UNIVERSITY OF CALIFORNIA SAN DIEGO

Signals and Conformational Changes in Trefoil Knotted SPOUT Methyltransferases

A thesis submitted in partial satisfaction of the requirements
for the degree of Master of Science

in

Chemistry

by

Tiange Tao

Committee in charge:

Professor Patricia A. Jennings, Chair
Professor Joseph Adams
Professor Simpson Joseph

2021

The Thesis of Tiange Tao is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

Table of Contents

Thesis Approval Page……iii

Table of Contents……iv

List of Figures and Tables……v

List of Abbreviations……vi

Acknowledgements……ix

Abstract of the Thesis……xi

Chapter 1. Introduction……1

Chapter 2. Optimization of Protein Purification of the Minimalist Methyltransferase, MTT_{TM}……10

Chapter 3. The Unfolding of Knotted Methyltransferase Protein Using Fluorescence Spectroscopy……22

Chapter 4. Monitoring Conformational Change of MTT_{SA} Unfolding Process by NMR Spectroscopy……35

Chapter 5. Conclusion and Discussion……43

Chapter 6. Materials and Methods……54

References……63

LIST OF FIGURES AND TABLES

Figure 1.1 Protein Visualization and Simplified Representation of MTT_{SA}……6

Figure 1.2 Protein Visualization and Simplified Representation of MTT_{TM}……7

Figure 2.1 SEC Chromatography Purification of T-WT……..17

Figure 2.2 SDS-PAGE Gel Image of T-WT Purification Results……..…18

Figure 2 .3 ESI-TOFMS Spectrum of Purified MTT_{TM}…….……20

Figure 2.4 Original and Deconvoluted Spectrum of Purified MTT_{TM} ……..… 21

Figure 3.1 Sequence of MTT_{SA} and MTT_{TM}…….30

Figure 3.2 Emission Spectrum of Denatured MTT_{SA}……..32

Figure 3.3 Emission Spectrum of Denatured MTT_{TM}……..34

Figure 4.1 Cartoon Visualization about Principle of NMR Spectroscopy……39

Figure 4.2 Chemical Shift of ^{1}H Spectrum in Amino Acids……40

Figure 4.3 Overlayed ^{1}H-^{15}N HSQC Spectrum of Native and Unfolded MTT_{SA}….42

Figure 5.1 Kinetics Spectrum of Unfolding MTT_{SA}……..49

Figure 5.2 Gnd-HCl Denaturation Curve of MTT_{TM}…….50

Figure 5.3 Contact Map and Cartoon Representation of MTT_{TM} by RaptorX…….52

LIST OF ABBRIEVIATIONS

1D	One dimensional
2D	Two dimensional
$AmSO_4$	Ammonium sulfate
B	external magnetic field
BME	2-mercaptoethanol
$CaCl_2$	Calcium chloride
Da	Dalton
DTT	Dithiothreitol
E.coli	*Escherichia coli*
ESI	electrospray ionization
$FeCl_3$	Iron(III) chloride
FPLC	Fast Protein Liquid Chromatography
FRET	Fluorescence resonance energy transfer
Γ, gamma	emissive rate of fluorophore
γ	gyromagnetic ratio of the nucleus
Gnd-HCl	guanidine hydrochloride
HSQC	heteronuclear single quantum coherence
Hz	Hertz
$I_{\parallel}$	vertically polarized emission
$I_{\perp}$	horizontally polarized emission
IPTG	Isopropyl β-D-1-thiogalactopyranoside

KH_2PO_4	Monopotassium phosphate
k_{nr}	rate of non-radioactive decay
LB	Miller Luria-Bertani Broth
μ	magnetic field vector
MD	molecular dynamics
$MgSO_4$	Magnesium sulfate
MM	Minimal media
MTase	Methyltransferase
MSA	Multiple sequence alignment
MTT_{SA}	Mesophile, M-WT
MTT_{TM}	Thermophile, T-WT
NaCl	Sodium chloride
NaOAc	Sodium acetate
$NaPO_4$	Sodium phosphate
Ni-NTA	Nickel-Nitrilotriacetic Acid
NMR	Nuclear Magnetic Resonance
pI	Isoelectric point
ppm	Parts per million
Q	Quantum yields
R	Fluorescence anisotropy
SAM	S-Adenosyl methionine
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size-exclusion column chromatography

SPOUT	SpoU-TrmD
τ, tau	fluorescence lifetime
τ_n	intrinsic lifetime
TOF	Time-of-flight
TROSY	Transverse relaxation optimized spectroscopy
UCH	ubiquitin C-terminal hydrolases
UV	Ultraviolet
$ZnCl_2$	Zinc chloride

ACKNOWLEDGEMENTS

This work was done as one of the most important milestones in my life. Without the help and support of following individuals, it would not be possible for me to complete my master's degree in Chemistry.

First of all, I would like to express my deepest appreciation to my advisor, Patricia Jennings, for always providing me supports and guidance. She is always considerate and patient whenever there are obstacles in this project. I deeply appreciate her for offering me a chance to do scientific research from the beginning back in 2017 when I was an undergraduate student. I would also like to thank my committee members, Professor Simpson Joseph and Professor Joseph Adams, for their valuable advice of reviewing my thesis, and their continues contributions to the field.

I would also like to extend my deepest grateful to Dr. Dominique Capraro, for always helping me with research ideas and teaching me new biochemistry techniques. It is my great pleasure to work with her together in this project, especially for her mentorship to my progress as a graduate student. In addition, there are also several members in Jennings lab that deserve the acknowledgement. I would like to thank Jiyue Chen, Maryam Ranjbar and Christopher Susanto for the advice that help me go through the struggles and difficulties in everyday lab work. I would also like to thank Kendra Hailey and Jason Stofleth for providing me advice in my early stage of scientific research. I also want to appreciate Dr Xuemei Huang and Dr Nicole Avakyan for their patience to assist me in the NMR and circular dichroism experiments. I have another special thanks to

Professor Wei Wang for his guidance and inspiration of valuable future research and career advice.

Finally, I am extremely grateful to my family and friends for their unconditional love, accompany and support for all my decisions. Especially my mother who has always been my role model as a great female in scientific research since my childhood. Without her giving and sacrificing for my education, I would not have chance to see the wider world. Another thank to my grandmother, the coolest old lady in my mind who always encourages me and takes care of me every time I go back home, I hope felicity and health will always be in your life.

ABSTRACT OF THE THESIS

Signals and Conformational Changes in

Unfolded SPOUT Methyltransferases withTrefoil Knot

By

Tiange Tao

Master of Science in Chemistry

University of California, San Diego, 2021

Professor Patricia A. Jennings, Chair

Understanding the complex knotted topology in protein folding is an important structural and functional indication as increasing families of protein discovered to contain deeply knotted structure. Methyltransferases in SpoU-TrmD (SPOUT) family catalyze the post-translational methylation of RNA assisted by the cofactor, S-Adenosyl methionine (SAM), and have a deeply fold trefoil knot that associated with the dimer interface and

ligand binding domain. Misfolding in neuronal ubiquitin C-terminal hydrolases (UCH), where SPOUT family protein associate with, can result in aggregation and cause neurodegenerative disease. “Minimalism” of knotted protein was used to investigate the unfolding/untying process mechanism and its functional relationship of the knot in a more straightforward manner. MTT_{TM} and MTT_{SA} are both SPOUT MTases that contain trefoil knot with similar structure but different stabilities. Thus, studying native and unfolded proteins can provide information about knot forming mechanism and functional approach.

MTT_{TM} purification protocol was modified, and the purification result was analyzed by mass spectroscopy. Fluorescence spectroscopy was utilized to monitor the tertiary structure change of native and unfolded MTases by tracing tryptophane signal for MTT_{SA} and tyrosine signal for MTT_{TM}. Highly refined structural and dynamics information of native and unfolded MTT_{SA} was obtained by NMR spectroscopy. Shifts in unfolded MTT_{SA} emission spectrum maximum as well as tryptophane signal in ^{1}H-^{15}N HSQC spectrum were observed, indicating possible solvent exposure and change in microenvironment of tryptophane. The unusual peak and shoulder appeared in the unfolded MTT_{TM} emission spectrum requires further investigation in tyrosine residues and its relationship with tertiary fold of the protein.

Chapter 1

Introduction

The Significance of Knotted Protein

Understanding how backbone knotted topology forms in proteins is a rare, complex, and unsolved phenomenon in protein folding. Over the last two decades of studies, more families of proteins have been identified to have a deeply knotted structure, as well as an indication of importance in protein structure and function (Flapan et al. 2019). Knotted proteins are found in a specific subclass of methyltransferases known as the SpoU-TrmD (SPOUT) family which catalyzes the post-translational methylation of RNA. The deep trefoil knot structure is uniquely classified as a class IV Rossman fold in the SPOUT family where the binding domain and the dimer interface reside within the knotted region (Hiroyuki 2017). Proteins within this family are associated with neurodegenerative diseases, such as the neuronal ubiquitin C-terminal hydrolases (UCH) (Virnau et al. 2011). Misfolding or untying those results in aggregation of UCH has been identified as the primary cause of some of the most common neurodegenerative diseases, such as Parkinson's disease, which results from deterioration of and death of dopaminergic neurons (Balestrino and Schapira 2020). According to the statistics from the Harvard NeuroDiscovery Center, there are 1 million people in the United States suffering from Parkinson's disease. Neurodegenerative diseases primarily happen among the middle-aged and elderly (Brambrink and Koerner 2007). The number is increasing as the population ages, where current clinical treatments are only able to control the symptoms as therapeutic methods are not yet available. Therefore, studying the complex mechanism and topology of forming a knot and untying it within the methyltransferase family is significant as any insight gained may contribute to therapeutic insights into mitigating disease states. In addition, understanding the roles of the knot in both the dimerization and modification of nucleosides is important in determining how the knot structure contributes to enzymatic activity (Hiroyuki 2017). In this work, two methyltransferase proteins (PDB ID: 1O6D and 1VH0) containing a deep trefoil knot

are studied. Using the less complex trefoil knot, details about the knotting mechanism and the functional relationship may be more straightforward and may lay the groundwork for more complex cases.

General Theory and Topology Description of the Knot

In order to study the topology of knotted proteins, it is essential to understand the topology description and theoretical pathway of knot formation. The standard annotation for knot description is composed by a numerical number, which describes the minimum number of crossing in the knot protein projections; a subscript, which identifies the particular knot within the crossing; and the + and – sign, which distinguishes the chirality of two knotted enantiomorphs (Cromwell 2004). According to the systematic reviews of the PDB, five distinct knot types have been identified, and those five subclasses of knots have been identified in proteins among the 23 families of knotted protein (Jackson et al. 2017), and 19 families of the protein contain -3_1 and $+3_1$ knots. The rest of the families include 4_1, -5_2, and $+6_1$ knots (Flapan et al. 2019). Taylor's twisted hairpin theory is a knotting mechanism that is commonly used to describe the knot folding process. The protein is assumed to first form a "twisted hairpin" of a loop as a "cinch" in the polypeptide chain or a mathematically proper term "zero knots" that happens relatively rare in proteins. Then, the C-terminus of the protein sequence is threaded through this loop to form a knot (Taylor 2007). Another theory representing the possible knot folding mechanism discovered more recently is called the loop-flipping mechanism (Flapan et al. 2019). Generally speaking, the unknotted open-chain first forms two zero knots and comes close together. Next, one end of the chain gets close to the two-loop region, and either loop flips over the terminus. The knot is formed by the terminus threading through the crossing loops.

The folding mechanism of the trefoil knots that commonly found in α/β knotted methyltransferase obtained from the computational modeling utilizing energy landscapes theory was indicated in the paper of Sulkowska's group. Firstly, the correctly twisted-native loop formed through the nucleation sites that overcomes the first energy barrier from unfolded polypeptide backbone to the intermediate state. Next, either a plug or slipknot is formed that must be based on the correct chirality of the loop, and then threaded by the C-terminus to overcome the next energy barrier and form the deeply tied trefoil knot (Noel and Sulkowska 2010). These folding steps and theories can describe some general folding mechanisms of the knotted proteins. However, to describe the more complex mechanism of the SPOUT tRNA MTases, particular parameters need to be considered and more detailed simulations and experiments are required to be done.

"Minimalism" of Knotted Protein

Two proteins within the SPOUT family identified as possible tRNA MTases, which contain deeply trefoil ($+3_1$) knots, are being investigated within this work. These two proteins are the thermophile (MTT_{TM}) with PDB code 1O6D and the mesophile (MTT_{SA}) with PDB code 1VH0. A straightforward sequence analysis by BLAST and protein sequence alignment by NCBI indicate that 1O6D contains 163 amino acids in the sequence and is found in the organism of T*hermotoga maritima* and may be a 23s rRNA methyltransferase (RImH). The organism taxonomy of T*hermotoga maritima* is first discovered in the marine geothermal area, resides in hot springs, and can grow in the high temperature range from 50℃ to 90℃ (Huber et al. 1986). Similarly, 1VH0 has a similar classification as a 23s rRNA methyltransferase with a sequence length of 161 amino acids. It has been identified within the OrfX gene of MRSA, an antibiotic resistant version of S*taphylococcus aureus*, a gram-positive bacterium that is positive

for catalase and denitrification and is frequently found in the skin and upper respiratory tract (Masalha et al. 2001). Both of these proteins are classified as a so-called “minimalist knot” as their overall structure contains the minimal core components that contribute to the knotted SPOUT classification. The SPOUT MTases are the S-adenosyl-L-methionine dependent enzymes with unique alpha helices and beta sheets folded to form a deeply folded knot (Tkaczuk et al. 2007). As the knot provides extra stability to the protein, despite the protein being denatured, the knot may still exist, and the protein becomes unfolded but tied (Andrews et al. 2013). Moreover, the intrigue in studying both the MTT_{Tm} and MTT_{Sa} proteins is that there is little information the methylation function. However, there are also challenges in studying the minimalist system of the trefoil knot due to their highly similar structure but different tying/untying mechanism. MTT_{TM} contains the knot that provides extra stability to the protein as an extreme example, while being able to survive under the high temperature. Even though MTT_{Sa} has a similar knot placement, the stability of the knot is very different from MTT_{TM}, and thus the knot forming mechanism between these two proteins may be very different. Therefore, by comparing the native and unfolded states of these two proteins, it would provide better understanding of how knot is formed and occurred.

Visualizing the biological assembly of the SPOUT MTases would be helpful to understand the structure and folding of the knot. Figure 1.1 and 1.2 includes the protein visualization and its simplified scheme. Both MTases are composed of five parallel aligned beta-sheets with five alpha-helices packed like a “sandwich”. The knot region of both proteins appears near the C-terminus and is characterized by the interactions of three loops, including crossing loop, threading loop, and wing loop. The three-loop region is also identified to form a binding pocket for SAM (Anantharaman et al. 2002). The 3D molecular representation and the simplified diagram of MTT_{SA} (PDB code 1VH0) and MTT_{TM} (PDB code: 1O6D) were shown in

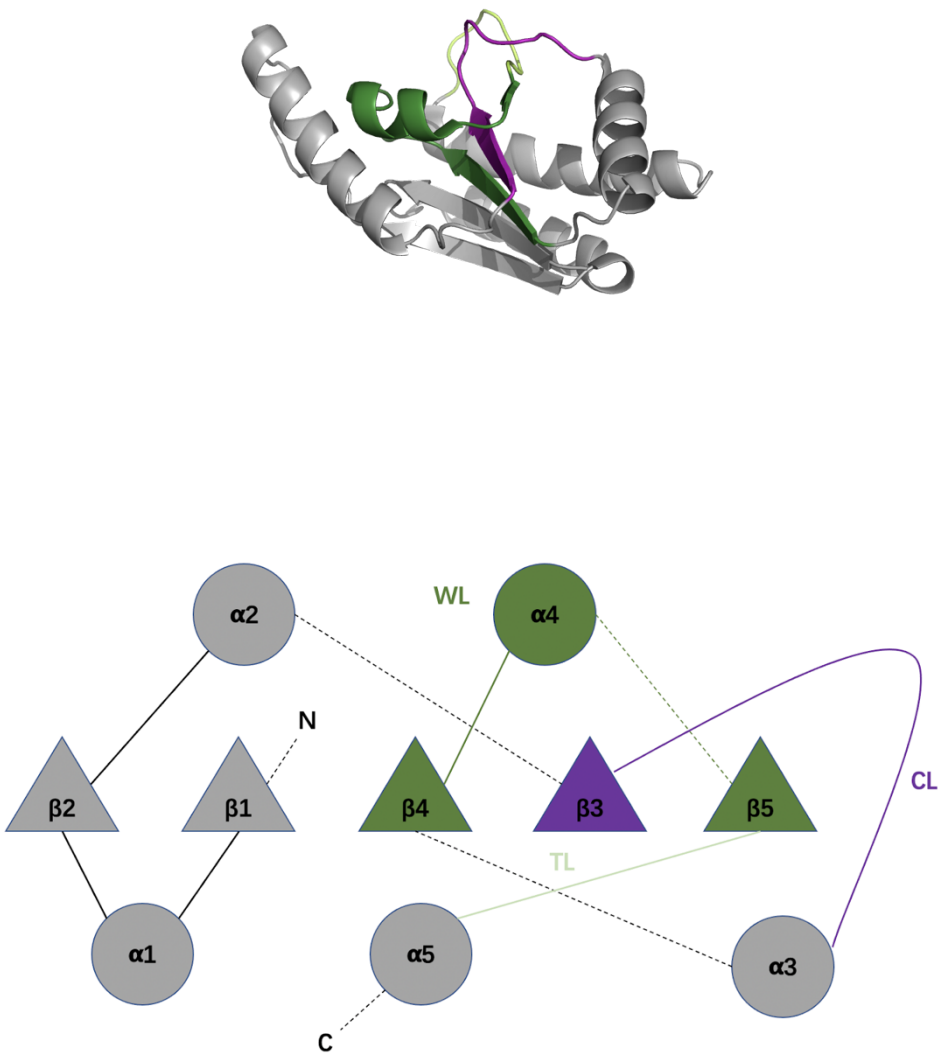


Figure 1.1

Protein Visualization using PyMol and Simplified representation of MTT_{Sa}. Knot contributing regions are highlighted using dark green (WL: wing loop), light green (TL: threading loop), and purple (CL: crossing loop). Beta-sheets are presented by the triangular shape; and alpha-helices are presented by circular shape, with numbers labeled as the manner from N-terminus to C-terminus. The loops are presented by solid (above the surface) and dash (below the surface) lines that connect alpha helices and beta sheets.

