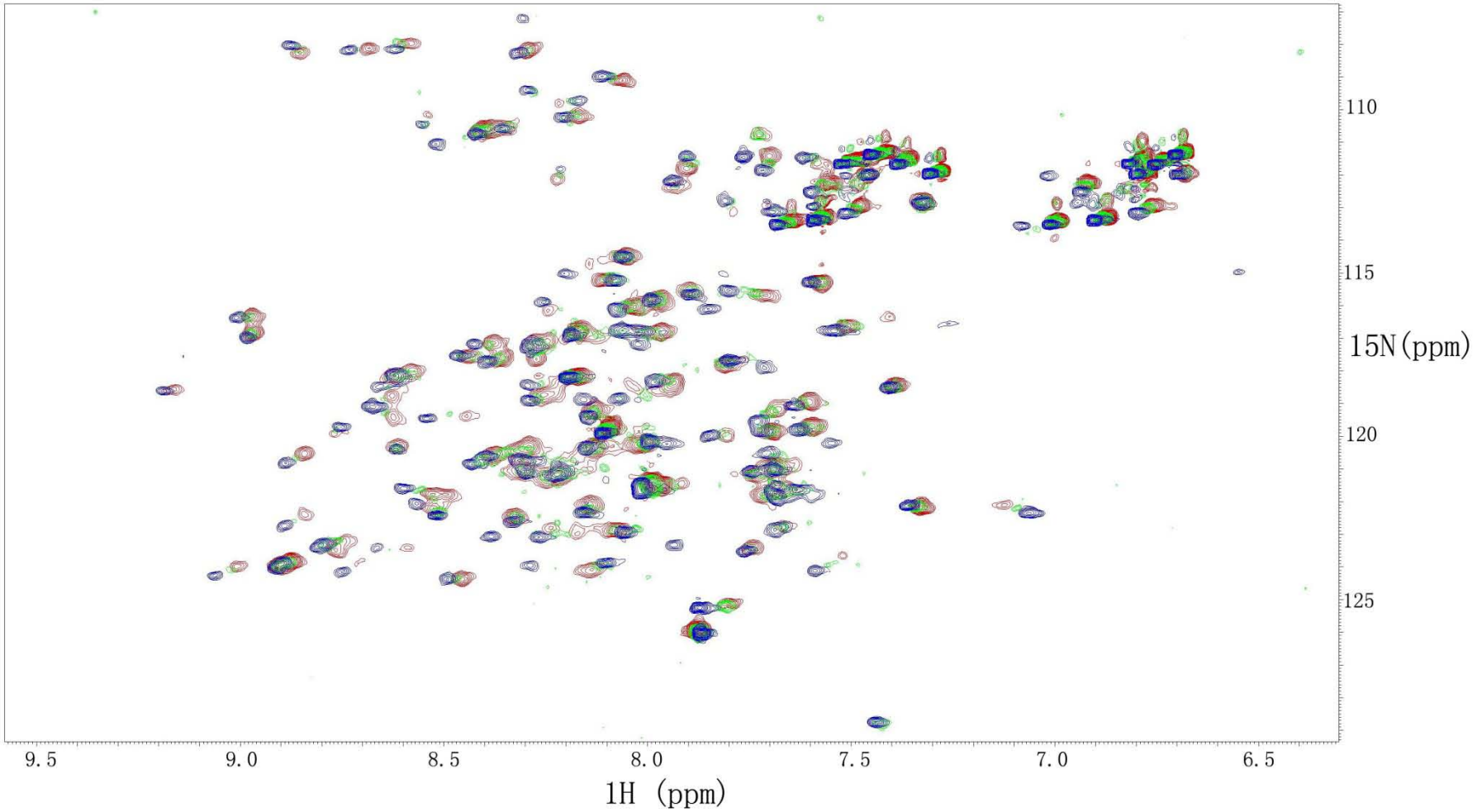


Figure 3.41 Overlay of ^{1}H, ^{15}N HSQC spectra of NTD_{Mi} at different sodium chloride concentrations: 0 mM (red), 100 mM (green), 300 mM (blue).

