

PROTEIN-PROTEIN INTERACTION OF PAI-H19Y MUTANT OF THE SLEEPING
BEAUTY TRANSPOSASE

by

Chenbo Yan

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Applied Physics

Charlotte

2020

Approved by:

Dr. Irina Nesmelova

Dr. Nathaniel Fried

Dr. Donald Jacobs

ABSTRACT

CHENBO YAN. Protein-protein interaction of PAI-H19Y mutant of the Sleeping Beauty transposase. (Under the direction of DR. IRINA NESMELOVA)

Due to the substantial interest in developing transposon technology for gene delivery and gene therapy, the Sleeping Beauty (SB) transposon system, the most active transposon in the vertebrate species, has been extensively studied. The SB transposon system is composed of the transposon DNA and the transposase enzyme. During the DNA transposition, one necessary step is the assembly of the transpososome. The transpososome is held by protein-protein interactions, and the formation of the transposase dimer is required for this process. PAI subdomain of the SB transposase has been found to mediate protein-protein interactions between the transposase subunits, but the specific conditions for this interaction remain unclear. In our recent study, we experimentally confirmed that the H19Y mutation in the PAI subdomain increased the structural stability of the subdomain, which allows studying protein-protein interactions at biologically relevant experimental conditions. In this study, we use microscale thermophoresis (MST), NMR spectroscopy, and crosslinking techniques to investigate the oligomerization of the PAI-H19Y mutant.

ACKNOWLEDGEMENTS

I want to thank my advisor Dr. Irina Nesmelova for her valuable advice and supervision in my past three years of research. Her support and guidance helped me greatly during both my Undergraduate and Master studies. All would have been impossible without her patient instructions. I would also like to thank Dr. Nathaniel Fried and Dr. Donald Jacobs for their valuable suggestions during my research and before this defense. Finally, I want to thank the Department of Physics and Optical Sciences at UNC Charlotte for Teaching and Research Assistantship. The NMR experiments were carried out at the Molecular Education, Technology, and Research Innovation Center (METRIC) at the North Carolina State University. This study was supported by the National Institute of Health grants R15GM112076 and R15GM128100 to IVN.

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF ABBREVIATIONS	vii
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: PRIOR RESEARCH	4
CHAPTER 3: TOPIC AND MOTIVATION	6
CHAPTER 4: METHODOLOGY	9
4.1 Protein expression, purification and sample preparation	9
4.2 Nuclear magnetic resonance (NMR) spectroscopy	9
4.3 Microscale thermophoresis (MST)	12
4.4 Crosslinking	15
CHAPTER 5: RESULTS	19
5.1 The association of PAI-H19Y by NMR spectroscopy	19
5.2 The association of PAI-H19Y by MST	24
5.2.1 Binding Affinity experiments using Monolith His-Tag Labeling Kit RED tris-NTA 2 nd Generation	24
5.2.2 Binding Affinity experiments using Monolith Protein Labeling Kit RED- NHS 2 nd Generation	28
5.3 The association of PAI-H19Y by crosslinking	30
5.4 DNA-binding affinity to PAI-H19Y determined by MST	33
5.4.1 Optimization of experimental condition for DNA-binding with PAI-H19Y by MST	34

5.4.2	Binding Affinity result for DNA-binding with PAI-H19Y	37
	CHAPTER 6: DISCUSSION	39
	CHAPTER 7: CONCLUSION	41
	REFERENCES	42

LIST OF FIGURES

FIGURE 1: Schematic representation of the DNA transposition	2
FIGURE 2: Schematic structure of Sleeping Beauty transposase	3
FIGURE 3: CD and NMR result of folded PAI-H19Y conformation	4
FIGURE 4: Cross-peak representation of [¹ H, ¹⁵ N]-HSQC spectrum	5
FIGURE 5: [¹ H, ¹⁵ N]-HSQC spectra of DNA-H19Y	6
FIGURE 6: Structural model of PAI subdomain binding with DNA	8
FIGURE 7: Pulse sequences for the fundamental HSQC sequence	11
FIGURE 8: Schematic representation of MST optic	13
FIGURE 9: Example of typical MST results and analysis	14
FIGURE 10: Reaction of EDC crosslinking	16
FIGURE 11: Chemical structure of BS3 crosslinker	17
FIGURE 12: Schematic gel result	18
FIGURE 13: NMR results for PAI-H19Y at pH 5	19
FIGURE 14: 3D-Model exemplifies structural changes at pH 5	20
FIGURE 15: Normalized proton chemical shifts changed at pH 5	21
FIGURE 16: NMR results for PAI-H19Y at pH 7	22
FIGURE 17: Normalized proton chemical shifts changed at pH 7	23
FIGURE 18: Binding affinity between Monolith His-Tag Labeling Kit and PAI-H19Y	24
FIGURE 19: Binding Affinity experiment using His-tag labeling dye	25
FIGURE 20: Dose Response for PAI-H19Y with His-tag labeling dye after wash	26
FIGURE 21: Dose Response for PAI-H19Y with His-tag labeling dye at pH 5 and 7	27
FIGURE 22: MST results for PAI-H19Y with lysine labeling dye	29