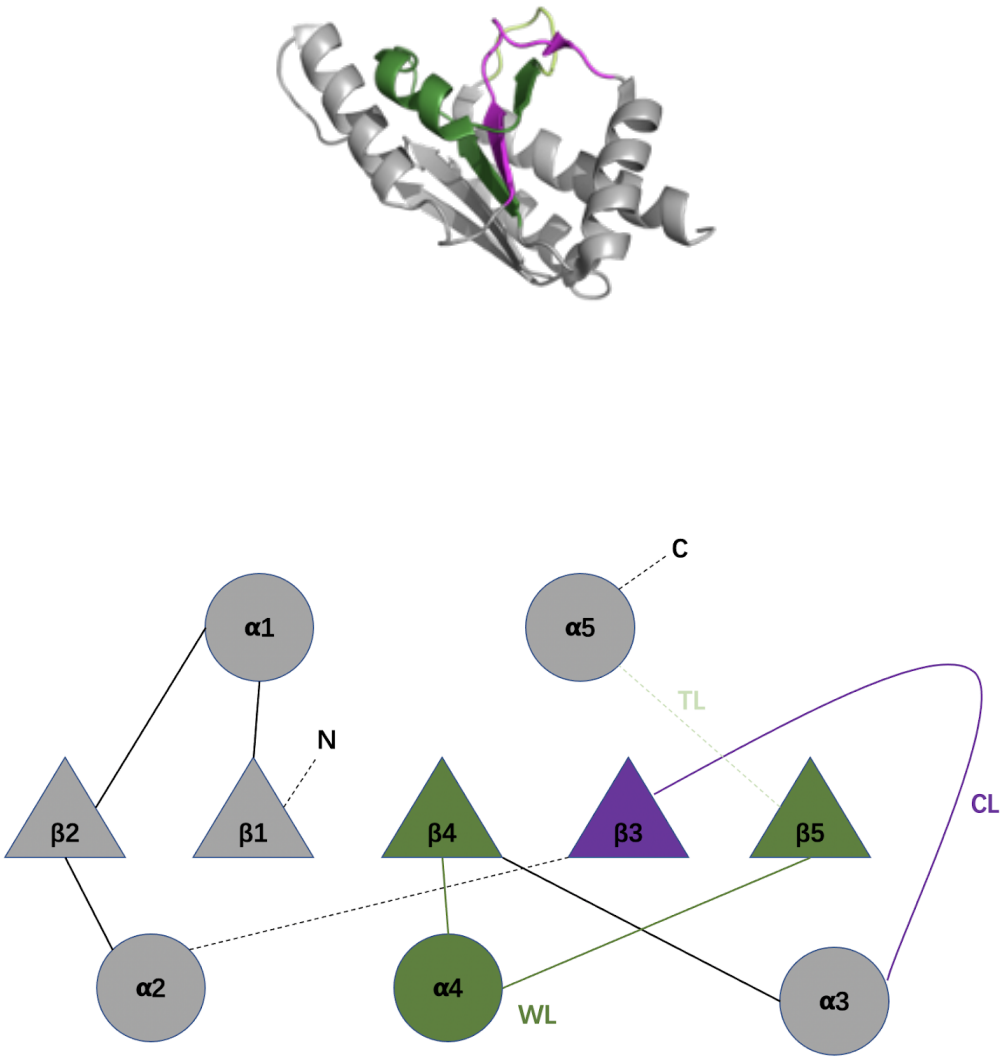


Figure 1.2

Protein Visualization using PyMol and Simplified representation of MTT_{TM}. Knot contributing regions are highlighted using dark green (WL: wing loop), light green (TL: threading loop) and purple (CL: crossing loop). Beta-sheets are presented by the triangular shape; and alpha-helices are presented by circular shape, with numbers labeled as the manner from N-terminus to C-terminus. The loops are presented by solid (above the surface) and dash (below the surface) lines that connect alpha helices and beta sheets.

Figure 1.1 and Figure 1.2. The crossing loop is from E76 to L82; the threading loop is from S125 to F131; secondary structural units of the trefoil knotted region were highlighted by different regions, and the folding layout of SPOUT family MTase was shown in the simplified diagram. In MTT_{SA}, the wing loop is from G107 to N121. Similarly, in MTT_{TM}, the crossing loop is from L112 to F122; the threading loop is from F65 to V76; the wing loop is from I95 to S116.

Another exceptional property of the knotted protein is that the knot itself can undergo the unfolding/untying process for long periods. Given that the knot is untied on a different time scale (at least 20-fold slower) than the unfolding of the protein, hysteresis introduced by the dynamic nature of the knot indicates that there is a gap between the reassembly of the protein when it is fully unfolded and untied (Andrew et al. 2013). While it is commonly stated from other research that the knotted MTases are unlikely to be untied from the denatured state of the protein (Mallam et al. 2010), it is still possible to untie the knot from the unfolded state if enough time is given. In addition, the topological change of the knot is a very complicated process that not only provides extra stability to the protein but also may contribute to the allosteric effects upon ligand binding (Capraro et al. 2020). Therefore, understanding the unfolding and refolding process and structural function of the trefoil knot is crucial for answering the question of how the conformation mechanism of the knot relates to the function of the protein due to its role in tRNA interactions.

In this work, we aim to identify the molecular determinants that contribute to the overall observation of slow unfolding and untying. To approach this, protein purification protocols were first modified and ameliorated. In addition, several sets of preliminary data of the unfolding experiments were obtained through different spectroscopic approaches. These include tracing the tryptophan and tyrosine signals within different denatured states using fluorescence spectroscopy; circular dichroism (CD) that is sensitive to alterations in the secondary structure

of the protein; and nuclear magnetic resonance (NMR) spectroscopy was utilized to obtain the structural and dynamic characteristics of the native fold and unfolded SPOUT MTase in the highly refined scale. Experiments were designed and conducted to elucidate key interactions within the backbone region of the proteins that contribute to the overall knot folding, as well as their functional correlations in the rRNA modification.

Chapter 2

Optimization of Protein Purification of the Minimalist Methyltransferase, MTT_{TM}

Introduction

Developing an effective protocol for the growing, isolating, and storage of a target protein and continual optimization of these steps is critical in biochemistry research. Obtaining pure protein samples at high concentrations is crucial as it facilitates the investigation of protein's functional and structural studies (Berg et al. 2002). Therefore, the selection of purification techniques is a crucial determinant for successfully separating the protein of interest from the unwanted mixtures. This chapter focuses on optimizing the protein expression and purification of the minimalist Methyltransferase from *Thermotoga maritima* (MTT_{TM}, T-WT, PDB ID 1O6D). The general process of protein purification includes cell harvesting, intracellular extraction, purification, isolation, and concentration (Nfor et al. 2007).

Initially, the protein gene is cloned into a plasmid vector and transformed into *Escherichia coli* (*E.coli*) for subsequent growth and protein expression. For our studies we use the pET24d vector as this expression system designed to yield high amounts of a target protein. The target protein is then extracted from the harvested cells by disrupting the cell wall using external forces. In most cases, a mechanical force is applied such as homonigizing or sonication. Here, sonication was conducted for the cell lysing process using the Sonic Dismembrator (Fisher Scientific) by converting the sound energy at 20 to 40 kHz frequency to the physical vibration, also known as ultrasonication (Shah et al. 2020). An ultrasonic probe that is suspended in a cell slurry is able to lyse open the cells and release the soluble protein in the buffer. However, care must be taken when using sonication as poor protein extraction may result from local overheating, bucking, and dislocation (Taylor, 2021). These issues can be alleviated by placing protein samples on ice and providing enough rest time to relieve the stress of the protein. Partial separation of impurities and cellular debris can also be accomplished physically, such as through filtration and centrifugation. Additionally, depending on whether the

protein is soluble in the buffer or the protein is in the insoluble inclusion bodies (Taylor 2021), additional steps may be necessary in order to prepare the target protein for downstream purification. For T-WT, targeted knotted protein is soluble upon release by sonication; thus centrifugation is an effective way to remove the cell debris.

In our case, we use affinity chromatography with a polyhistidine protein tag (His-tag protein) that has six histidines attached at the C-terminus of the protein sequence, which has a high affinity for the metal ions. Thus, metal affinity using a Ni-NTA matrix is considered to be an effective technique for protein purification. General steps of Ni-NTA affinity chromatography include binding, washing, and eluting with the preference of around pH 8.0 for protein binding with Ni-NTA resin (QIAGEN) effectively and increasing imidazole concentration for protein elution. While the Ni-NTA matrix is effective in isolating the target protein, additional purification steps are needed for the protein purity required for the subsequent optical experiments. Therefore, a post affinity Ni-NTA chromatography step was explored in order to maximize protein purity.

In this chapter, a high-resolution size-exclusion column chromatography (SEC) is introduced and applied as an optimization step in the overall purification of MTT_{Tm}. Using Fast Protein Liquid Chromatography (FPLC), the protein samples can be separated based on their molecular weight (MW) or molecular weight distribution (MWD) (Mori and Barth 1999). Several key components are essential for SEC purification. The stationary phase is packed in a column with polymer beads containing identical pore size. The elution from the Ni-NTA step that contains target protein is loaded onto the SEC column and flushed with a mobile phase at fixed pressure and flow rate. Protein with smaller size is easier to be retained in the pore of the column, and thus separation due to protein size becomes resolute (Wu 2003). A sensitive UV concentration detector after the column can detect the protein's concentration following the

absorbance signal at 280 nm (A280). The protein concentration is proportional to the A280 signal; thus, the amount of protein is proportional to the area under the peaks of the FPLC trace.

Choosing a buffering system in protein purification and storage is also one of the key determinants for the purification results. Several factors determine whether the buffer is effective, including the pH, salts, reducing agents, and stabilizing elements. The pH of the buffer depends on both the purification method and the isoelectric points (pI) of the protein. As mentioned earlier in Ni-NTA column chromatography, an adequate protein solution that enables His-tag protein to bind with the resin is around pH 7 - 8, although these conditions may not be ideal for protein storage. Therefore, the pI of the protein and the subsequent experimental conditions may help determine a suitable buffer to avoid protein precipitation upon storage.

Furthermore, different pKa values of the buffer system affect the buffer's performance, and thus the pKa of the chosen buffer should be within the range of the pH of choice. Reducing agents are also added to the buffer to avoid the unwanted linkage via intra and intermolecular disulfide bonds. For example, DTT is used in buffers to avoid unwanted disulfide linkage and prevent protein from oxidation. However, DTT might affect the binding ability to the Ni-NTA resin, thus should be avoided in the affinity purification step. Stabilizing agents such as glycerol are able to help prevent aggregation and stabilize the protein, which are also suggested to consider when choosing the buffer.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a low-cost and commonly used tool to analyze protein samples by separating due to molecular weight between 5 and 250 kDa (Laemmli 1970). In this chapter, SDS-PAGE is used to detect the protein's existence and purity. Sodium dodecyl-sulfates (SDS) and 2-mercaptoethanol (BME) are required in the denaturation step to unfold the protein into a linear form (Nowakowski, 2014).

Heating protein samples at 70-80 ℃ is also essential to assist the linearization of the native protein. The SDS-PAGE is implemented by the appropriate voltage (from 100-120 V) that migrates the negatively charged protein to a positively charged anode. Proteins with different molecular weights are separated as it migrates through the polyacrylamide gel, which maintains the charge of the protein. Coomassie blue is used to visualize the SDS-PAGE results through the steps of staining and destaining.

Another approach to analyze the purity of the purification results is protein mass spectroscopy, which provides a more sensitive and accurate analysis in protein identification, size quantification, and characterization (Domon and Aebersold 2006). Activated by ionization of the gas phase molecule, samples can be separated by mass during the step of acceleration in the magnetic field. The history of biomolecular mass spectroscopy started in 1989, enabling the ionization of unimpaired polypeptides through electrospray ionization (ESI) (Fenn et al. 1989). Time-of-flight (TOF) analyzers are able to calculate the mass-to-charge ratio of an analyte ion by measuring its flight time through a vacuum-sealed tube of a certain length. Thus, ToF analyzer performance has vastly improved, especially resolution and mass accuracy (Domon and Aebersold 2006). The instrument of mass spectroscopy used in this chapter is the Agilent 6230 Accurate-Mass TOF MS system. It enables the separation with high accuracy and sensitivity as well as fast-speed data acquisition. The TOF MS system is combined with a liquid chromatography system (LC) called Agilent 1260 Infinity Binary LC for mass analysis, which makes this approach, LC-ESI-TOF mass spectroscopy, suitable for protein analysis.

Results

The purified samples eluted by imidazole from Ni-NTA column chromatography were concentrated to ~10mL and loaded in the FPLC and further purified by S200 size exclusion

column chromatography. Figure 2.1 is the FPLC trace indicating the separation of proteins based on their size. It is observed that the major peak elutes from volume ~280 mL to 318mL, which is the expected fractions containing isolated T-WT. In addition, the FPLC trace is also able to reflect the approximate protein concentration for each fraction through the A280 absorbance (unit: mAU). The major peak for this FPLC trace have the A280 maxima for ~860 mAu at ~310 mL, indicating the relatively high protein concentration at that volume. Also, the larger area under the curve generated in the FPLC trace represents the higher amount of protein in the fractions. The UV detector on FPLC is able to provide an approximation on protein concentration. However, a more accurate and precise protein concentration is measured by Nanodrop Spectrometer or UV-Vis spectrophotometer.

Samples of fractions collected from the eluted S200 size exclusion column were linearized and loaded on SDS-PAGE gel. Figure 2.2 shows the SDS-PAGE gel image of the protein samples before injected to the FPLC and after eluted from the FPLC. As the gel image shows that protein concentration was diluted after isolation by SEC on the FPLC. However, the S200 size exclusion column chromatography is able to produce relatively pure proteins. This result indicates that the modified purification protocol was successful.

To verify the isolated protein is the target T-WT and has not degraded or been modified, protein samples were analyzed on the LC-Agilent 6230-ESI-TOF MS Spectrum. Figures 2.3 and 2.4 illustrate the spectrum of purified T-WT samples. Only one major peak is observed, the samples that were detected at the specific molecular weight. The deconvoluted mass spectrum of T-WT the measured molecular weight of the sample, in which the molecular weight of the purified protein sample is 18813 Da. The theoretical molecular weight value from Protein Data Bank (PDB) of the 1O6D is 18.9 kDa. Therefore, the observed molecular weight of our purified protein is consistent with the theoretical mass value as the difference in value is consistent with

the loss of methionine after translation, indicating the protein purification is successful with high purity.