3.9.3 NTD-LK-RP-LK-CTD$_{Mi}$

To understand the functional role of NTD$_{Mi}$ in the native silk protein, the construct NTD-LK-RP-LK-CTD$_{Mi}$ (52 kD) which includes all the four different domains of the native MiSp was created. This protein could be easily expressed by *E. coli*. The concentration of NTD-LK-RP-LK-CTD$_{Mi}$ could reach >200 mg/ml in 10 mM Tris buffer with 100 mM NaCl at pH 7.0. No precipitation of this protein was noticed during the concentrating process. When the concentration of the protein was higher than 200 mg/ml, the protein sample became sticky gel-like solution which could not be further concentrated. Given that LK domain is hydrophobic and has poor solubility, the NTD clearly has the effect to improve the solubility of the whole protein because adding NTD-LK to RP-LK-CTD could slightly increase the protein solubility from 150 mg/ml to >200 mg/ml. Since the oligomerization of NTD$_{Mi}$ is most likely mediated by charge-charge interaction, most of the charged residues of NTD$_{Mi}$ should be exposed on the surface which would make its surface very hydrophilic. This might explain the effect of NTD$_{Mi}$ on improving the solubility of the whole protein.

Similar to RP-LK-CTD$_{Mi}$, NTD-LK-RP-LK-CTD$_{Mi}$ also forms soluble aggregates when concentration is high (Figure 3.42d). In addition, the aggregation feature of NTD-LK-RP-LK-CTD$_{Mi}$ is concentration dependent according to the DLS results (Figure 3.42). As shown in Figure 3.42, the DLS intensity of molecules with large diameters (gray columns) went up from 6.7% to 80.5%, when the protein concentration was increased from 1 mg/ml to 20 mg/ml. This concentration dependent aggregation feature of NTD-LK-RP-LK-CTD$_{Mi}$ was also noticed during purification by a Superdex^{TM} 200 size exclusion column. The running buffer used in these SEC experiments was 10 mM Tris with 100 mM NaCl at pH 7.0. When the protein sample with relative lower concentration (2mg/ml) was loaded to the column, the elution peak of target protein showed at 73 ml which corresponding to ~100 kD. In this case, the protein mainly exists as dimers because the CTD forms stable dimer. If the concentration of the loading sample was increased to 10 mg/ml, the elution

peak would shift to 57 ml and became broader, indicating non-uniform large oligomers were formed. The formation of oligomers could assist the protein to maintain its solubility in aqueous solution by burying the hydrophobic parts (LK and RP) and exposing the hydrophilic parts (CTD and most likely NTD), especially when concentration is high.

NTD-LK-RP-LK-CTD_{Mi} is also sensitive to shear force. Well defined silk-like fibers were quickly formed when 4 mg/ml protein was gently shook (Figure 3.43). Since RP-LK-CTD_{Mi} already has the ability to form well defined fibers under shear force, the functional roles of NTD in this shear-induced fiber formation need to be further investigated.