FIGURE 23: SDS-PAGE gel result for EDC crosslinking	30
FIGURE 24: SDS-PAGE gel result for EDC plus sulfo-NHS crosslinking	31
FIGURE 25: SDS-PAGE gel result for BS3 crosslinking	33
FIGURE 26: MST result for Pretest with His-Tag labeled DNA	34
FIGURE 27: MST result for Binding Check for His-Tag labeled DNA and PAI-H19Y	35
FIGURE 28: MST result for Binding Check using Premium Capillaries	36
FIGURE 29: MST result for Binding Check using detergent	37
FIGURE 30: MST results for DNA-binding experiments with PAI-H19Y	37

LIST OF ABBREVIATIONS

SB	Sleeping Beauty
CD	circular dichroism
NMR	nuclear magnetic resonance
MST	microscale thermophoresis
HSQC	heteronuclear single quantum coherence
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride
BS3	Bis(sulfosuccinimidyl) suberate
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis

CHAPTER 1: INTRODUCTION

Gene therapy utilizes the principle that disease can be treated by the introduction of foreign genetic material into cells of patients to modify gene expression and cell behavior. This delivery of gene is especially effective against diseases where natural immunity is abnormal or ineffective, such as cancer or AIDS [1]. Many gene delivery systems have been developed in the past years using either viral or nonviral approach. Although viral vectors offer the potential of efficient gene transfer, their preparations are time-consuming and expensive. Using viral vectors also have the risk of contamination and causing unwanted immune responses. In contrast, nonviral DNA plasmids can be manufactured at a reduced cost and with lower immunogenicity, however, their rate of delivery to target cell nuclei and the rate of integration is low [2].

The *Sleeping Beauty* (SB) DNA transposon offers a new nonviral vector with increased rate of transposition. As the most active *Tc1/mariner*-like transposon in vertebrate species, SB transposon has been developed as the leading nonviral vector for gene therapy, and it's the first and only transposon that is currently in clinical trials for human gene therapy [3]. *Tc1/mariner* is one of the 23 recognized super-families of DNA transposons, and all transposons within the family shares a similar structure and mechanism when comes to DNA transposition. The transposition is achieved through a cut-and-paste mechanism, where two transposase enzymes bind to the two ends of the transposon DNA, and they join together to form a synaptic complex, which is then integrated into the target DNA, as shown in Figure 1 [4].

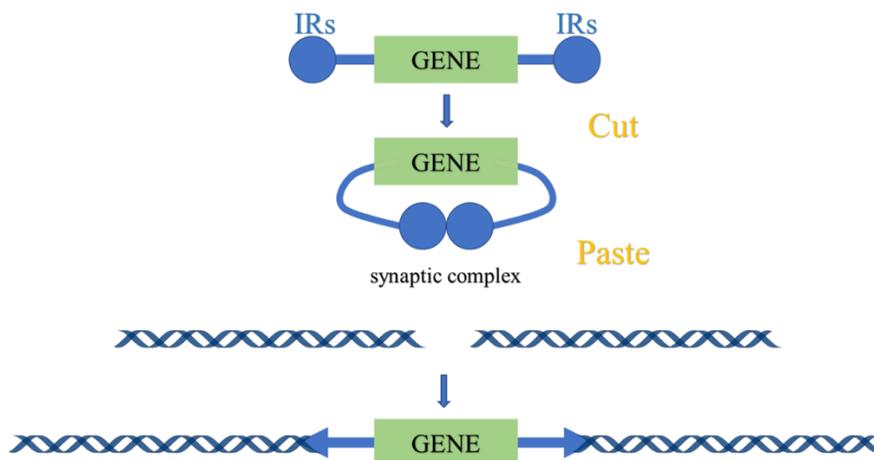


Figure 1. Schematic representation of the DNA transposition process for SB transposon system.

The efficiency of the transposon system depends on how well the transposon DNA and the transposase enzyme work together, both components can be modified to achieve higher transposition efficiency. The focus of this study is the transposase enzyme, which consists of two structurally independent domains, the DNA-binding domain and the catalytic domain. The DNA-binding domain further consists of two subdomains, PAI and RED (Fig. 2). Previous studies have shown that the PAI subdomain is the primary DNA-recognition domain, and only its folded conformation can bind to the DNA [5,6]. However, the folding conformation can only be achieved at non-physiologic environmental conditions, such as low temperature (5°C), high concentration of salt (more than 600 mM NaCl), and pH values greater than 7.0 [7]. We addressed this problem in our previous research by inducing a stabilizing mutation H19Y, where a histidine is replaced by a tyrosine.

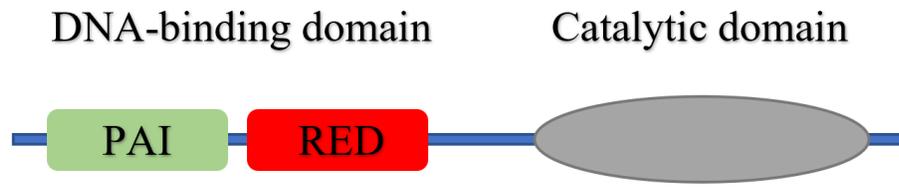


Figure 2. Schematic structure of *Sleeping Beauty* transposase.