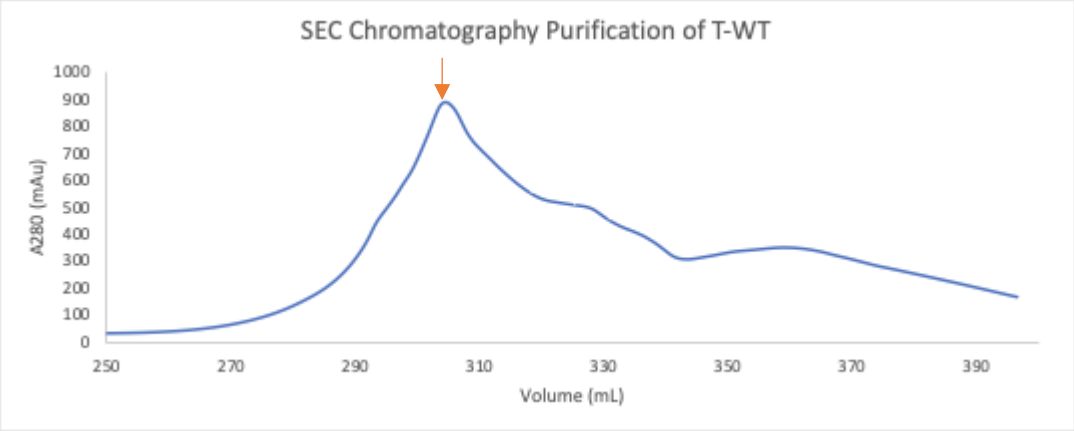


Figure 2.1

A. FPLC Trace of T-WT Purification is presented in this figure, in which x-axis is the elution volume (mL), and y-axis represents the A280 signal (mAU) detected by UV detector. Fractions were collected and pooled from each peak. The orange arrow represents the sample on the SDS-PAGE gel, Lane 4, in Figure 2.2.

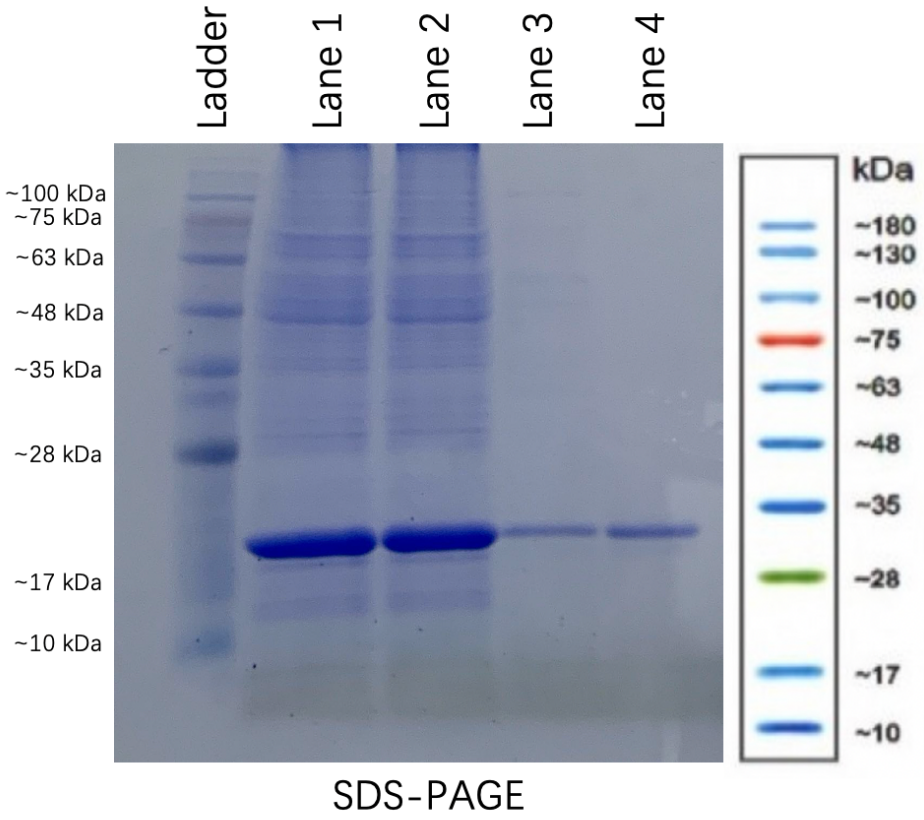


Figure 2.2

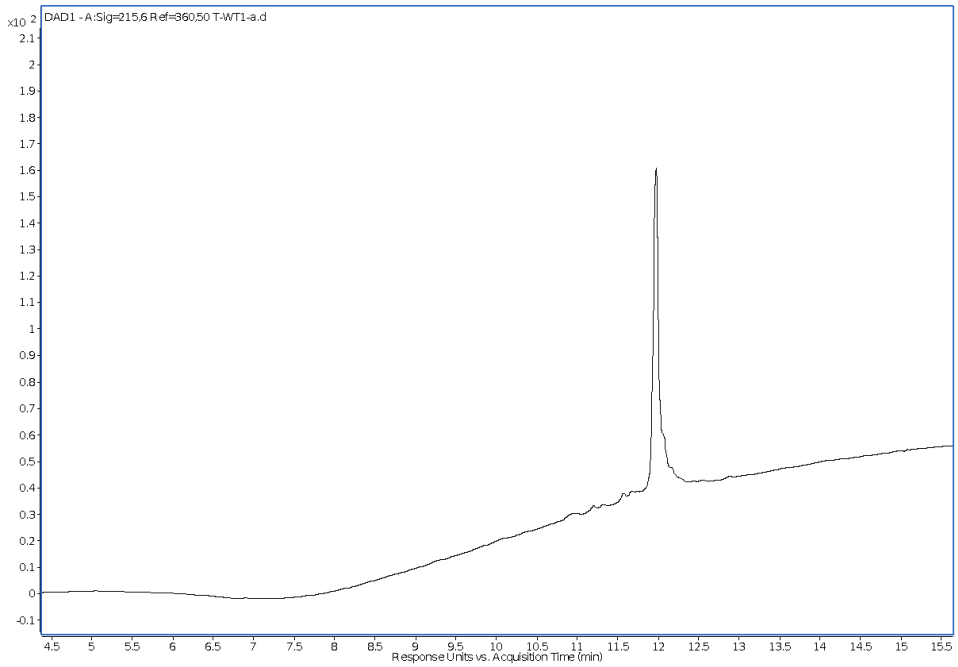
SDS-PAGE gel image of the purified T-WT from FPLC results. Purified protein samples from FPLC eluted fractions were run on the SDS-PAGE gel to assess the purity and existence, which used G02101 as protein ladder. Samples were loaded to the gel in the following way. Lane 1: T-WT Elution from Ni-NTA column chromatography. Lane 2: Concentrated and filtered T-WT Elution. Lane 3: Fraction eluted at 295 mL from the FPLC. Lane 4: Fraction eluted at 310 mL (orange arrow) from the FPLC.

Figure 2.3

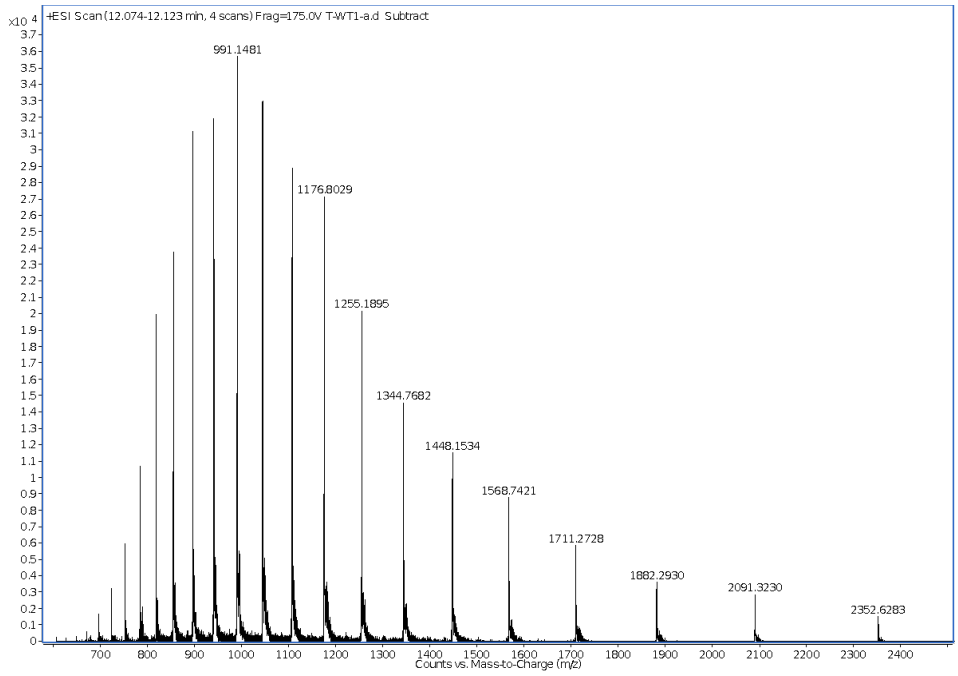
A. LC-Agilent 6230-ESI-TOF MS Spectrum of purified WT of Thermophilic Methyltransferase (T-WT). From the spectrum of Response Units v.s. Acquisition time, one majority peak is identified at 12 minutes. The signal was obtained at UV 210nm.

B. LC trace peak of the identified major peak from LC-Agilent 6230-ESI-TOF MS Spectrum. The graph shows the results of a sample separated by liquid chromatography, which recorded the counts vs. mass-to-charge of the protein sample.

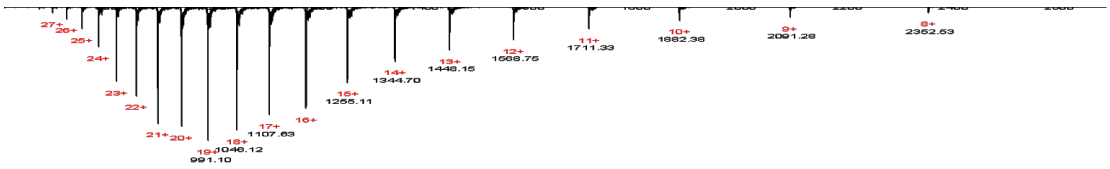
A.


DAD1 - A:Sig=215,6 Ref=360,50 T-WT1-a.d
x10 2
Response Units vs. Acquisition Time (min)

B.


+ESI Scan (12.074-12.123 min, 4 scans) Frag=175.0V T-WT1-a.d Subtract
x10 4
991.1481
1176.8029
1255.1895
1344.7682
1448.1534
1568.7421
1711.2728
1882.2930
2091.3230
2352.6283
Counts vs. Mass-to-Charge (m/z)

A.



B.

.

Figure 2.4

A. Original ESI-TOFMS spectrum of purified T-WT obtained from LC-Agilent 6230-ESI-TOF Mass Spectroscopy, with mass-to-charge (mz) value of each peak separated by LC labeled in black color text. The charges of each fragment shown on the LC trace were labeled in red text. The final molecular weight of the majority peak was determined using data shown in this graph.

B. Deconvoluted mass spectrum of T-WT from LC-Agilent 6230-ESI-TOF Mass Spectroscopy. This mass spectrum shows the processed data of the purified T-WT sample, which has only one majority peak by mass, 18813 Da. The result from Mass Spectroscopy matches the literature value on PDB (18.9kDa), which indicates the purification was successful.

Chapter 3

The Unfolding of Knotted Methyltransferase Protein Using Fluorescence Spectroscopy

Introduction

Fluorescence and phosphorescence are two types of molecular luminescence that occur in nature as due to the process of emission from electronically excited states to lower states (Lakowicz 2017). The fact that fluorescence luminescence absorbs energy, such as ultraviolet (UV) light, and emits immediately after absorption, is different from phosphorescence luminescence, which is able to store the absorbed light and emit it later. Fluorophore refers to chemical compounds or materials that are able to absorb light and re-emit the energy absorbed. In 1933, Polish physicist Aleksander Jabłoński established the Jabłoński diagram, which illustrates the molecular process of absorption and emission of the light, and can be represented by two terms, quantum yields (Q) and fluorescence lifetime (τ, tau) (Jablonski 1933). The quantum yield (Q) represents the ratio of the number of emitted photons to the number of absorbed photons. The lifetime (τ, tau) of the fluorophore is defined by the average time the molecule spends during the energy excitation and emission to the ground state. Equation 3-1 describes the relationship between the fluorophore emission rate and the quantum yield, where k_{nr} represents the rate of non-radioactive decay; $\boldsymbol{\Gamma}$ (gamma) represents the emissive rate of fluorophore; Q represents the quantum yield of the photon. When k_{nr} is much less than $\boldsymbol{\Gamma}$ ($k_{nr} \ll \boldsymbol{\Gamma}$), the quantum yield is approaching 1.

3-1 $$Q = \frac{\Gamma}{\Gamma + k_{nr}}$$

Equation 3-2 shows the expression of the fluorescence lifetime. τ represents the lifetime of the fluorophore, which is usually around 0.1 - 4 ns for a protein fluorophore.

3-2 $$\tau = \frac{1}{\Gamma + k_{nr}}$$

Natural lifetime or intrinsic lifetime (τ_n) refers to the lifetime of fluorophore in the absence of nonradiative processes. It is an essential property that measures the probability of finding an electron and an electron hole in the same low-dimensional excitation state (Chen et al. 2013). The natural lifetime of a fluorophore can be expressed as a function of lifetime and quantum yield in equation 3-3:

3-3 $$\tau_n = \frac{1}{\Gamma} = \frac{\tau}{Q}$$

This concept correlates with fluorescence quenching, which represents the decreased intensity of fluorescence. The timescale of quenching is best illustrated by the dynamic process in biomolecules by understanding the role of an excited-state lifetime. Collisional quenching refers to the decreased fluorescence intensity during the deactivation of the excited-state fluorophore that collides with the quencher (such as oxygen, halogens, amines, and electron-deficient molecules) in solution. Since excitation is an instantaneous process, the absorption spectrum is not able to provide enough information about the non-radioactive path to the ground state. However, the collision quenching is able to provide time and distance expansion through the solvent relaxations, which is a relatively long timescale compared to excitation. The solvent molecules reorient at 10^{-10}s (around excited states) and shift the emission states to longer wavelengths by lowering their energy level.

Anisotropy refers to the variable physical properties of substances along different molecular axes, and it is measured based on the principle of photoselective fluorophores excitation, which is initiated by polarized light. Fluorescence anisotropy (r) can be expressed in the equation 3-4, where r represents fluorescence anisotropy, $I_{\parallel}$ represents the vertically polarized emission, and $I_{\perp}$ represents the horizontally polarized emission.

3-4 $$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}}$$

Therefore, fluorescence anisotropy provides insight in the dimensional, conformational, and rigidity properties of the protein in different environments. Randomly oriented fluorophores have the preference to absorb light and photons, which have electric vectors parallel to the fixed-oriented transition. After considering the factor of rotational diffusion that decreases anisotropy value and displaces the emission dipole of the fluorophore, the expected anisotropy (r) can be determined using the Perrin equation (3-5). Here, r_0 refers to anisotropy without the effect of rotational diffusion and θ refers to the rotational correlation time. The rotational diffusion can provide information about the angular displacement between absorption and emission.

3-5 $$r = \frac{r_0}{1 + (\frac{\tau}{\theta})}$$

Fluorescence resonance energy transfer (FRET) is a distance-dependent energy transfer that occurs when the emission spectrum of one donor fluorophore overlaps with the absorption spectrum of the other acceptor fluorophore. The rate of resonance energy transfer depends on the distance between the donor and acceptor, which is affected by their overall concentration coupled by dipole-dipole interaction. The rate of energy transfer is expressed in equation 3-6, where k_T refers to the rate of energy transfer, τ_D refers to the lifetime of donor, r refers to the distance between donor and acceptor, and R_0 refers to the spectrum overlap.

3-6 $$k_T(r) = \frac{1}{\tau D} \left(\frac{R_0}{r}\right)^6$$

The efficiency of FRET (E) can also be expressed using distance (r) as a variable in equation 3-7:

3-7 $$E = \frac{R_0^6}{R_0^6 + r^6}$$

With measuring the spectrum overlap, FRET is considered an approach for measuring distances on different protein and biological macromolecules sites.

There are two types of biochemical fluorophores, intrinsic (occurs naturally) and extrinsic (added to the sample) fluorophores. Different types of biomolecules have different wavelength absorption and emission. For example, the indole group in tryptophan and aromatic amino acids appears to be a dominant intrinsic fluorophore. Other biomolecules, such as DNA have weak fluorophores that require staining with fluorescent dye in order to be visualized. Some other substances, such as NADH and oxidized flavins, can be used as extrinsic probes, which have high intensities and stabilities but do not distract the labeled macromolecules.

There are three types of intrinsic fluorophores in proteins that can be used in fluorescence spectroscopy. Aromatic amino acids in proteins are able to absorb and emit energy at identical wavelengths due to the effect of sidechain, solvent pH, concentration, and folding or unfolding events. These amino acids include phenylalanine (Phe, F), tyrosine (Tyr, Y), and tryptophan (Trp, W), which are relatively rare in comparison to other amino acids in the protein. As a result of the limiting presence of aromatic amino acids within a given polypeptide chain, following the fluorescence signal is efficient and straightforward, depending on which aromatic residue is present. Phenylalanine has a relatively weak absorption of ultraviolet light and emission at around 282 nm. Tyrosine is a relatively simple fluorophore with absorption at around 275 nm and emission at around 303 nm. Of the three amino acids fluorophores, tryptophan has the most significant signal in both the absorption and emission spectra. The absorption of UV fluorescence is broader as it goes from 295-305 nm, while the emission is at

350 nm. This is a dominant signal in the emission spectrum that is highly sensitive to the local micro-environment as changes in the emission spectrum through unfolding events, conformational changes, or ligand binding (Lakowicz 2017). Therefore, tracing the signals of intrinsic protein fluorophores in fluorescence spectroscopy is a practical approach in biophysics due to the different levels of sensitivity in the respective fluorophore's local environment. For example, tyrosine has relatively low sensitivity to the local environment because its emission occurs from a single electronic state. The single electronic transition has constant anisotropy over a long absorption wavelength range (Ross et al. 1992). Tyrosine can also easily lose a proton on the aromatic hydroxyl group and undergo an excited ionization-state at neutral pH, where the fluorescence is quenched (Creed 1984). On the other hand, the emission spectrum of tryptophan has a relatively high sensitivity to the local environment due to its two nearby isoenergetic transitions. The anisotropy of the indole group in tryptophan is not constant over the long-wavelength range. Since the anisotropy depends on the fractional absorption distribution of each state, the emission maximum is more sensitive in tryptophan, and therefore it can usually reflect the structural information of a protein (Eftink et al. 1990).