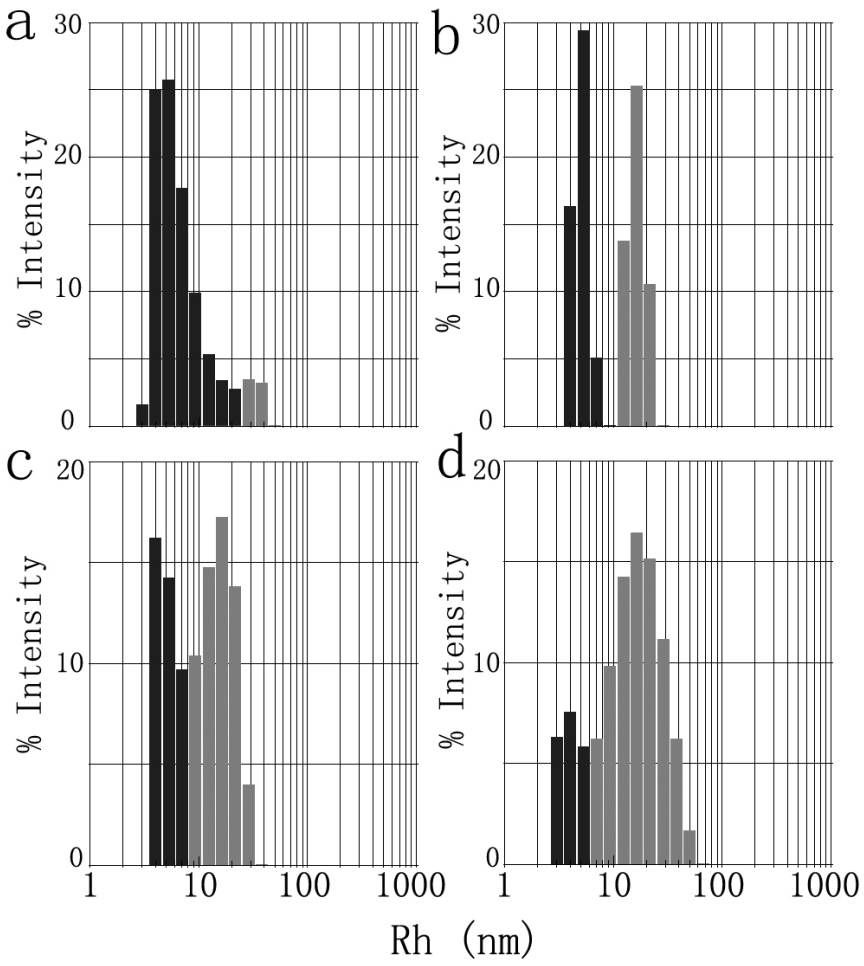


Figure 3.42 DLS results of NTD-LK-RP-LK-CTD$_{Mi}$ at different concentrations: 1 mg/ml (a), 5 mg/ml (b), 10 mg/ml (c) and 20 mg/ml (d). All the protein samples were prepared in 10 mM Tris, 100 mM NaCl at pH 7.0. The average size of black bars is corresponding to dimer, while the gray bars are corresponding to oligormers bigger than dimer.

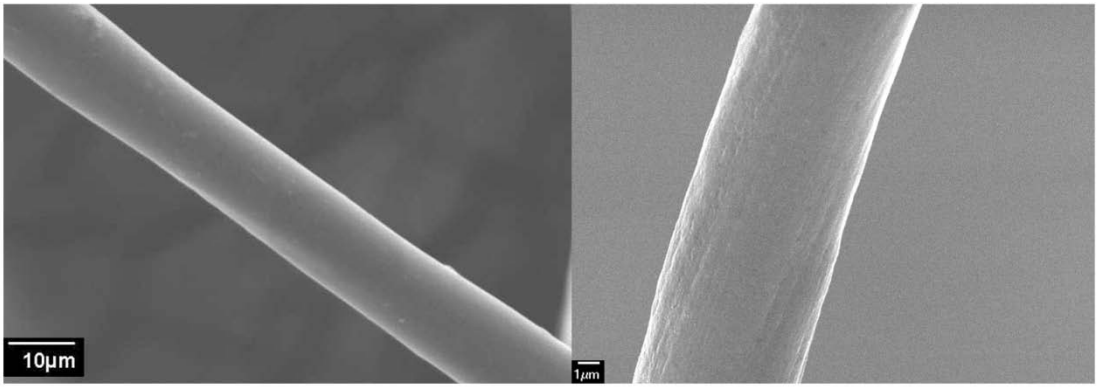


Figure 3.43 SEM pictures of fibers formed by NTD-LK-RP-LK-CTD$_{Mi}$.

4 Chapter 4: Conclusion

This study explored the solution structures of isolated MiSp domains of *N. antipodiana*. The structure of CTD_{Mi} adopts a globular fold of two twisted five-helix bundles which pack in parallel to form a homodimer. Although CTD_{Ma} and CTD_{Mi} share 44% sequence identity and high similarity in 3D structure, CTD_{Mi} is more stable than CTD_{Ma}. The stability difference can be explained by the structural differences mainly in surface charge and hydrophobic interactions. Since only CTD_{Ma} contains an inter-molecular disulfide linkage, stabilization of the dimeric structure of relatively conserved spider CTDs should be achieved through hydrophobic interactions. LK_{Mi} is intrinsically disordered and has aggregation-prone feature. In aquous solution, the structure of RP_{Mi} also adopts a globluar fold which is composed of a seven-helix bundle similar to those of $RP1_{Tu}$, $RP2_{Tu}$ and RP_{Ac}. However, a detailed comparison of these four structures revealed that the orientation of helix 6 from RP_{Mi} is different from that of the other three, which results in the relative weak hydrophobic interactions between helix 6 and the rest of helices in RP_{Mi}. The low stability of RP_{Mi} results from the weak hydrophobic interaction between helixes. In addition, the surface exposed hydrophobic patch of RP_{Mi} could be the reason why its water solubility is lower than the other RPs. The backbone assignment of NTD_{Mi} was achieved and its secondary structure was predicted from the assigned chemical shift values. The structure of NTD_{Mi} contains five helices, similar to the reported structure of NTD_{Ma}s. In addition, NTD_{Mi} exhibits a salt-dependent conformational change. Taken together, all the structural studies on the MiSp domains are of considerable importance since this is the first time that structure characterization is performed on MiSp and they provide valuable information to explain their functional roles.