Result

In this chapter, taking advantage of the different fluorophores that are intrinsically available in each protein, the emission spectra were obtained from the native and partially unfolded states of the knotted proteins. Figure 3.1 shows the sequence of MTTSA and MTTTM with the intrinsic fluorophores highlighted. The protein sequence of MTTSA has a tryptophan (W18) but also has several tyrosines (Y17, Y25, Y32, Y125, Y148, and Y159) and phenylalanines (F105, F107, F129, F134, and F151). On the other hand, MTTTM lacks tryptophan in the sequence but contains several tyrosines (Y24, Y105, Y151, and Y153) and

phenylalanines (F16, F27, F31, F67, F83, F86, F112, F119, F126, F140, and F143). Using the specificity in the aromatic amino acids emission maxima, a wavelength profile of the respective proteins in native and partially unfolded (through incubating in denaturants over time) conditions was collected. Due to the resonance energy transfer from phenylalanine to tyrosine to tryptophan as their spectral properties, the emission spectrum of MTT_{SA} was traced dominantly from the tryptophan signal, which was excited at 295 nm and collected the emission spectrum in the range of 280 nm to 450 nm with 3 nm slit width. Moreover, the emission spectrum of T-WT was traced dominantly from the tyrosine signal, which was excited at 275 nm, and the emission spectrum was collected from 265 nm to 450 nm with a slit width of 5 nm.

Fractions of purified samples from S200 column chromatography were collected and concentrated to at least 0.3 mg/mL for optimal fluorescence signal. Native knotted protein was incubated in a denaturation buffer. Emission spectra were collected over various timeframes (1 Day, 5 Days and 11 Days), as it has been previously revealed that unfolding and untying of the knotted structure is remarkably slow because of o the kinetically uncoupled process of folding and tying the knot (Capraro and Jennings 2016). Interestingly, given the presence of a single tryptophan and sensitivity of the microenvironment of this fluorophore to changes, small amounts of denaturant may reveal subtle changes in tertiary interactions. Thus, the effects of denaturant on the tryptophan signal may be observed from an emission spectrum of native M-TW. The emission spectrum of native protein is shown (Figure 3.2, blue) where the emission maximum, lambda max, was determined and labeled. Additional spectra were collected at a different denaturant concentration and on varying timescales in order to assess changes in the lambda max. For the M-WT incubated in 0.5M guanidine hydrochloride denaturation buffer, lambda max of the native protein is 333nm. The emission maximums at 1 day, 5 days and 11 days of denaturant incubation of the M-WT are 333nm, 336nm and 335 nm respectively. Similarly, the emission maximum of M-WT shifted is monitored by raising the

concentration of the denaturant of 1.5M guanidine hydrochloride. The emission maximum of 1 day, 5 days and 11 days of the incubated protein are 338 nm, 339 nm and 338 nm respectively. Examination of the lambda max at two different, but low concentrations of denaturant may reveal conformational changes in the native state prior to unfolding or untying (REF-Ervin, 2002). For some proteins, a hyperfluorescent intensity maxima may be observed during protein unfolding, indicative of a conformational state that may precede unfolding, or here, untying. For M-WT, the change in the lambda max at low denaturant is observed (Figure 3.2A). According to the overlay of the time-dependent emission spectra (Figure 3.2A), there is a distinct hyperfluorescent change over time, where an observed shift in lambda max and increase in signal intensity is observed over time. Interestingly, while there is an observed decrease in relative signal intensity, a greater change, but different trend is observed with more equilibration at a higher concentration of denaturant. While there is a noticeable increase of shift in wavelength at 1.5M denaturant, and the fluorescence signal intensity is more significantly quenched, a hyperfluorescent trend is still observed (Figure 3.2B) . These preliminary indications reveal a unique trend that suggests that changes in the tertiary interactions of M-TW are slow and differences in the solvent exposure of the tryptophan are apparent. These differences may be important in both the initial unfolding of the protein and untying of the knot, over time.

Concurrently, similar experiments were conducted on T-WT, although the excitation and emission data reflect observations following the fluorescent characteristics of tyrosine, as T-WT does not contain a tryptophan. Here, emission spectra were collected at the denaturant concentration of 3.2M and 4.0M over a similar timeframe. These equilibration points were selected based on previous observations indicating the presence of hysteresis at these points (Capraro, 2016). The hysteretic gap may be attributed to differences in tertiary interactions within T-WT, although only data highlighting the secondary structure has thus far been analyzed

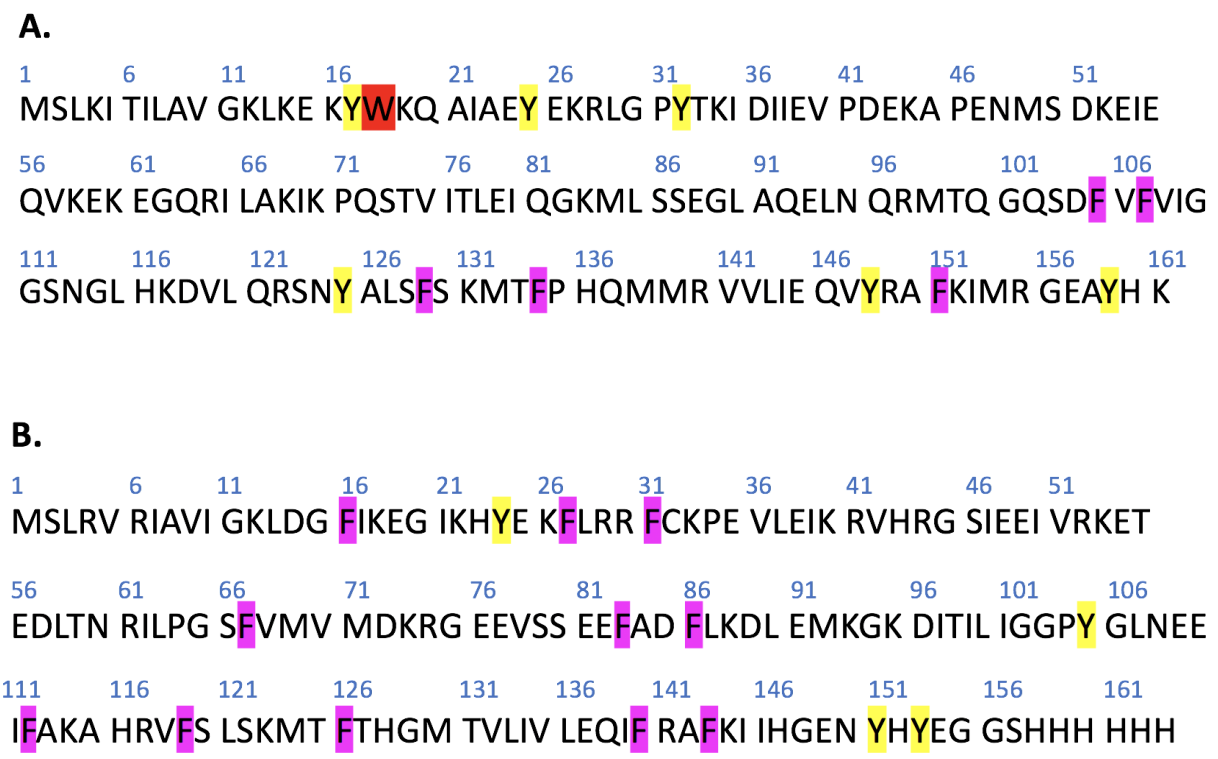


Figure 3.1

Sequence of MTT_{Sa} (Figure 3.1A) and MTT_{Tm} (Figure 3.1B) with highlighted intrinsic fluorophore: tryptophan (W) in red, tyrosine (Y) in yellow, and phenylalanine (F) in purple. The intrinsic fluorophore that was followed for MTT_{Sa} (M-WT) was W18. For MTT_{Tm} (T-WT) the intrinsic fluorophores that were followed were: Y24, Y105, Y151, and Y153.

and presented (Capraro, 2016). Initial data collected investigating tertiary and/or conformational differences included equilibrated denaturant samples over a 1 day, 5 days, and 11 days timeframe using fluorescence spectroscopy based on the tyrosine emission signal. Figure 3.3 highlights the emission spectra of T-WT in 3.2M and 4.0M guanidine hydrochloride denaturant for 1 day, 5 days and 11 days. Interestingly, the spectra reveal a similar red-shift in the spectra over dilution and time, as well as a hyperfluorescent trend (Figure 3.3). Initially, an emission maximum for native T-WT was observed at 342 nm. The lambda max T-WT at 3.2M denaturant over time, 1 day, 5 days and 11 days, shifts to 354nm, 354nm and 355 nm respectively. Similarly, lambda max at 4.0M denaturant shifts to 347 nm, 356 nm and 356 nm respectively. Additionally, unlike M-WT, the emission spectra of T-WT reveal an additional shouldering in the data between 305 and 310 nm that may correlate to changes in the solvent exposure of the available tyrosine. Given the unique trend in the emission spectra of T-WT, further investigation is required to better understand the relationship of the tyrosine interactions with the overall fold and the transition to unfolding of the protein. The initial indication is that the shifting and shouldering observed in the emission spectra may be contributions of the exposure of tyrosine residues to the buffer and the appearance of the phenylalanine signals (there are 4 tyrosines and 11 phenylalanines) during the initial unfolding process of the protein.

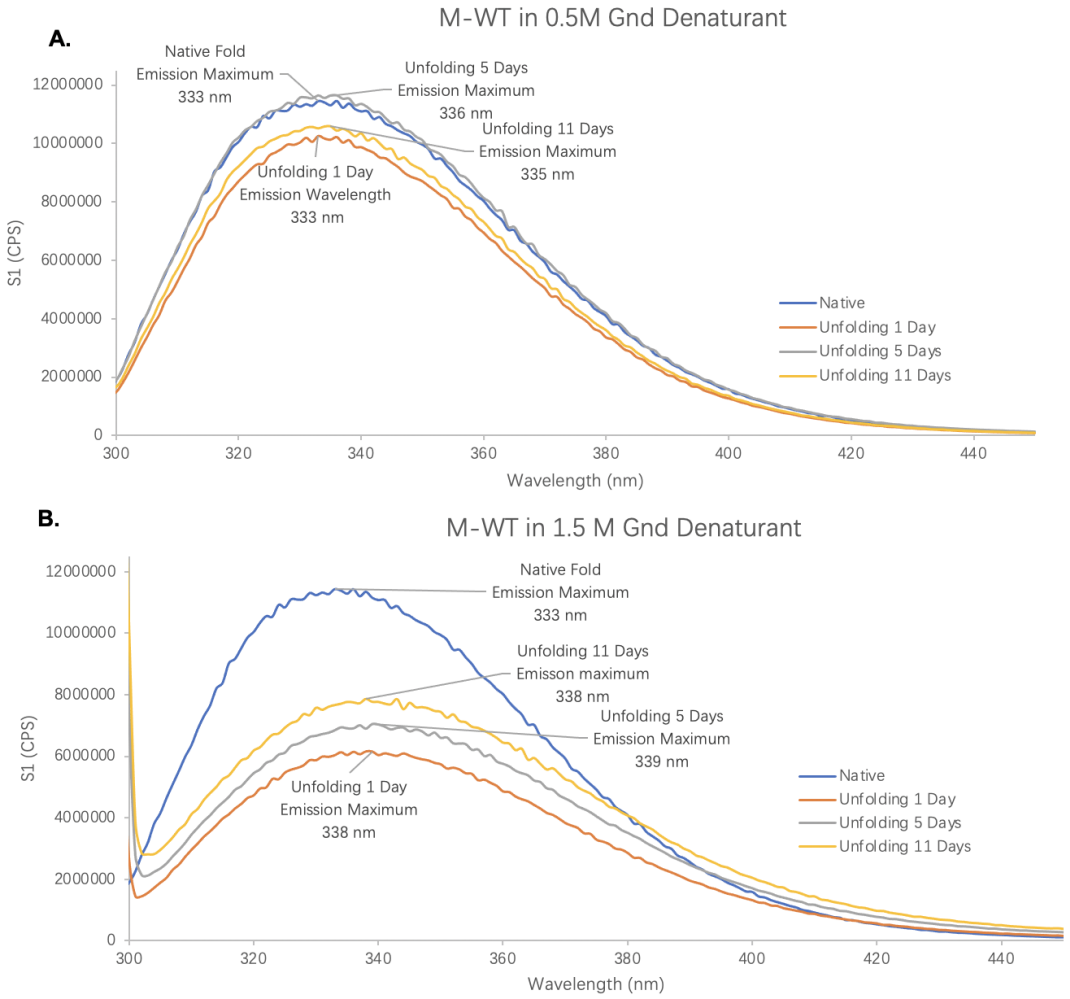


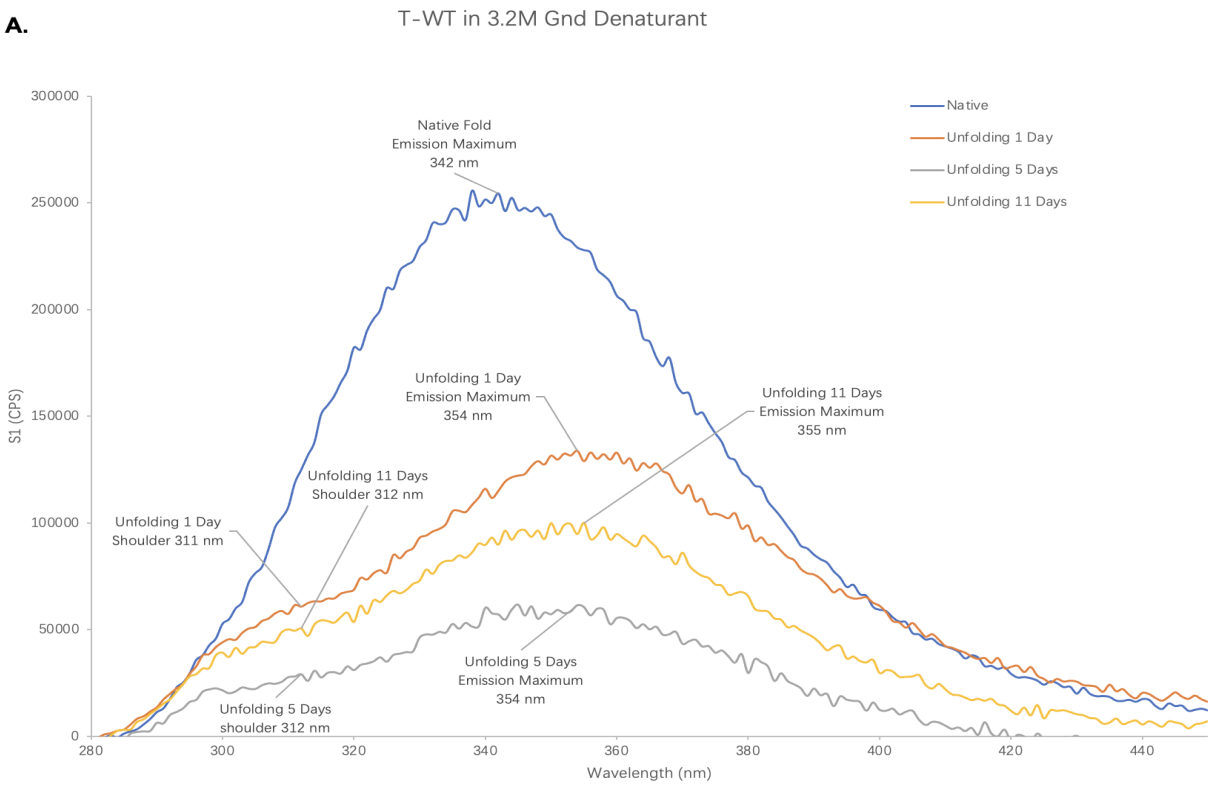
Figure 3.2 A

Emission spectra of M-WT monitored fluorescence spectroscopy of tryptophan excitation. Native M-WT protein (blue) was incubated in (A) 0.5M guanidinium hydrochloride buffer for 1 day (orange), 5 days (gray), and 11 days (yellow). (B) The emission spectra of the native protein and the 1.5M guanidinium hydrochloride buffer were labeled in the following the same coloring scheme.