Another aim of this study was to reveal the mechanism of fiber formation and protein storage of MiSp. CTD_{Mi}, RP_{Mi} and LK_{Mi} have distinct stability and solubility, which can be explained by their distinct structural properties. Interestingly, the presence of sodium chloride can stabilize RP_{Mi} against chemical denaturation and inhibit the oligomerization of NTD_{Mi}, indicating sodium chloride has a positive effect

on maintaining the high solubility of silk protein during its storage. The solubility studies shows that both CTD and NTD can maintain the multi-domain protein fragment in a highly water-soluble state. In addition, the lower solubilities of RP_{Mi} and LK_{Mi} suggest that they initiate the aggregation process during fiber formation. The shear-force-aggregation study showed that shear force can greatly accelerate protein aggregation by partial unfolding. Moreover, in the presence of mechanical force, the aggregation rate of a folded protein is inversely proportional to its thermal or chemical stability; while the aggregation rate of a multi-domain protein containing both folded and disordered domains is determined mainly by the property of the disordered domain and the solubility of the entire protein. Although all MiSp domains investigated in this study could self-assemble in the presence of shear force, only the RP-LK-CTD_{Mi} and NTD-LK-RP-LK-CTD_{Mi} formed well defined silk-like fibers, indicating that all the domains play distinct roles in fiber formation.

In view of all the information obtained in this study, the functional roles of MiSp domains in protein storage and fiber formation could be proposed. During storage, the micelle formation could maintain the protein as a water-soluble state by exposing the hydrophilic parts (NTD and CTD) and burying the hydrophobic RP and LK. When the protein enters the spinning duct, sodium chloride is removed, which cause the oligomerization of NTD and unstabilization of RP. The oligomerization of NTD may assist the fusion of micelles by gathering different micelles and the unstabilization of RP could expose more hydrophobic surface to facilitate the aggregation process. In the presence of shear force, the protein undergoes a self-assembly process. With the assistance of RP domain, LK domain serves as a nucleation site to assemble different molecules together and CTD domains enable the arrangement of the assembled molecules in a highly ordered manner.

All the structural and functional studies have provided new insights on the roles played by different MiSp domains during protein storage and fiber formation. They also shed light on the fundamental knowledge of sequence-structure-function relationship. In addition, the study of MiSp may provide guidelines for the study of other types of spider silk proteins.