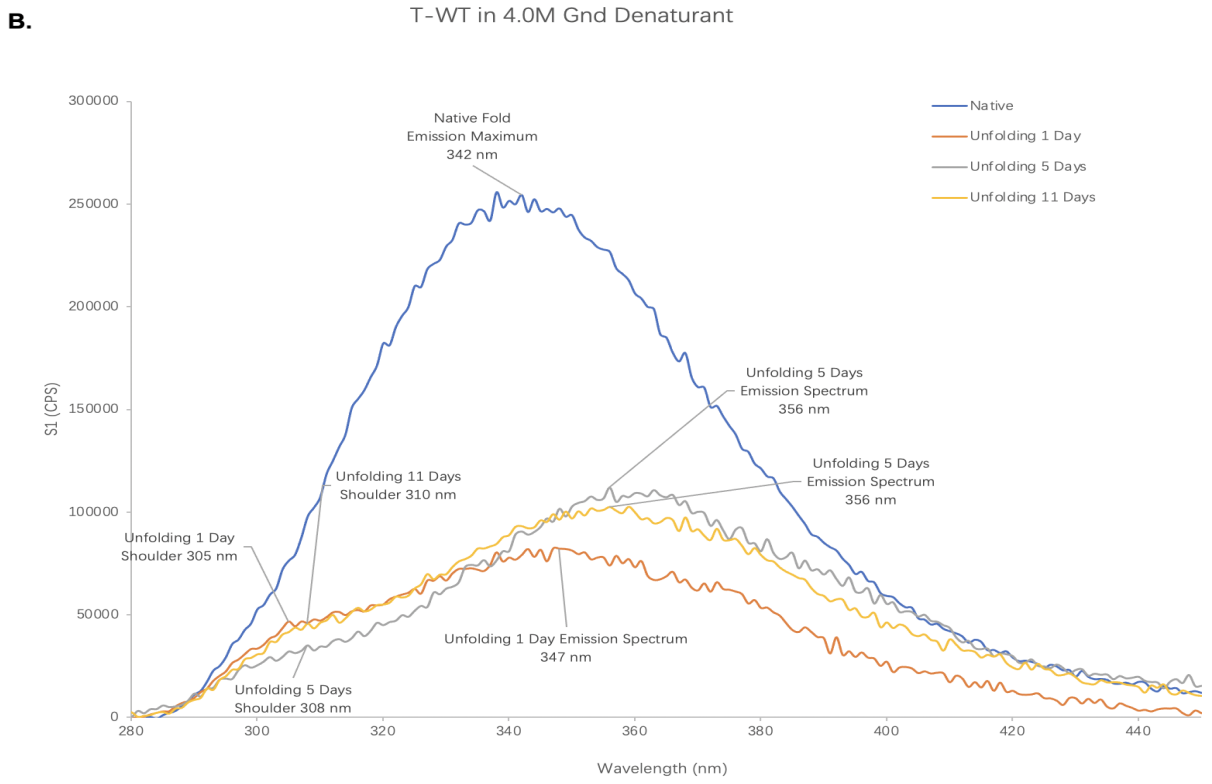
Figure 3.3 A

The emission spectra of T-WT are monitored by fluorescence spectroscopy of tyrosine excitation m. Native T-WT (blue) protein was incubated in (A) 3.2M guanidinium hydrochloride n buffer for 1 day (orange), 5 days (gray) and 11 days (yellow). The lambda max of the emission spectra is indicated. (B) T-WT incubated in 4.0M guanidinium hydrochloride buffer following the same timescale and coloring scheme as (a). Lambda max for each spectrum is indicated.

A.


T-WT in 3.2M Gnd Denaturant
Native
Unfolding 1 Day
Unfolding 5 Days
Unfolding 11 Days
Native Fold
Emission Maximum
342 nm
Unfolding 1 Day
Emission Maximum
354 nm
Unfolding 11 Days
Emission Maximum
355 nm
Unfolding 11 Days
Shoulder 312 nm
Unfolding 1 Day
Shoulder 311 nm
Unfolding 5 Days
Emission Maximum
354 nm
Unfolding 5 Days
shoulder 312 nm
S1 (CPS)
Wavelength (nm)

B.


T-WT in 4.0M Gnd Denaturant
Native
Unfolding 1 Day
Unfolding 5 Days
Unfolding 11 Days
Native Fold
Emission Maximum
342 nm
Unfolding 5 Days
Emission Spectrum
356 nm
Unfolding 5 Days
Emission Spectrum
356 nm
Unfolding 11 Days
Shoulder 310 nm
Unfolding 1 Day
Shoulder 305 nm
Unfolding 1 Day Emission Spectrum
347 nm
Unfolding 5 Days
Shoulder 308 nm
S1 (CPS)
Wavelength (nm)

Chapter 4

Monitoring Conformational Change of the Knotted Methyltransferase's Unfolding Process by NMR Spectroscopy

Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy is a commonly used technique to identify and analyze the structure and dynamics of biomolecular compounds. It is a powerful tool to investigate the molecular motion of proteins in solution on a multidimensional scale (Rule and Hitchens 2010). The general theory of NMR spectroscopy is the utilization of different magnetic spins of the various nuclei in order to identify atomic interactions that may relate to bonding interactions of spatial proximity in the protein (Howard 1998). Different from X-ray crystallography, which is also a protein structural determination approach, NMR samples are prepared in solution and are faster in terms of data acquisition than crystalized samples. The increasing size of the biochemical compound introduces complexity to spectral analysis due to the numbers of magnetically active nuclei, therefore the size of a protein or protein complex is a limiting factor in protein NMR spectroscopy. With the development and optimization of relaxation techniques and selective labeling strategies of amino acids within protein domains has broadened the size range that is capable of being analyzed, increasing to nearly 100 kDa (Yu 1999).

The principle of NMR spectroscopy comes from the odd mass number of atomic nuclei that are able to rotate around a given axis. The rotation of the charged nuclei creates a magnetic field which is visualized in Figure 4.1. In NMR spectroscopy, the spinning charged nuclei is able to create a magnetic field vector, The derivative of the nuclear magnetic moment vector produces a spin angular moment vector, dμ, which is proportional to the nuclear magnetic moment (Waterman 1997). When an external magnetic field, B, is applied to the nuclear magnetic moment, the spinning nucleus will precess around the axis of the external magnetic field. This phenomenon is called Larmor precession. The precession is presented by angular frequency, which is proportional to the external magnetic field given by the equation, $\varpi = -\gamma B$, in which refers to the gyromagnetic ratio of the nucleus (Tekely et al. 2002). The

Larmor precession is observable by bringing the external magnetic field perpendicular to the nuclear magnetization vector. The molecular electrons in this field will produce local currents, which will also produce an alternative field. The external magnetic field can be reduced by this alternative field, and this phenomenon is known as chemical shift, which can be used to identify the structural property at the atomic level (Silverstein et al. 2014). However, not all charged nuclei are suitable for NMR spectroscopy. For example, ^{12}C, ^{14}N and ^{16}O are not active in NMR spectroscopy due to the even mass number of atomic nuclei, so the nucleus is not able to spin and create the magnetic moment. Several commonly used isotopes that can be used to label proteins that trigger magnetic spin are ^{1}H, ^{15}N, ^{13}C and ^{31}P (Howard 1998). Due to this property, protein samples are required to be isotopically labeled, which is easily achieved during the fermentation process of protein overexpression and subsequent purification and isolation steps.

A chemical shift can be used as a reference signal to identify the nuclei and characterize the environment of a given residue. Fourier transform is a mathematical transformation that converts the spatial and temporal function into spectrum signals in spectroscopy and is reflected as a frequency (Bailey and Swarztrauber 1994). The unit of chemical shift in NMR spectroscopy is measured in hertz (Hz). The difference in a chemical shift is divided by Larmor frequency due to the difference in frequency of the external magnetic field. Thus, the frequency of chemical shift difference is much smaller than the external magnetic field frequency and is therefore usually presented by the unit of parts per million (ppm), where the frequency in the unit of Hz is multiplied by one million. For instance, if the two signals in 1H spectrum are 8000 Hz apart and the Larmor frequency is 800 million Hz, the chemical shift difference is determined to be (8000 Hz / 800 MHz)106 ppm = 10 ppm. Visually, an NMR spectrum is a left oriented axis, in which the frequency measured in ppm is increasing from right to left. Figure 4.2 provides a description of the different frequency range the various nuclei within an amino acid may appear

in a given1H spectrum. Specifically, alkyl groups on the side chain of amino acids have a relatively low frequency and low peak intensity. The central carbon signal, Calpha, has a larger frequency than the carbons on the side chain. The imine and amide groups in amino acids have the highest frequency signal but relatively low peak intensity. Aromatic side chains have a lower frequency, but higher intensity compared to the amide frequency. Both the alpha proton, HA, and the water signals have the greatest intensity, and are located approximately at the midpoint of the spectrum. Generally, a combination of experiments yielding characteristic signals for each of the amino acids in the polypeptide chain are used to identify, assign, and generate an NMR solution structure. While chemical shift frequencies are affected largely by the local environment of each residue, such as whether the amino acid is making a hydrogen bond interaction or packed up against an aromatic ring, complicating backbone assignments, the advantage of acquiring atomic details for a given protein outweigh these potential complications.

In this chapter, protein NMR spectroscopy was used to acquire data for M-WT. Specifically, a two-dimensional 1H-15N heteronuclear single quantum coherence (HSQC) spectrum, with one axis of 1H and another axis of 15N, was collected to determine the chemical shift dispersion of M-WT. In this particular experiment, an additional time pulse was introduced creating a variable in time (t) systematically between two pulses. In this way, the 15N nuclei signal can be observed additionally by combining with 1H signal to form a two-dimensional HSQC spectrum. The principle of this spectrum is based on the magnetization transfer from the proton (1H) to the heteronuclear (15N) through the J-coupling. J-coupling, also referred to as spin-spin coupling, is a process where two nuclei perturb the valence electrons respectively generating an observed splitting in peaks in the NMR spectrum (Cavanagh 2006). Significantly, an HSQC represents a unique chemical shift dispersion, much like a fingerprint, for a given protein.

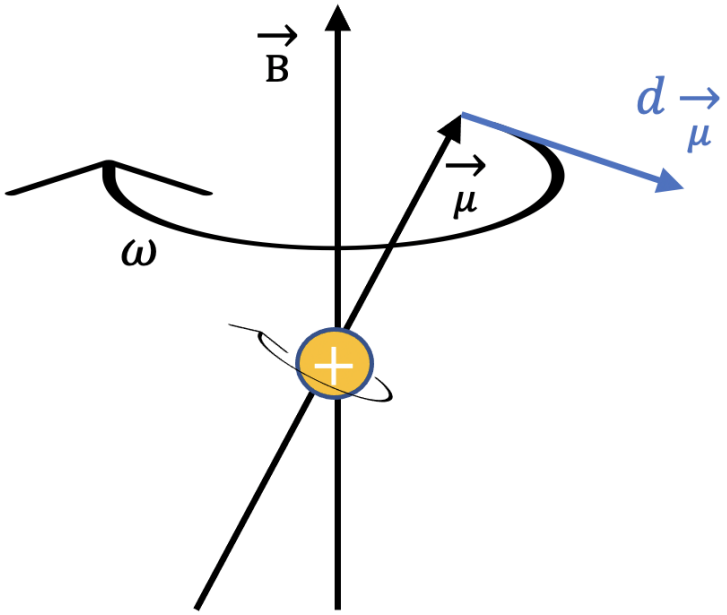


Figure 4.1

Cartoon visualization about the principle of NMR spectroscopy that the spinning charged nucleus creates a magnetic field. In this figure, the yellow circle with "+" represents the charged nucleus that undergoes a spinning motion. B represents the external magnetic field that is perpendicular to the nuclear magnetization vector. represents the nuclear magnetic moment vector that characterizes the spin of the charged nuclei, and with blue arrow represents the spin angular momentum vector that is proportional to. The black circular arrows represent the Larmor precession and represents the angular frequency of the precession.

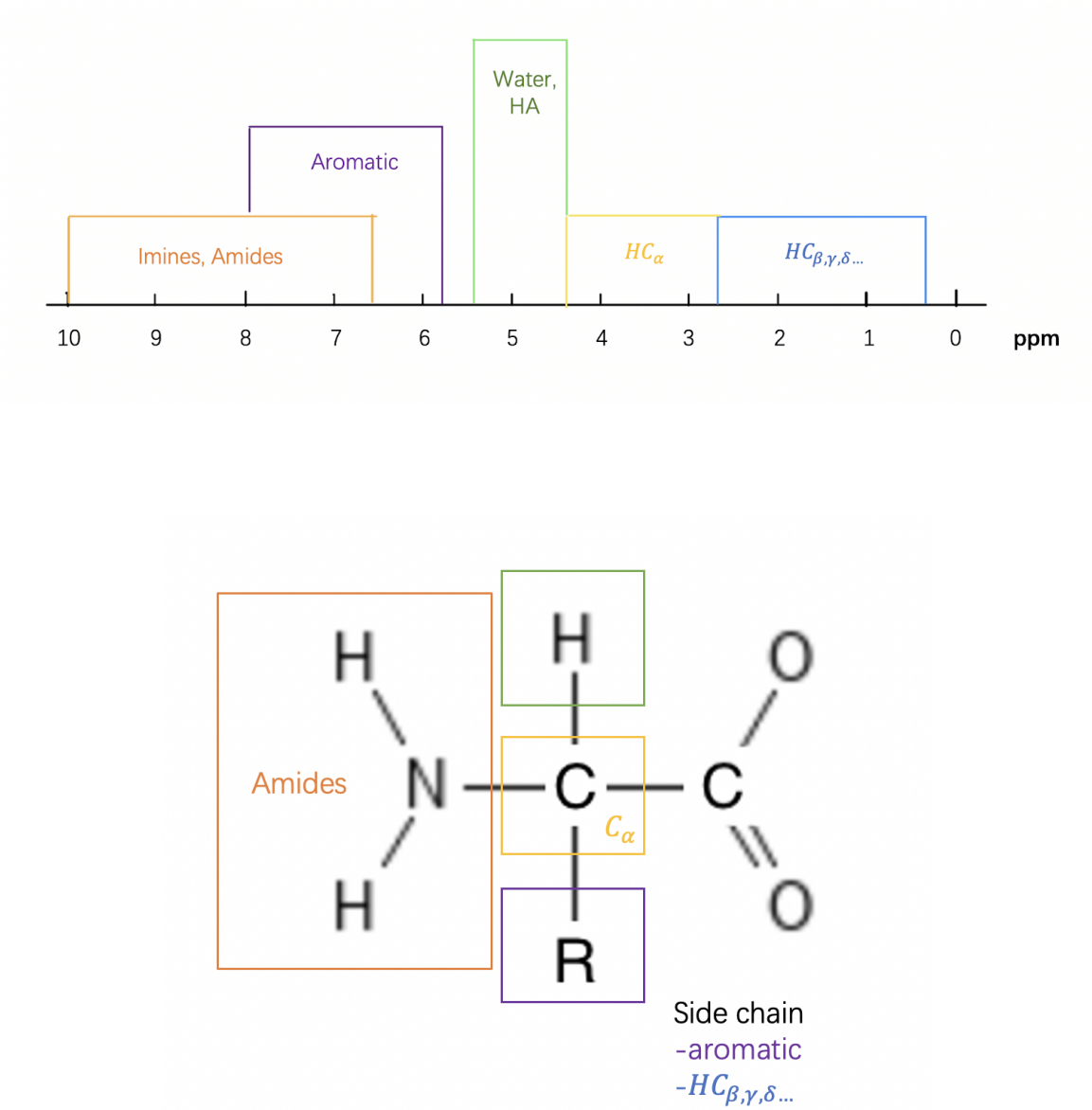


Figure 4.2

1H spectrum of the nucleus signal approximate peaks in amino acids. The nucleus signals in amino acids are presented with the frequency range and approximate peak intensity. Blue represents the alkyl signal $HC_{\beta,\gamma,\delta...}$on the side chain. Yellow represents the central C_α signal. Green represents the water or the HA group. Purple represents the aromatic signal on the sidechain of amino acids. Orange represents the amides signal in amino acids. Each group is presented and circles with respective colors in the chemical structure diagram below the spectrum.