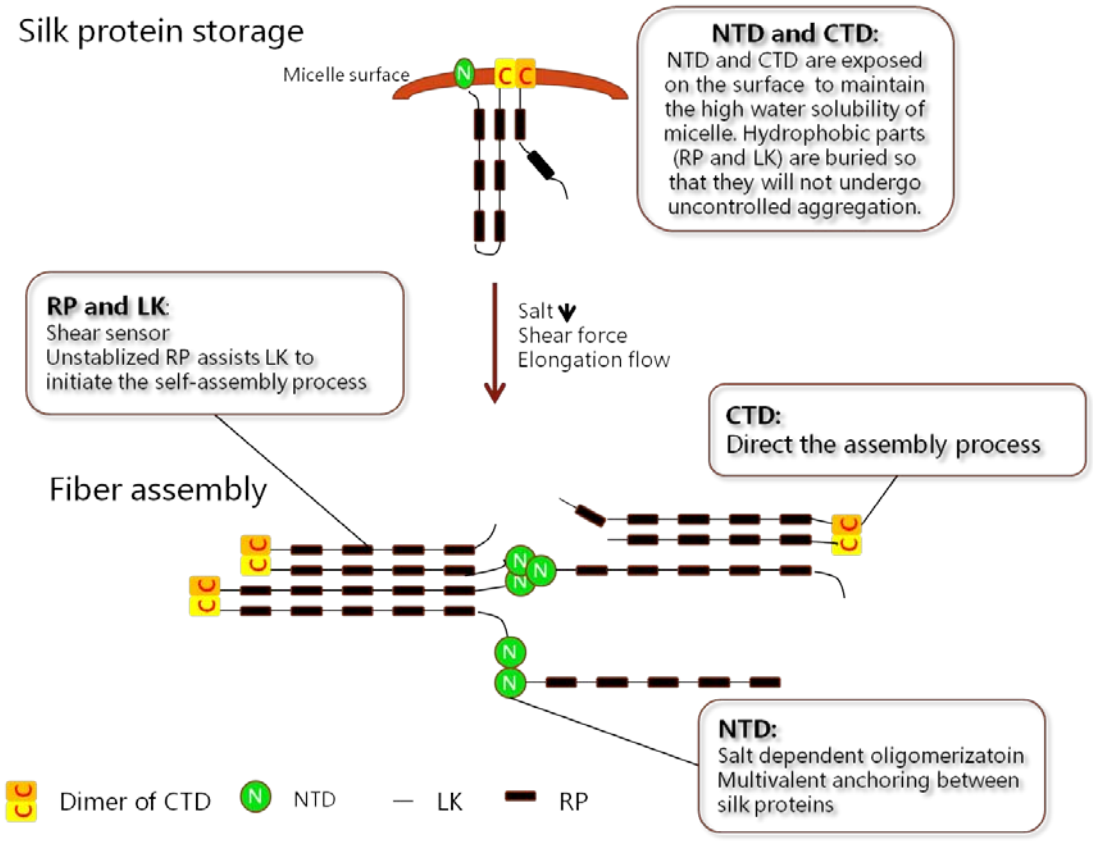


Figure 4.1 A proposed model of silk assembly.

Due to time limitation, the investigation of NTD_{Mi} was not deep enough. Therefore, NTD_{Mi} should be studied to a lager extent in the future. It will also be interesting to thoroughly study the effects of pH and salt concentration on both silk protein storage and fiber formation. In addition, how the length of silk protein and different combinations of domains contribute to the final mechanical properties of minor ampullate silk is worth detailed study. For instance, the constructs with two terminal domains and only numbers of LK or RP in between can be prepared. Thus, by comparing the mechanical properties of the fibers spanned from those constructs, the information about the sequence-mechanical properties relationship might be obtained. Moreover, besides silk fibers, silk protein can also be processed into different morphologies. Thus, our spider silk protein could be applied to some practical areas, such as drug delivery and tissue engineering.

5 Chapter 5: Reference

Askarieh, G., M. Hedhammar, et al. (2010). "Self-assembly of spider silk proteins is controlled by a pH-sensitive relay." Nature **465**(7295): 236-238.

Ayoub, N. A., J. E. Garb, et al. (2013). "Ancient properties of spider silks revealed by the complete gene sequence of the prey-wrapping silk protein (AcSp1)." Mol Biol Evol **30**(3): 589-601.

Ayoub, N. A., J. E. Garb, et al. (2007). "Blueprint for a high-performance biomaterial: full-length spider dragline silk genes." PLoS One **2**(6).

Beckwitt, R. and S. Arcidiacono (1994). "Sequence conservation in the C-terminal region of spider silk proteins (Spidroin) from Nephila clavipes (Tetragnathidae) and Araneus bicentenarius (Araneidae)." J Biol Chem **269**(9): 6661-6663.

Beena, K., J. B. Udgaonkar, et al. (2004). "Effect of signal peptide on the stability and folding kinetics of maltose binding protein." Biochemistry **43**(12): 3608-3619.

Bell, A. L. and D. B. Peakall (1969). "Changes in fine structure during silk protein production in the ampullate gland of the spider Araneus sericatus." J Cell Biol **42**(1): 284-295.

Bittencourt, D., B. M. Souto, et al. (2007). "Spidroins from the Brazilian spider Nephilengys cruentata (Araneae : nephilidae)." Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology **147**(4): 597-606.

Blackledge, T. A., M. Kuntner, et al. (2011). The Form and Function of Spider Orb Webs: Evolution from Silk to Ecosystems. Advances in Insect Physiology. C. Jérôme, Academic Press. **Volume 41:** 175-262.