Results

Acquisition of a 1H-15N HSQC is an effective first step at determining the overall fold of a given protein. Chemical shift dispersion can indicate the quality of the overall fold, unique interactions, and potential aggregation. Initial indications of changes in the overall structure can also be assessed through comparison of the changes in the fingerprint pattern in different solution conditions. Here, we probe chemical shift changes by comparing a folded and unfolded spectrum of M-TW. The experiments were collected at 305K in similar solutions with the unfolded sample containing 7M urea. The overlay of folded (blue) and unfolded (red) M-WT is presented in figure 4.3. Comparatively, the chemical shift dispersion is significantly different. Generally, the spectrum of native M-TW has well resolved peaks that are well dispersed indicative of a folded protein. This chemical shift dispersion is due in part to ring currents and dipole interactions between amino acids (Yao et al. 1997). The unfolded spectrum (red) revealed a noticeable difference in chemical shift dispersion where many of the peaks are clustered in the middle, consistent with amino acids that are not making additional interactions or a protein that is generally unfolded. Specific interactions standout in the 1H-15N HSQC overlay. There is a significant shift in the spectra where the resonance peak appearing at 10.51ppm (a probable tryptophan) has shifted in the unfolded state to 10.17ppm. As there is a lone tryptophan in M-WT (W18) upfield shift may result from the loss of hydrogen bonding due to decrease in deshielding because of the unfolded backbone. These lost intramolecular hydrogen bonds may include Y17 and M138 (proximal residues in the 1vh0 structure) as the protein is unfolded.

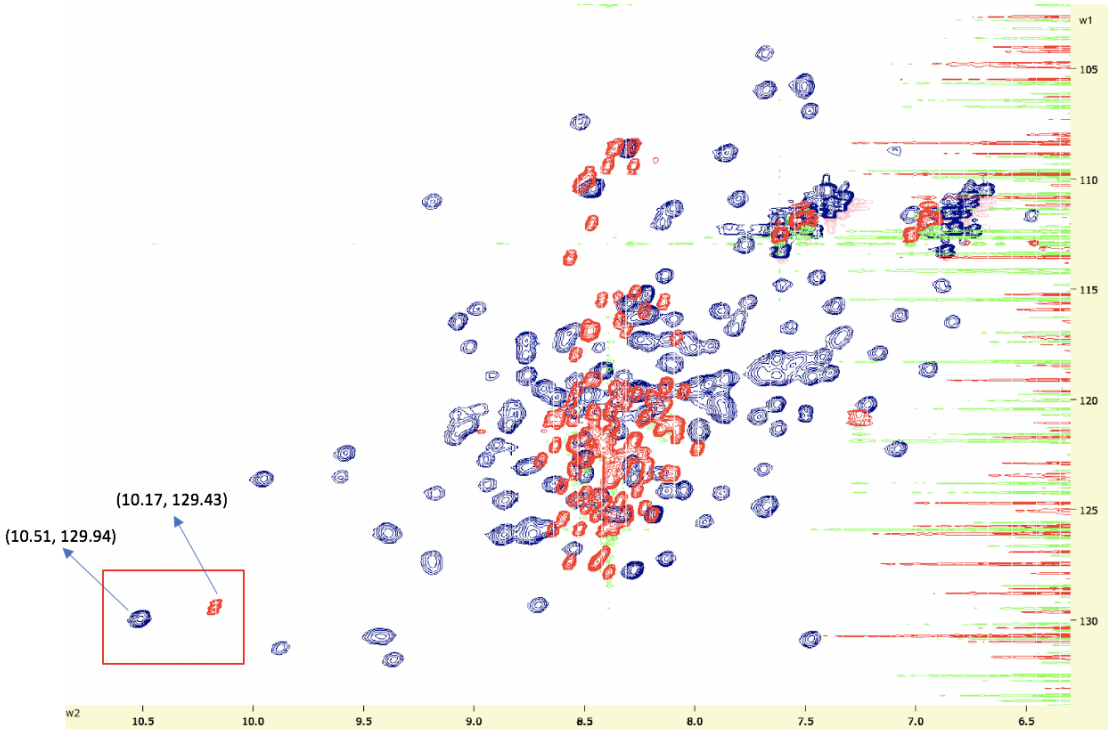


Figure 4.3

Overlay of the ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum of native M-WT(black) and unfolded M-WT in 7.0M urea denaturant (red). The location and shift of the possible tryptophan residue is boxed, and the respected shifts indicated.

Chapter 5

Discussion and Future Direction

Discussion

The protein expression, purification and subsequent preliminary experiments of T-WT were critical in generating foundational information regarding the overall fold and solution dynamics for the unique class of knotted proteins. The initial purification protocol was modified and optimized in order to maximize the yield of the protein of interest for concurrent downstream experiments. The resulting isolated protein was analyzed by mass spectroscopy indicating the presence of a single peak shown at 18813 Da, consistent with the theoretical values. This result verified the protein of interest in the subsequent experiments

The initial unfolding experiments of SPOUT MTase using fluorescence spectroscopy which provided preliminary data based on the emission spectrum signal of tryptophan (M-WT) and tyrosine (T-WT) were informative and interesting. For M-WT, the protein samples that were incubated in low and moderate levels of denaturant (0.5M and 1.5M guanidine hydrochloride) over a lengthy period of time (5 days and 11 days) revealed changes in the lambda max. These differences suggest that while the unfolding of the protein resulted in fluorophore quenching, consistent with loss of static folded interactions, the hyper fluorescence observed may correlate to the knot movements as the protein loosens the route to unfolding. With an increase in the population of unfolded protein (1.5M), a trend is still observed, although it is less pronounced. Given that the emission maximum shift to larger wavelength is a possible response for the hydrogen in tryptophan exposure or bonding to the water in the buffer solution (Lakowicz 2017) because of unfolding, more experiments are needed to assess the use of emission spectral analysis as a knot signature signal.

Similarly in the unfolding experiments following the tyrosine emission signal (T-WT) revealed multiple trends in the spectra over time. For T-WT, the change in lambda max is

consistent with previous results using circular dichroism (CD). While CD monitors the secondary structural changes in a protein, where the helical signal is most pronounced when it is fully formed, over time and in the denaturant, that signal is altered (Capraro and Jennings 2016). Since the emission maximum shift was observed to be increasing at the middle range (2.5M to 4.5M) of denaturation buffer, the emission spectrums of T-WT incubating in 3.2M and 4.0M guanidine hydrochloride denaturation buffer were obtained. A special trend of splitting the emission maximum peak to an approximate 355 nm peak and an approximate 310 nm shoulder after the native samples were incubated in the denaturant for 11 days. However, the reasons for this phenomenon of splitting peaks are complicated due to the multiple tyrosines in the sequence of T-WT and might also reflect the signal of phenylalanine showing up on the emission spectrum. Another unusual observation for the T-WT emission spectrum is from the emission maximum peak of the native protein. Normally speaking, tyrosine is excited at around 275 nm and emitted at around 304 nm (Yang et al. 2017). However, our observed native T-WT with several tyrosines but absence of tryptophan emission spectrum showed the emission maximum at 342 nm, which was easily confused with tryptophan fluorescence signal. One possible reason for this special phenomenon is the tyrosinate emission at around 345 nm in the protein that lacks tryptophan residues in neutral solution (Szabo et al. 1978). This guess requires further solidations and evidence since tyrosine is able to be ionized to tyrosinate in neutral and weak basic solutions (Schnarr and Helene 1982). However, the pH of our T-WT samples is 5.6, which conflicts with the guess of excited-state ionization of tyrosine to tyrosinate. Therefore, the reason for the emission maxima of the protein samples with tyrosine but lack tryptophan requires further investigation.

The two dimensional ^{1}H-^{15}N HSQC spectrum provided evidence in investigating the native and unfolded M-WT residue fluctuations. The unfolded M-WT HSQC spectrum is observed to have a narrower chemical shift dispersion in comparison to the native protein. It

could be explained by the fact of losing secondary structure. In addition, W18 is observed to have the most significant change in chemical shift when the protein is unfolded. This observation could indicate the change in surrounding environment including the decrease of electronegativity and hydrogen bond effect of W18 with the nearby residues, including Y17 and M138. However, further calculations and investigation should be done, including the backbone assignment and calculations of structural parameters, including orientation, intramolecular distance, and coupling constants (Chatterjee et al. 2005).

The work presented in this thesis mainly focuses on obtaining the preliminary data of the knotted SPOUT MTase during the unfolding process. The purification protocol was modified and analyzed by mass spectroscopy to assess the purity of the purification results. Several approaches including the fluorescence spectroscopy and NMR spectroscopy were applied to characterize the unfolding process as the native protein samples were incubated in the denaturation buffer over time.

Future Directions

Given that the NMR spectrums of M-WT in native, partially unfolded, and unfolded states. It would be significant to monitor the real-time unfolding process of the knotted protein. Therefore, a kinetic spectrum was obtained to monitor the time for the initial unfolding step of the M-WT. Figure 5.1 illustrates the initial unfolding kinetics of the native M-WT in 3.5M and 4.0M Urea denaturation buffer. From the kinetic spectrum, it could be observed that the decrease in fluorescence signal intensity in the first 200 seconds was the greatest, representing the drastic unfolding event in the native knotted protein. However, the signal intensity was still having a decreasing trend when the native protein was in the 3.5M urea denaturation buffer

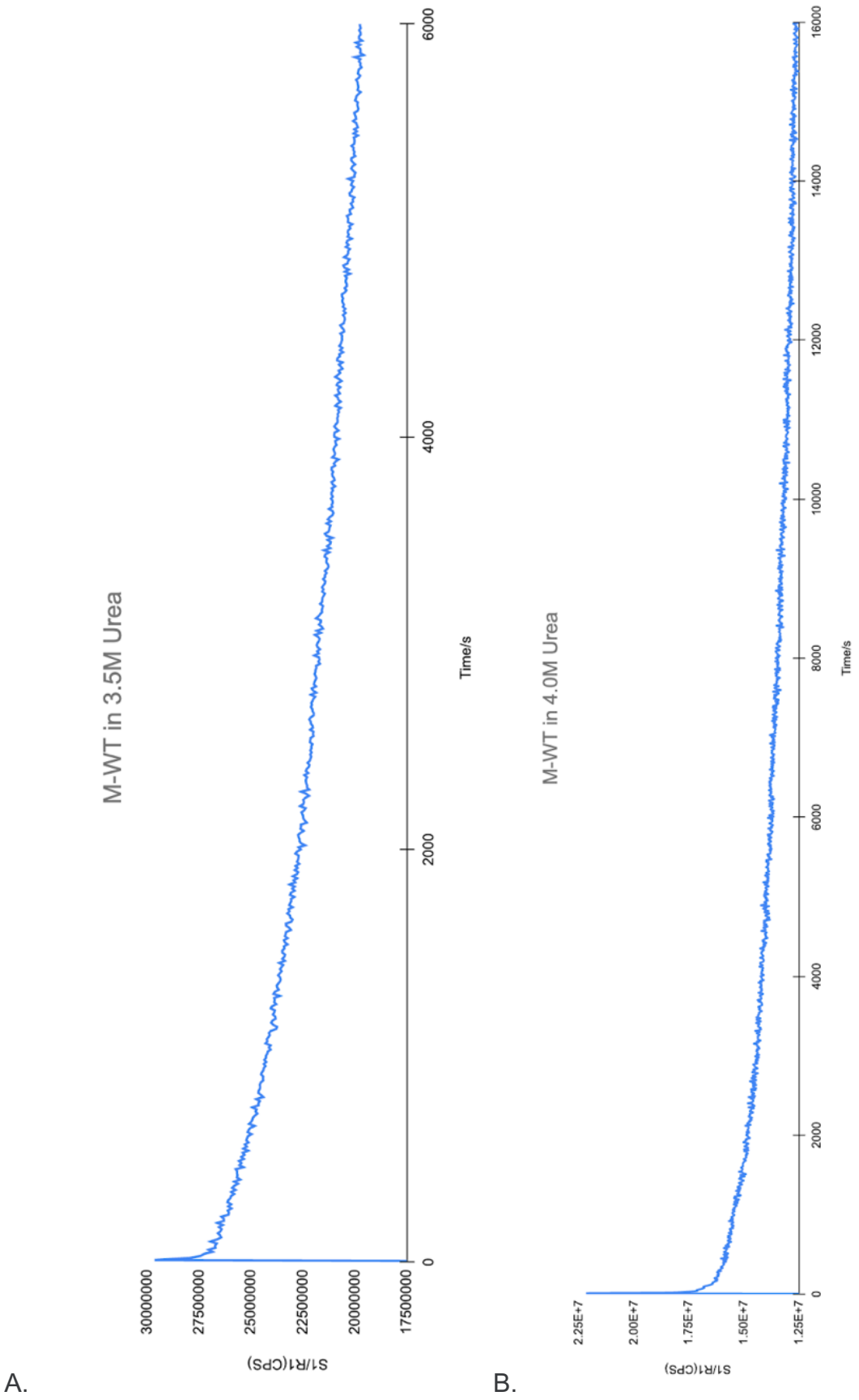
after 6000s. Thus, another kinetics unfolding experiment in 4.0M urea denaturation buffer was set up in the same method but increased the time frame to 16,000 seconds. The decrease in fluorescence signal intensity of native M-WT in 4.0M urea denaturation buffer was determined to be 1.46-fold than the 3.5M trial, indicating that the native knotted M-WT is unfolding faster in 4.0M urea. Similarly, the decrease in fluorescence signal intensity slowed down after 12,000 seconds, but we could still observe the minor decrease in the intensity at 16,000 seconds. Given the results from the emission spectrum experiment in Chapter 3 of this thesis, the emission maximum shift was observed at 5 days incubation in the 1.5M guanidine denaturation buffer. The unfolding process of M-WT was expected to be long given that the trefoil knot provides extra stability to the protein and makes the unfolding process longer. The kinetic unfolding experiment is still able to provide us a preliminary idea of the approximate time frame for us to observe the signal shift in real-time NMR spectroscopy in order to capture the signal change during the initial unfolding process at the atomic level.

In order to create an unfolding profile for the knotted T-WT, protein samples were incubated in various concentrations of guanidine hydrochloride denaturation buffer from 0M to 6.0M at room temperature for 7 days and 14 days. The emission spectrums of each sample were obtained by tracing the tyrosine signal. The emission maximums of protein in various denaturant concentrations were determined and plotted in Figure 5.2. From the results shown in this figure, there is no significant difference of the unfolding profile between 7 days and 14 days incubation in guanidine hydrochloride. The previous unfolding data monitored by circular dichroism spectroscopy indicated a curve with a more significant trend showing denaturation since it reflects the signal change of diminishing secondary structure (Carpraro and Jennings 2016). In the case of fluorescence spectroscopy, the instrument monitored the protein's tertiary fold by tracing the tyrosine signal. Therefore, there might still be some interactions between the

Figure 5.1

(A) illustrates the kinetics spectrum of the unfolding process of native M-WT. Urea denaturation buffer stock was added to the native protein to make the final concentration of urea 3.5M, and the kinetic spectrum was obtained immediately after the sample was mixed with the urea denaturation buffer. This figure shows the kinetics for the first 6000 seconds of the unfolding protein by observing the fluorescence intensity decrease.

(B) was obtained using the same method. The native protein was mixed with the urea denaturation buffer stock to make the final concentration of urea 4.0M. The kinetic spectrum that monitors the unfolding process was obtained for the first 16,000 seconds.