Bloch, F. (1946). "Nuclear Induction." Physical Review **70**(7-8): 460-474.

Brunger, A. T., P. D. Adams, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-921.

Bunning, J. D. and J. E. Lydon (1996). "The cellular optical texture of the lyotropic nematic phase of the caesium pentadecafluoro-octanoate (CsPFO) water system in cylindrical tubes." Liquid Crystals **20**(4): 381-385.

CACACE, M. G., E. M. LANDAU, et al. (1997). "The Hofmeister series: salt and solvent effects on interfacial phenomena." Quarterly Reviews of Biophysics **30**(03): 241-277.

Chen, G., X. Liu, et al. (2012). "Full-length minor ampullate spidroin gene sequence." PLoS One **7**(12): 14.

Colgin, M. A. and R. V. Lewis (1998). "Spider minor ampullate silk proteins contain new repetitive sequences and highly conserved non-silk-like "spacer regions"." Protein Science **7**(3): 667-672.

Delaglio, F., S. Grzesiek, et al. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.

Dicko, C., J. M. Kenney, et al. (2006). β–Silks: Enhancing and Controlling Aggregation. Advances in Protein Chemistry. J. M. S. Andrey Kajava and A. D. P. David, Academic Press. **Volume 73:** 17-53.

Dicko, C., D. Knight, et al. (2004). "Secondary structures and conformational changes in flagelliform, cylindrical, major, and minor ampullate silk proteins. Temperature and concentration

effects." Biomacromolecules **5**(6): 2105-2115.

Exler, J. H., D. Hümmerich, et al. (2007). "The Amphiphilic Properties of Spider Silks Are Important for Spinning." Angewandte Chemie International Edition **46**(19): 3559-3562.

Gaines Iv, W. A. and W. R. Marcotte Jr (2008). "Identification and characterization of multiple Spidroin 1 genes encoding major ampullate silk proteins in Nephila clavipes." Insect Molecular Biology **17**(5): 465-474.

Gaines, W. A., M. G. Sehorn, et al. (2010). "Spidroin N-terminal domain promotes a pH-dependent association of silk proteins during self-assembly." J Biol Chem **285**(52): 40745-40753.

Garb, J. E. and C. Y. Hayashi (2005). "Modular evolution of egg case silk genes across orb-weaving spider superfamilies." Proc Natl Acad Sci U S A **102**(32): 11379-11384.

Gatesy, J., C. Hayashi, et al. (2001). "Extreme diversity, conservation, and convergence of spider silk fibroin sequences." Science **291**(5513): 2603-2605.

Gosline, J. M., M. E. DeMont, et al. (1986). "The structure and properties of spider silk." Endeavour **10**(1): 37-43.

Gosline, J. M., M. W. Denny, et al. (1984). "Spider silk as rubber." Nature **309**(5968): 551-552.

Greenwood MJ, M., AR, and Harding, JS (2010). Behav Ecol.

Guerette, P. A., D. G. Ginzinger, et al. (1996). "Silk properties determined by gland-specific expression of a spider fibroin gene family." Science **272**(5258): 112-115.

Guinea, G. V., M. Elices, et al. (2003). "Self-tightening of spider silk fibers induced by moisture." Polymer **44**(19): 5785-5788.

Hagn, F. (2012). "A structural view on spider silk proteins and their role in fiber assembly." J Pept Sci **18**(6): 357-365.

Hagn, F., L. Eisoldt, et al. (2010). "A conserved spider silk domain acts as a molecular switch that controls fibre assembly." Nature **465**(7295): 239-242.

Hagn, F., C. Thamm, et al. (2011). "pH-dependent dimerization and salt-dependent stabilization of the N-terminal domain of spider dragline silk--implications for fiber formation." Angew Chem Int Ed Engl **50**(1): 310-313.

Hayashi, C. Y., T. A. Blackledge, et al. (2004). "Molecular and mechanical characterization of aciniform silk: uniformity of iterated sequence modules in a novel member of the spider silk fibroin gene family." Mol Biol Evol **21**(10): 1950-1959.

Heim, M., D. Keerl, et al. (2009). "Spider Silk: From Soluble Protein to Extraordinary Fiber." Angewandte Chemie-International Edition **48**(20): 3584-3596.

Herrmann, T., P. Guntert, et al. (2002). "Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA." J Mol Biol **319**(1): 209-227.

Hijirida, D. H., K. G. Do, et al. (1996). "13C NMR of Nephila clavipes major ampullate silk gland." Biophysical journal **71**(6): 3442-3447.

Hinman, M. B. and R. V. Lewis (1992). "Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes dragline silk is a two-protein fiber." J Biol Chem **267**(27): 19320-19324.

Holm, L. and P. Rosenström (2010). "Dali server: conservation mapping in 3D." Nucleic Acids Research **38**(suppl 2): W545-W549.

Huang, W., Z. Lin, et al. (2006). "Characterization and expression of a cDNA encoding a tubuliform silk protein of the golden web spider Nephila antipodiana." Biochimie **88**(7): 849-858.

Jaudzems, K., G. Askarieh, et al. (2012). "pH-dependent dimerization of spider silk N-terminal domain

requires relocation of a wedged tryptophan side chain." J Mol Biol **422**(4): 477-487.

Jin, H.-J. and D. L. Kaplan (2003). "Mechanism of silk processing in insects and spiders." Nature **424**(6952): 1057-1061.

La Mattina, C., R. Reza, et al. (2008). "Spider minor ampullate silk proteins are constituents of prey wrapping silk in the cob weaver Latrodectus hesperus." Biochemistry **47**(16): 4692-4700.

Landreh, M., G. Askarieh, et al. (2010). "A pH-dependent dimer lock in spider silk protein." J Mol Biol **404**(2): 328-336.

Laskowski, R. A., J. A. Rullmannn, et al. (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR **8**(4): 477-486.

LaVallie, E. R., E. A. DiBlasio, et al. (1993). "A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm." Biotechnology **11**(2): 187-193.

Lazo, N. D. and D. T. Downing (1999). "Crystalline Regions of Bombyx mori Silk Fibroin May Exhibit β-Turn and β-Helix Conformations." Macromolecules **32**(14): 4700-4705.

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

Lin, Z., W. Huang, et al. (2009). "Solution structure of eggcase silk protein and its implications for silk fiber formation." Proc Natl Acad Sci U S A **106**(22): 8906-8911.

Pawar, A. P., K. F. DuBay, et al. (2005). "Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases." Journal of Molecular Biology **350**(2): 379-392.

Pettersen, E. F., T. D. Goddard, et al. (2004). "UCSF Chimera--a visualization system for exploratory research and analysis." J Comput Chem **25**(13): 1605-1612.

Plazaola, A. and G. C. Candelas (1991). "Stimulation of fibroin synthesis elicits ultrastructural modifications in spider silk secretory cells." Tissue and Cell **23**(2): 277-284.

Purcell, E. M., H. C. Torrey, et al. (1946). "Resonance Absorption by Nuclear Magnetic Moments in a Solid." Physical Review **69**(1-2): 37-38.

Riekel, C., C. Branden, et al. (1999). "Aspects of X-ray diffraction on single spider fibers." Int J Biol Macromol **24**(2-3): 179-186.

Rising, A., G. Hjalm, et al. (2006). "N-terminal nonrepetitive domain common to dragline, flagelliform, and cylindriform spider silk proteins." Biomacromolecules **7**(11): 3120-3124.

Romer, L. and T. Scheibel (2008). "The elaborate structure of spider silk: structure and function of a natural high performance fiber." Prion **2**(4): 154-161.

Roseman, M. A. (1988). "Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds." J Mol Biol **200**(3): 513-522.

Scheibel, T. (2004). "Spider silks: recombinant synthesis, assembly, spinning, and engineering of synthetic proteins." Microb Cell Fact **3**(1): 14.

Shao, Z. and F. Vollrath (1999). "The effect of solvents on the contraction and mechanical properties of spider silk." Polymer **40**(7): 1799-1806.

Shao, Z., F. Vollrath, et al. (1999). "Analysis of spider silk in native and supercontracted states using Raman spectroscopy." Polymer **40**(10): 2493-2500.

Slotta, U. K., S. Rammensee, et al. (2008). "An Engineered Spider Silk Protein Forms Microspheres." Angewandte Chemie International Edition **47**(24): 4592-4594.

Spiess, K., A. Lammel, et al. (2010). "Recombinant Spider Silk Proteins for Applications in

Biomaterials." Macromolecular Bioscience **10**(9): 998-1007.