A.
M-WT in 3.5M Urea
S1/R1(CPS)
30000000
27500000
25000000
22500000
20000000
17500000
0
2000
4000
6000
Time/s
B.
M-WT in 4.0M Urea
S1/R1(CPS)
2.25E+7
2.00E+7
1.75E+7
1.50E+7
1.25E+7
0
2000
4000
6000
8000
10000
12000
14000
16000
Time/s

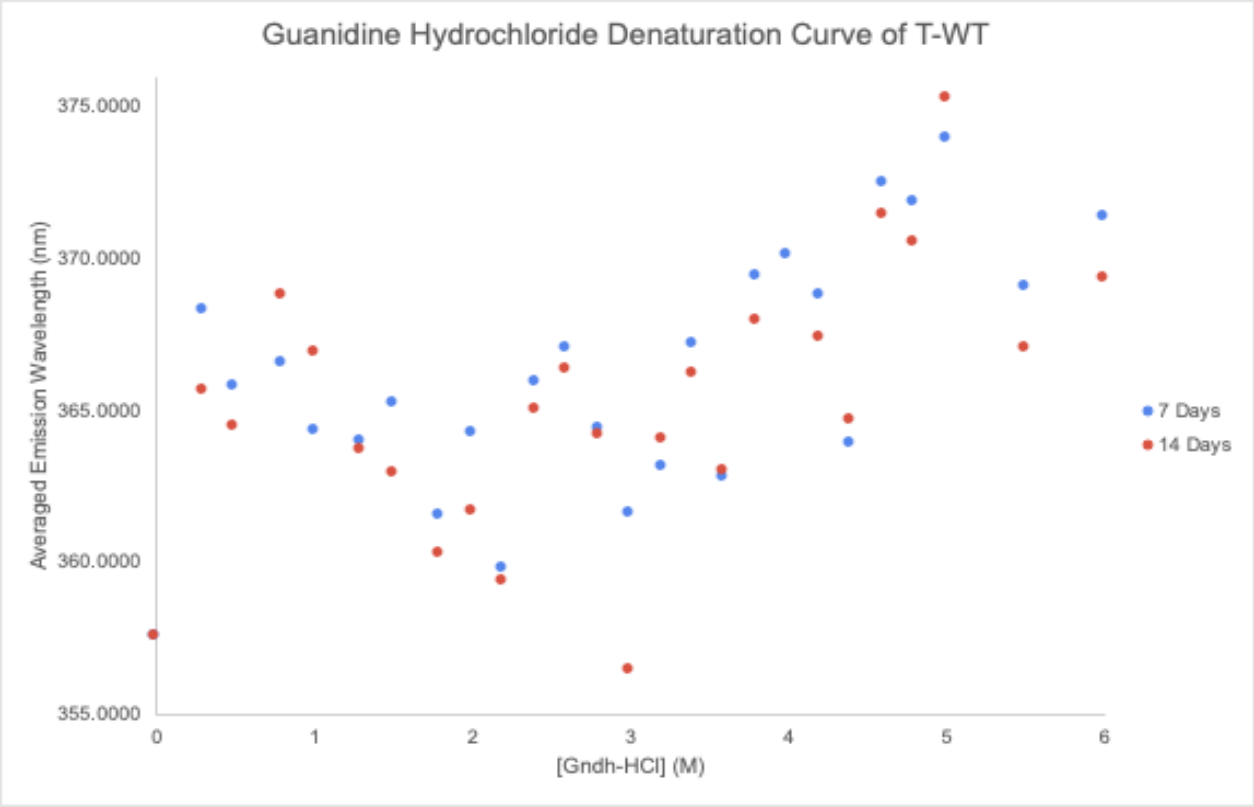


Figure 5.2

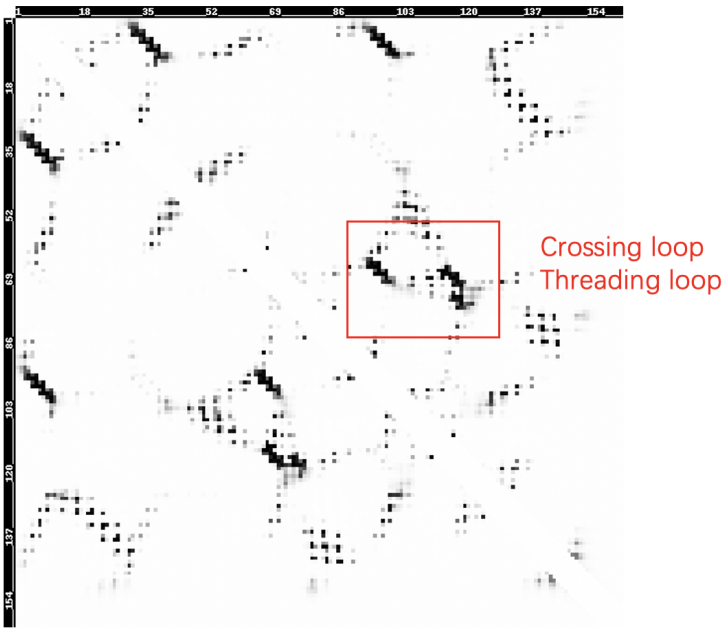
Guanidine hydrochloride denaturation curve was plotted by averaging emission maximum wavelength of T-WT in various concentrations of guanidine hydrochloride denaturation buffer ranges from 0.0M to 6.0M, with x-axis showing guanidine hydrochloride concentration and y-axis showing averaged emission maximum wavelength calibrated by buffer respectively. Protein samples were incubated in the denaturant at room temperature for 7 days (blue) and 14 days (red). The fitted curves were presented by the dash lines with identical colors.

environment and the tyrosine residues during the unfolding step of MTT_{Tm}, leaving more investigations to be done in terms of the tertiary fold and the microenvironment of the protein.

Another future direction of this work could be introducing the computational modeling tools as a supporting approach to investigate the structural and functional effect before and after the ligand (SAM) binds to the SPOUT MTases protein, in which the ligand binding site is very close to the knotted region. In order to be familiar with the principle and function of computational simulations, RaptorX developed by Xu's group was used to predict the possible conformations for the native MTT_{TM} by generalizing a contact map and probability matrix from the Multiple sequence alignment (MSA) result (Xu et al. 2017). The contact was defined by Cb-Cb distance, that is ≤ 8Å, and the contact probability was colored by a different scale of grey: the darker color represents the higher probability of contact between two residues. Figure 5.3 (A) shows the contact map of the first model that RaptorX generated. The knotted region including crossing loop and threading loop is circled by a red box. It shows a relatively high probability of contact in the knotted region, and a unique turn in the helix, which is commonly found in knotted protein. RaptorX provided us in total five models based on the inter-residue distance and PDB sequence, and they are shown in Figure 5.3 (B). What can be observed is that most of the regions in the model have little differences except for the c-terminus is very flexible.

RaptorX is able to provide a high-quality structure without the homologues in PDB. Therefore, it would be possible to simulate the knotted protein structure of the protein mutations by substituting different residues into Alanine, which would make the side chain smaller. The mutations that have the most conformational difference might indicate the residues that might contribute to the structural function of the knotted SPOUT MTase. Starting from this point, it

A.



B.

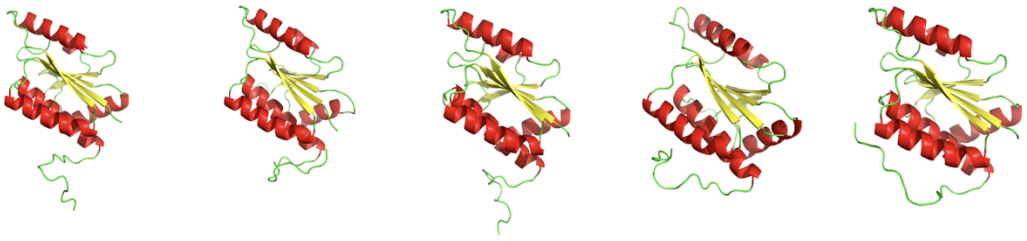


Figure 5.3

A. The contact map predicted by RaptorX produced from Xu's group for SPOUT MTT$_{TM}$. The red box indicates the location of the crossing loop and threading loop of the knot.

B. Cartoon representation of the top 5 models generated from the contact map and MSA probability matrix.

would be possible and more efficient to investigate differences among the wild-type and mutations through the experimental approach.

Molecular Operating Environment (MOE) would also be a useful tool to simulate the interactions and possible conformational change between the ligand and the protein. Therefore, it is possible to use the MOE docking visualization prediction of the binding ability of the ligand to the binding site in the aspect of ligand conformation and orientation. Additionally, MDAnalysis could be run with the results of MOE docking. This analysis is able to provide structural and temporal properties of MD simulation trajectories and individual protein structures (Michaud-Agrawal et al. 2011). In order to have a more detailed and accurate result, additional MD simulation of the unfolding mechanism is necessary to be done, and several factors for this approach discussed here are necessary to be considered. Firstly, we need to consider the high stability of the knotted core, which takes 20-fold slower to untie the knot in comparison to only unfolding it. Also, our previous research observed a long-range allosteric change in helices, which is consistent with tying and untying in helical propensity, when the knot is binding to the SAM ligand. Thus, we also need to consider the movement of surrounding domains and opening the hydrophobic core (Xu and Li 2018). The simulation results can be compared with the previous NMR data of the ligand binding experiment to assess the accuracy of the simulations.

Chapter 6

Material and Method

Protein Expression, Purification and Sample Preparation

General

Unlabeled Mesophile and Thermophile samples were expressed in Miller Luria-Bertani Broth (LB media) (Fisher BioReagents) by growing in the *E.coli* cells. Isotopically labeled Mesophile were expressed in Minimal Media that contains 2.0 g/L $AmSO_4$ (Fisher Chemical) and 3.0 g/L D-glucose (Cambridge Isotope Laboratories) that isotopically labeled ^{15}N and ^{13}C.

M-WT (MTT_{SA})

Unlabeled protein is expressed and purified in the following protocol. The wild type protein of mesophile (PDB code: 1Vh0) was transformed using BL21(DE3) Competent *E.coli* (Invitrogen). After transforming the M-WT genes into *E.coli*, the cells were grown in LB media with 50 mg/mL Kanamycin (Fisher BioReagents) as an antibiotic at 37℃ until OD_{600} reached 0.6-1.2. The protein expression was induced in 1mM IPTG (Teknova) and induced overnight (16-18 hours) at room temperature (25°C). Cell pellets were collected by centrifugation at 3000 rpm for 30 minutes using Beckman-JLA 8.1 (Beckman Coulter). The cell pellets were first frozen at -80℃ for at least 1 hour and then thawed in order to ultimately produce the lysing result. Next, cell pellets were resuspended using 10mL lysis buffer (50 mM $NaPO_4$ pH 8, 300 mM NaCl, 5 mM B-ME) per 1L of cell growth and sonicated on ice for total of 25 minutes (10 seconds on, 15 seconds off) to lyse the protein from cells. Supernatants that contain the protein are collected by centrifuging the lysate at 13000 rpm for 30 minutes using Beckman-JA 14.5 (Beckman Coulter). Protein purification is implemented using the method of manual purification of 6xHis-tag proteins from *E.coli* using Ni-NTA column chromatography (QIAGEN). Ni-NTA resin was equilibrated with 2 beds volume using the lysis buffer. Next, 4 beds volume of the cell lysate was loaded to the column and incubated with the resin on the shaker for 1 hour at 25℃. The supernatant was

collected by gravity flow and marked as flow-through (M-FT). The column was washed twice with 2.5 beds volume of wash buffer (50mM NaPO4 pH8, 300 mM NaCl, 5 mM BME, 10 mM Imidazole) and collected the fractions (M-W1 and M-W2). Purified protein was collected by eluting with 1 bed volume of elution buffer (150 mM NaOAc pH 5.6, 600 mM NaCl, 10 mM BME, 300 mM Imidazole) for 3 times. Fractions containing M-WT were pooled and labeled as M-E. Purity and amount were accessed by SDS-PAGE in order to confirm protein purification. M-E was concentrated to approximately 10mL for further purification using size exclusion column chromatography on FPLC (Bio-Rad). S200 size exclusion column was equilibrated using filtered and degassed SE buffer (75 mM NaOAc pH 5.6, 1% glycerol). Concentrated protein samples were injected and ran through the S200 column. Peaks shown on the FPLC trace with A280 absorbance were accessed for protein presence using SDS-PAGE gel, and approximate protein concentration were measured using ND-1000 Nanodrop Spectrophotometer. Fractions that were confirmed of protein presence were pooled and stored for next step analysis.

T-WT (MTT_{TM})

The wild type protein of thermophile (PDB code: 1O6D) is expressed and purified in the following protocol. DH5 competent cells are used to maximize the transformation efficiency of the genes followed by transformation to *E.coli* cells using BL21 (DE3). The initial growth started in LB media with 50mg/mL Kanamycin (Fisher BioReagents) as antibiotics at 37℃ until the OD_{600} reached 0.6-1.2. IPTG is added to the final concentration of 1mM to inducting overnight (16-18 hours) at room temperature (25℃). Cell pellets containing the protein were collected by centrifugation at 3000 rpm for 30 minutes using Beckman-JLA 8.1 (Beckman Coulter). The total cell pellets were resuspended by 1X lysis buffer (75mM NaOAc, pH 8.0, 150mM NaCl, 5mM BME), and the homogenous was sonicated on ice for total of 25 minutes, with 10 seconds on and

15 seconds off in order to lysed the protein out from the pellet and make it solvable to the buffer. Supernatant containing T-WT was obtained by centrifuging the homogenous after sonication at 13000 rpm for 40 minutes using Beckman-JA 14.5 (Beckman Coulter). 4 beds volume clear supernatant was then transferred to the equilibrated Ni-NTA column using at least 2 beds volume lysis buffer, and 6X His-tag protein was set to bind with the resin on shaker for 1 hour at room temperature. Flow through (FT) was collected by gravity flow. Column was washed twice with 2.5 beds volume 1X wash buffer (75mM NaOAc, pH 8.0, 150mM NaCl, 10mM imidazole, 5mM BME). Protein was eluted by 1 bed volume of 1X Elution buffer (75mM NaOAc, pH 8.0, 150mM NaCl, 150mM imidazole, 5mM BME) for three times, and each fraction of elutions were pooled together. Flow though(T-FT), Wash (T-W1 and T-W2), and Elution (T-E) samples were collected for SDS-PAGE analysis. After confirmation with SDS-PAGE, the T-E fraction was concentrated to ~10mL and loaded to the S200 size exclusion column on FPLC. The size exclusion column chromatography was run in the SE buffer (75mM NaOAc, pH 5.6, 1% Glycerol, 1mM DTT). Fractions were collected from the trace shown on the FPLC. Each peak was checked for protein presence on SDS-PAGE, and the protein concentration was measured through A280 using the Nanodrop microvolume spectrophotometer.

M-WT Labeling with ^{15}N

^{15}N labelled protein is prepared and purified in the following protocol for a 4 liters growth. Necessary glassware and 4L filtered DI water were autoclaved and cooled down to room temperature. Reagents were added to the DI water in order to prepare the Minimal Media (MM)

in following amount: 8.0 g $AmSO_4$ (Fisher Chemical), 26.8g Na_2PO_4 (Fisher Chemical), 12.0g KH_2PO_4 (CALBIOCHEM), 6.0g NaCl (Fisher Chemical), 6.0g D-glucose (Cambridge Isotope Laboratories), 400 10mM $FeCl_3$ (Sigma-Aldrich), 400 1M $CaCl_2$ (Spectrum), 400 10mM $ZnCl_2$ (Sigma-Aldrich), 2mL 1M $MgSO_4$ (Fischer Chemical), and 40mL Vitamin E (Sigma). The solution was verified to be pH around 7 and then was sterile filtered. The growth was started by 100 mL Minimal Media overnight per 2L growth with 50mg/mL Kanamycin as antibiotics. 100 mL MM overnights were transferred to the 2L flasks as the MM turned opaque. Sufficient amount of cell stock was suggested to start in order to reach optimal growth results. Isotopically labelled protein expression induced by IPTG (Teknova) at OD_{600} reached 0.500-0.700 at room temperature for 16-18 hours at 25. Cell pellets were collected by centrifugation at 3000 rpm for 30 minutes using Beckman-JLA 8.1 (Beckman Coulter), and then resuspended in 1x Lysis buffer (50 mM $NaPO_4$ pH 8, 300 mM NaCl, 5 mM B-ME). The soluble protein was then extracted by Sonic Dismembrator (Fisher Scientific) for 25 minutes on ice (10 seconds on and 15 seconds off) and collected by centrifuging the lysate at 13000 rpm for 30 minutes using Beckman-JA 14.5 (Beckman Coulter). Labeled M-WT protein was purified using the same method as unlabeled M-WT protein. The first run of purification was done using Ni-NTA column chromatography (QIAGEN), with 1x washing buffer (50mM NaPO4 pH8, 300 mM NaCl, 5 mM BME, 10 mM Imidazole) and 1x elution buffer (150 mM Na-Acetate pH 5.6, 600 mM NaCl, 10 mM BME, 300 mM Imidazole). Fractions obtained from Ni-NTA column chromatography and protein expression were analyzed by SDS-PAGE. The purified M-WT was concentrated to about 10 mL and was further purified by size exclusion column chromatography on FPLC (Bio-Rad) with 1x SE buffer (75 mM NaOAc pH 5.6, 1% glycerol). The A280 peaks were analyzed for protein existence and purity by SDS-PAGE. 1x SE buffer was then switched to isotopically labelled SE buffer (75mM NaOAc-d3, pH 5.6, 1% Glycerol) by PD-10 desalting column (GE Healthcare) using gravity protocol. The final protein concentration was measured using the ND-

1000 Nanodrop Spectrophotometer. The sample was concentrated to approximately 8 mg/mL and stored at 4℃.

Fluorescence Spectroscopy Experiment

Protein fluorescence spectroscopy experiment of M-WT and T-WT was obtained through FluoroMax-4 spectroscopy (HORIBA). The emission spectrum of M-WT was obtained by excitation of the tryptophan fluorophore at 295 nm with slit width 3nm and collecting the emission signal of tryptophan residue from 280 nm to 450 nm. The emission spectrum of T-WT was obtained by excitation of the tyrosine fluorophore at 275 nm, with slit width 5nm, and collecting the emission signal of tyrosine from 265 nm to 450 nm. Both native state and unfolded processes of the proteins were generated by denaturant incubation and were reflected through the emission spectrum within different time frames using the same methods.