Sponner, A., E. Unger, et al. (2004). "Conserved C-termini of Spidroins are secreted by the major ampullate glands and retained in the silk thread." Biomacromolecules **5**(3): 840-845.

Terry, A. E., D. P. Knight, et al. (2004). "pH Induced Changes in the Rheology of Silk Fibroin Solution from the Middle Division of Bombyx mori Silkworm." Biomacromolecules **5**(3): 768-772.

Teule, F., A. R. Cooper, et al. (2009). "A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning." Nature Protocols **4**(3): 341-355.

Tugarinov, V., W. Y. Choy, et al. (2005). "Solution NMR-derived global fold of a monomeric 82-kDa enzyme." Proc Natl Acad Sci U S A **102**(3): 622-627.

van Beek, J. D., S. Hess, et al. (2002). "The molecular structure of spider dragline silk: Folding and orientation of the protein backbone." Proceedings of the National Academy of Sciences **99**(16): 10266-10271.

van Beek, J. D., J. Kummerlen, et al. (1999). "Supercontracted spider dragline silk: a solid-state NMR study of the local structure." Int J Biol Macromol **24**(2-3): 173-178.

Vasanthavada, K., X. Hu, et al. (2007). "Aciniform spidroin, a constituent of egg case sacs and wrapping silk fibers from the black widow spider Latrodectus hesperus." J Biol Chem **282**(48): 35088-35097.

Vollrath, F. (2000). "Strength and structure of spiders' silks." J Biotechnol **74**(2): 67-83.

Vollrath, F. and D. P. Knight (1999). "Structure and function of the silk production pathway in the spider Nephila edulis." Int J Biol Macromol **24**(2-3): 243-249.

Vollrath, F. and D. P. Knight (2001). "Liquid crystalline spinning of spider silk." Nature **410**(6828): 541-548.

Wang, S., W. Huang, et al. (2012). "NMR structure note: repetitive domain of aciniform spidroin 1 from Nephila antipodiana." J Biomol NMR **54**(4): 415-420.

Williamson, M. P., T. F. Havel, et al. (1985). "Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry." J Mol Biol **182**(2): 295-315.

Wishart, D. S., B. D. Sykes, et al. (1991). "Relationship between nuclear magnetic resonance chemical shift and protein secondary structure." J Mol Biol **222**(2): 311-333.

Xu, L., M. L. Tremblay, et al. (2012). "1H, 13C and 15N NMR assignments of the aciniform spidroin (AcSp1) repetitive domain of Argiope trifasciata wrapping silk." Biomol NMR Assign **6**(2): 147-151.

Xu, M. and R. V. Lewis (1990). "Structure of a protein superfiber: spider dragline silk." Proceedings of the National Academy of Sciences **87**(18): 7120-7124.

Xu, Y., Y. Zheng, et al. (2006). "A new strategy for structure determination of large proteins in solution without deuteration." Nat Methods **3**(11): 931-937.

Xu, Y. Q., D. Long, et al. (2007). "Rapid data collection for protein structure determination by NMR spectroscopy." Journal of the American Chemical Society **129**(25): 7722-7723.

Xu, Y. Q., Y. Zheng, et al. (2006). "A new strategy for structure determination of large proteins in solution without deuteration." Nature Methods **3**(11): 931-937.

Xue, B., R. L. Dunbrack, et al. "PONDR-FIT: A meta-predictor of intrinsically disordered amino acids." Biochimica Et Biophysica Acta-Proteins and Proteomics **1804**(4): 996-1010.

Yang, D. W., Y. Zheng, et al. (2004). "Sequence-specific assignments of methyl groups in high-molecular weight proteins." Journal of the American Chemical Society **126**(12):

3710-3711.

Yang, S. A., C. G. Noble, et al. (2009). "Characterization of DLC1-SAM Equilibrium Unfolding at the Amino Acid Residue Level." Biochemistry **48**(19): 4040-4049.

Zhang, S. and A. Rich (1997). "Direct conversion of an oligopeptide from a β-sheet to an α-helix: A model for amyloid formation." Proceedings of the National Academy of Sciences **94**(1): 23-28.

Zwahlen, C., P. Legault, et al. (1997). "Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex." Journal of the American Chemical Society **119**(29): 6711-6721.