Unfolding Emission Spectrum of M-WT

Six samples from size exclusion column chromatography that have been analyzed for the existence and purity from SDS-PAGE were concentrated to at least 0.30 mg/mL to do the fluorescence experiment. Concentrated M-WT samples were incubated in 0.5M guanidine hydrochloride denaturation buffer (0.5M guanidine hydrochloride, 75mM NaOAc, pH 5.6, 1% Glycerol) and 1.5M guanidine hydrochloride (Alfa Aesar) denaturation buffer (1.5M guanidine hydrochloride, 75mM NaOAc, pH 5.6, 1% Glycerol), and incubated at room temperature for 1 day, 5 days and 11 days. The emission spectrum of the native and unfolded protein was obtained using Fluoromax-4 Spectrophotometer (HORIBA). Protein samples were loaded in the 200 L cylindrical cuvette and excited at 295 nm with slit width 3nm. The emission wavelength signal was collected with the range from 280 nm to 450 nm, with a slit width 3 nm. The emission

spectrum data were calibrated by running the same emission spectra using the buffer and denaturant that protein samples are stored and incubated. The emission spectra signals of the protein were subtracted by the signals of the buffer to minimize the effects of the environment.

Unfolding Kinetics of M-WT

The unfolding process of M-WT is initiated by preparing the protein samples with a 6.0M Ultra pure Urea (VWR Life Science) denaturation buffer stock (6.0M Urea, 75mM NaOAc, pH 5.6, 1% Glycerol) to desired denaturant concentration. To check whether the protein is able to be unfolded in the 3.5M Urea denaturation buffer, emission spectra of unfolded protein using Fluoromax-4 Spectrophotometer (HORIBA) were obtained to see the emission maximum wavelength shift using the same method as the previous section of unfolding emission spectrum of M-WT (excitation at 295 nm, slit width 3nm, emission from 280 nm to 450 nm). The unfolding process in the timescale perspective was monitored using the kinetics measurements on the Fluoromax-4 Spectrophotometer (HORIBA). In order to capture the initial unfolding process, 6.0M denaturant buffer stock was added to the native protein samples in a 200 L cylindrical cuvette right after the kinetics experiment was initiated (excitation at 295 nm). The intensity of emission signal (S1) was collected in a 10 seconds time interval for a total of 6000 seconds, and the emission detector signal was corrected by reference quantum counter signal (R1) using S1/R1 correction. The processed data is able to be captured in FluorEssence ™ software on the PC connected to the Fluoromax-4 spectrophotometer (HORIBA).

Unfolding Emission Spectrum of T-WT

Five samples from size exclusion column chromatography were concentrated to at least 0.3 mg/mL in order to obtain an appropriate protein fluorescence signal. The protein samples are

incubated in 3.2 M Guanidine denaturation buffer (3.2M guanidine hydrochloride, 75mM NaOAc, pH 5.6, 1% Glycerol, 1mM DTT) and 4.0M Guanidine denaturation buffer (4.0M guanidine hydrochloride, 75mM NaOAc, pH 5.6, 1% Glycerol, 1mM DTT) for 1 day, 5 days and 11 days. The emission spectrum of the native and unfolded protein was obtained using the Fluoromax-4 Spectrophotometer (HORIBA). Protein samples were loaded in the 200 L cylindrical cuvette and excited at 275 nm with slit width 5nm to trace the tyrosine signal. The emission wavelength signal was collected with the range from 265 nm to 450 nm, with a slit width 5 nm. The emission spectrum data were calibrated by running the same emission spectra using the buffer and denaturant that protein samples are stored and incubated. The emission spectra signals of the protein were subtracted by the signal of the buffer to minimize the effects of the environment.

Guanidine Hydrochloride Unfolding Experiment

In this experiment, the unfolding process of the protein was monitored by Fluoromax-4 Spectrophotometer (HORIBA) in different concentrations of guanidine denaturation buffer. The 7M guanidine hydrochloride denaturation buffer was prepared as a stock solution and was diluted in different titrations from 0.0 M to 6.0 M in the protein samples with consistent concentration and the SE buffer. The samples were incubated in the guanidine hydrochloride denaturation buffer for 7 days and 14 days at room temperature, and the emission spectrum of the samples with different titrations of denaturant were obtained by tracing the tyrosine signal (excitation: 275 nm; emission: 265 nm to 450 nm; slit width: 5 nm). Same as the previous fluorescence experiment protocol, the emission spectra signals were calibrated by the corresponding buffer that the protein samples are in.

NMR Experiment of M-WT

The NMR experiment of M-WT was done through the 4-channel, z-gradient Bucker Avance 600 with 5 mm triple-resonance indirect Cryoprobe. ^{15}N isotopically labeled M-WT were prepared in the following states: native M-WT, and M-WT incubated in 7.0M Urea denaturation buffer. 550μL native M-WT samples in isotopically labelled SE buffer were prepared by adding 50μL D_2O (Cambridge Isotope Laboratories) to make the final volume 600μL. The two-dimensional 1H-15N heteronuclear single quantum coherence (HSQC) spectrum was obtained at the temperature of 305K.

References

Altschul, S.F. & Gish, W. (1996) "Local alignment statistics." Meth. Enzymol. 266:460-480.

Anantharaman, V.; Koonin, E. V.; Aravind, L. SPOUT: A Class of Methyltransferases That. Includes SpoU and TrmD RNA Methylase Superfamilies, and Novel Superfamilies of Predicted Prokaryotic RNA Methylases. J. Mol. Microbiol. Biotechnol. 2002, 4 (1), 71–75.

Andrews, B. T.; Capraro, D. T.; Sulkowska, J. I.; Onuchic, J. N.; Jennings, P. A. Hysteresis as a Marker for Complex, Overlapping Landscapes in Proteins. J. Phys. Chem. Lett. 2013, 4 (1), 180–188.

Bailey, D. H.; Swarztrauber, P. N. A Fast Method for the Numerical Evaluation of Continuous. Fourier and Laplace Transforms. SIAM J. Sci. Comput. 1994, 15 (5), 1105–1110.

Balestrino, R.; Schapira, A. H. V. Parkinson Disease. Eur. J. Neurol. 2020, 27 (1), 27–42.

Benítez-Páez, A.; Villarroya, M.; Armengod, M.-E. The Escherichia Coli RlmN Methyltransferase Is a Dual-Specificity Enzyme That Modifies Both RRNA and TRNA and Controls Translational Accuracy. RNA 2012, 18 (10), 1783–1795.

Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th edition. New York: W H Freeman; 2002. Section 4.1, The Purification of Proteins Is an Essential First Step in Understanding Their Function.

Brambrink, A. M.; Koerner, I. P. Parkinson's Disease (PD). In Decision Making in. Anesthesiology; Elsevier, 2007; pp 178–179.

Burban, D. J.; Haglund, E.; Capraro, D. T.; Jennings, P. A. Heterogeneous Side Chain. Conformation Highlights a Network of Interactions Implicated in Hysteresis of the Knotted Protein, Minimal Tied Trefoil. J. Phys. Condens. Matter 2015, 27 (35), 354108.

Capraro, D. T.; Burban, D. J.; Jennings, P. A. Unraveling Allostery in a Knotted Minimal. Methyltransferase by NMR Spectroscopy. J. Mol. Biol. 2020, 432 (9), 3018–3032.

Capraro, D. T.; Jennings, P. A. Untangling the Influence of a Protein Knot on Folding. Biophys. J. 2016, 110 (5), 1044–1051.

Cavanagh, J.; Skelton, N.; Fairbrother, W.; Rance, M.; Palmer, A., III. Protein NMR. Spectroscopy, 2nd ed.; Academic Press: San Diego, CA, 2006.

Chatterjee, A.; Kumar, A.; Chugh, J.; Srivastava, S.; Bhavesh, N. S.; Hosur, R. V. NMR of. Unfolded Proteins. J. Chem. Sci. (Bangalore) 2005, 117 (1), 3–21.

Chen, S.; Yoshita, M.; Ishikawa, A.; Mochizuki, T.; Maruyama, S.; Akiyama, H.; Hayamizu, Y.; Pfeiffer, L. N.; West, K. W. Intrinsic Radiative Lifetime Derived via Absorption Cross Section of One-Dimensional Excitons. Sci. Rep. 2013, 3 (1), 1941.

Creed, D. The Photophysics and Photochemistry of the Near-UV Absorbing Amino Acids-Ii. Tyrosine and Its Simple Derivatives. Photochem. Photobiol. 2008, 39 (4), 563–575.

Cromwell, P. R. Knots and Links; Cambridge University Press, 2004.

Dayhoff, M.O., Schwartz, R.M. & Orcutt, B.C. (1978) "A model of evolutionary change in. proteins." In "Atlas of Protein Sequence and Structure, vol. 5, suppl. 3." M.O. Dayhoff (ed.), pp. 345-352, Natl. Biomed. Res. Found., Washington, DC.

Deepak. Factors Influencing Chemical shifts of NMR active nuclei https://lab-training.com/2015/12/15/factors-influencing-chemical-shifts-of-nmr-active-nuclei/ (accessed May 28, 2021).

Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. Science 2006, 312 (5771), 212–217.

Eftink, M. R.; Selvidge, L. A.; Callis, P. R.; Rehms, A. A. Photophysics of Indole Derivatives: Experimental Resolution of 1La and 1Lb Transitions and Comparison with Theory. J Phys Chem 1990, 94, 3469–3479.

Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for. Mass Spectrometry of Large Biomolecules. Science 1989, 246 (4926), 64–71.

Flapan, E.; He, A.; Wong, H. Topological Descriptions of Protein Folding. Proc. Natl. Acad. Sci. U. S. A. 2019, 116 (19), 9360–9369.

Handbook of Size Exclusion Chromatography and Related Techniques: Revised and Expanded; Wu, C.-S., Ed.; CRC Press: Boca Raton, FL, 2003.

Hancock, J.M. & Armstrong, J.S. (1994) "SIMPLE34: an improved and enhanced. implementation for VAX and Sun computers of the SIMPLE algorithm for analysis of clustered repetitive motifs in nucleotide sequences." Comput. Appl. Biosci. 10:67-70.

Hori, H. Transfer RNA Methyltransferases with a SpoU-TrmD (SPOUT) Fold and Their Modified Nucleosides in TRNA. Biomolecules 2017, 7 (4), 23.

Howard, M. J. Protein NMR Spectroscopy. Curr. Biol. 1998, 8 (10), R331-3.

Huber, R.; Langworthy, T. A.; König, H.; Thomm, M.; Woese, C. R.; Sleytr, U. B.; Stetter, K. O. Thermotoga Maritima Sp. Nov. Represents a New Genus of Unique Extremely Thermophilic Eubacteria Growing up to 90 ℃. Arch. Microbiol. 1986, 144 (4), 324–333.

Jablonski, A. Efficiency of Anti-Stokes Fluorescence in Dyes. Nature 1933, 131 (3319), 839-840.

Jackson, S. E.; Suma, A.; Micheletti, C. How to Fold Intricately: Using Theory and Experiments. to Unravel the Properties of Knotted Proteins. Curr. Opin. Struct. Biol. 2017, 42, 6–14.

Jinbo Xu. Distance-based protein folding powered by deep learning. PNAS, 2019.

Jinbo Xu and Sheng Wang. Analysis of distance-based protein structure prediction by deep. learning in CASP13, PROTEINS 2019.

Karlin, S. & Altschul, S.F. (1993) "Applications and statistics for multiple high-scoring segments. in molecular sequences." Proc. Natl. Acad. Sci. USA 90:5873-5877.

Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of. Bacteriophage T4. Nature 1970, 227 (5259), 680–685.

Lakowicz, J. R. Principles of Fluorescence Spectroscopy; Lakowicz, J. R., Ed.; Springer: New. York, NY, 2017.

Masalha, M.; Borovok, I.; Schreiber, R.; Aharonowitz, Y.; Cohen, G. Analysis of Transcription of. the Staphylococcus Aureus Aerobic Class Ib and Anaerobic Class III Ribonucleotide Reductase Genes in Response to Oxygen. J. Bacteriol. 2001, 183 (24), 7260–7272.

Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the. Analysis of Molecular Dynamics Simulations. J. Comput. Chem. 2011, 32 (10), 2319–2327.

Mori, S.; Barth, H. G. Size Exclusion Chromatography, 1999th ed.; Springer: Berlin, Germany, 1999.

Nfor, B. K.; Ahamed, T.; van Dedem, G. W. K.; van der Wielen, L. A. M.; van de Sandt, E. J. A. X.; Eppink, M. H. M.; Ottens, M. Design Strategies for Integrated Protein Purification Processes: Challenges, Progress and Outlook. Journal of Chemical Technology & Biotechnology 2008, 83 (2), 124–132.

Noel, J. K.; Sułkowska, J. I.; Onuchic, J. N. Slipknotting upon Native-like Loop Formation in a Trefoil Knot Protein. *Proc. Natl. Acad. Sci. U. S. A.* 2010, *107* (35), 15403–15408.

Nowakowski, A. B.; Wobig, W. J.; Petering, D. H. Native SDS-PAGE: High Resolution. Electrophoretic Separation of Proteins with Retention of Native Properties Including Bound Metal Ions. Metallomics 2014, 6 (5), 1068–1078.

Park Y, Sheetlin S, Ma N, Madden TL, & Spouge JL. (2012) "New finite-size correction for local. alignment score distributions." BMC Res Notes. 2012 Jun 12;5:286.

Ross, J. B. A.; Laws, W. R.; Rousslang, K. W.; Wyssbrod, H. R. Tyrosine Fluorescence and. Phosphorescence from Proteins and Polypeptides; Plenum Press: New York, 1992; Vol. 3, pp 1–63.

Rule, G. S.; Hitchens, T. K. Fundamentals of Protein NMR Spectroscopy; Springer: Dordrecht, Netherlands, 2010.

Schnarr, M.; Helene, C. Effects of Excited-State Proton Transfer on the Phosphorescence of. Tyrosine-Phosphate Complexes. Photochem. Photobiol. 1982, 36 (1), 91–93.

Shah, T. R.; Koten, H.; Ali, H. M. Performance Effecting Parameters of Hybrid Nanofluids. In. Hybrid Nanofluids for Convection Heat Transfer; Elsevier, 2020; pp 179–213.

Sheng Wang, Siqi Sun, Zhen Li, Renyu Zhang and Jinbo Xu. Accurate De Novo Prediction of. Protein Contact Map by Ultra-Deep Learning Model. PLoS Computational Biology, 2017.

Sheng Wang, Zhen Li, Yizhou Yu and Jinbo Xu. Folding membrane proteins by deep transfer. learning. Cell Systems, 2017

Sheng Wang, Siqi Sun and Jinbo Xu. Analysis of deep learning methods for blind protein. contact prediction in CASP12. PROTEINS, 2017.

Silverstein, R. M.; Webster, F. X.; Kiemle, D. Spectrometric Identification of Organic. Compounds, 7th ed.; Wiley, 2014.

Szabo, A. G.; Lynn, K. R.; Krajcarski, D. T.; Rayner, D. M. Tyrosinate Fluorescence Maxima at. 345 Nm in Proteins Lacking Tryptophan at PH 7. FEBS Lett. 1978, 94 (2), 249–252.

Taylor, A. C. Advances in Nanoparticle Reinforcement in Structural Adhesives. In Advances in. Structural Adhesive Bonding; Elsevier, 2010; pp 151–182.

Taylor, W. R. Protein Knots and Fold Complexity: Some New Twists. Comput. Biol. Chem. 2007, 31 (3), 151–162.

Tekely, P. Malcolm H. Levitt. Spin Dynamics: Basics of Nuclear Magnetic Resonance. John. Wiley & Sons, Chichester, UK, 2001, 686 Pp. Price £34.95. ISBN 0-471-48921-2: Book Review. Magn. Reson. Chem. 2002, 40 (12), 800–800.

Tkaczuk, K. L.; Dunin-Horkawicz, S.; Purta, E.; Bujnicki, J. M. Structural and Evolutionary. Bioinformatics of the SPOUT Superfamily of Methyltransferases. BMC Bioinformatics 2007, 8, 73.

Waterman, P. G. Multidimensional NMR in Liquids. Phytochemistry 1997, 44 (6), 1183.

Wootton, J.C. & Federhen, S. (1996) "Analysis of compositionally biased regions in sequence. databases." Meth. Enzymol. 266:554-571.

Wootton, J.C. & Federhen, S. (1993) "Statistics of local complexity in amino acid sequences and sequence databases." Comput. Chem. 17:149-163.

Xu, Y.; Li, S.; Yan, Z.; Luo, Z.; Ren, H.; Ge, B.; Huang, F.; Yue, T. Stabilizing Effect of Inherent. Knots on Proteins Revealed by Molecular Dynamics Simulations. Biophys. J. 2018, 115 (9), 1681–1689.

Yang, H.; Xiao, X.; Zhao, X.; Wu, Y. Intrinsic Fluorescence Spectra of Tryptophan, Tyrosine and Phenylalanine. In Selected Papers of the Chinese Society for Optical Engineering Conferences held October and November 2016; Lv, Y., Le, J., Chen, H., Wang, J., Shao, J., Eds.; SPIE, 2017.

Yao, J.; Dyson, H. J.; Wright, P. E. Chemical Shift Dispersion and Secondary Structure. Prediction in Unfolded and Partly Folded Proteins. FEBS Lett. 1997, 419 (2–3), 285–289.

Yu, H. Extending the Size Limit of Protein Nuclear Magnetic Resonance. Proc. Natl. Acad. Sci. U. S. A. 1999, 96 (2), 332–334.