CHAPTER 2: PRIOR RESEARCH

In our previous research, we studied the PAI subdomain of the *Sleeping Beauty* transposase with the purpose of identifying the structure-stabilizing mutation and defining the binding interface between the PAI subdomain and the transposon DNA. We have chosen the H19Y mutation since we observed the pH induced folding for wildtype PAI is likely related to the deprotonation of histidine side chain(s), and the H19Y mutation was predicted to have the strongest stabilizing effect based on Gibbs free energy of folding. Using circular dichroism (CD) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy, we have experimentally confirmed that the H19 to tyrosine mutation stabilizes the PAI subdomain of the SB transposase, causes it to remain folded at physiologic environmental conditions.

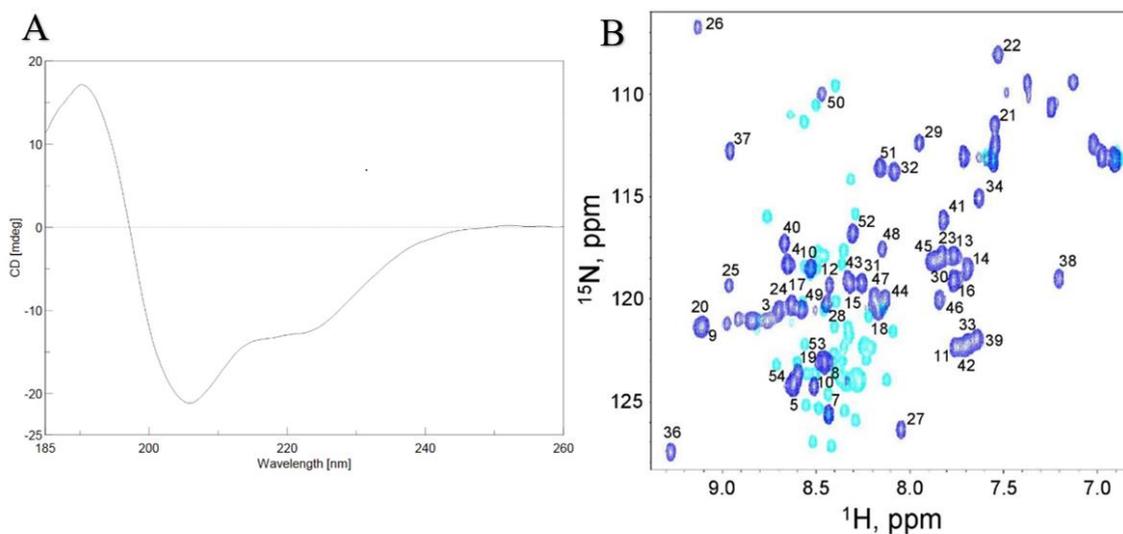


Figure 3. (A) CD spectrum of PAI-H19Y mutant showing folded conformation. (B) [^1H , ^{15}N]-HSQC spectra of the wildtype PAI subdomain (cyan cross-peaks) and H19Y mutant (blue cross-peaks) showing the folding of the protein induced by the H19Y mutation. All spectra were collected in 25 mM sodium phosphate buffer at 35°C, pH 5.0.

Figure 3 shows the CD and NMR spectra of PAI-H19Y subdomain collected in the 25 mM sodium phosphate buffer, at 35°C and pH 5.0. The CD spectrum (Figure 3A) clearly

shows a folded conformation of PAI-H19Y judged by the positive peak at 198 nm and the two negative peaks at 208 and 222 nm. Figure 3B shows the $[^1\text{H},^{15}\text{N}]$ -HSQC spectra of the wildtype PAI subdomain (cyan cross-peaks) and H19Y mutant (blue cross-peaks). $[^1\text{H},^{15}\text{N}]$ -HSQC detects the amide proton that attached to a nitrogen in the peptide bond, since all amino acid residues (except proline) has amide proton, therefore, the residue can be visualized as a peak in the $[^1\text{H},^{15}\text{N}]$ -HSQC spectrum (Fig. 4).

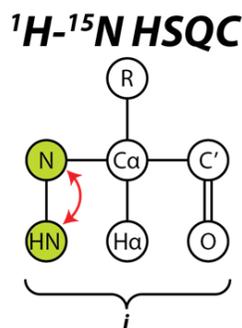


Figure 4. The peptide bond between amid proton and nitrogen is detected and shown as cross-peak in $[^1\text{H},^{15}\text{N}]$ -HSQC spectrum.

As shown in Figure 3B, the spectrum of the wildtype PAI subdomain (cyan) reveals very narrow distribution of cross-peaks, indicating that the respective amino acid residues are in random coil like conformation, thus the unfolded structure [5,7]. In contrast, H19Y mutant (blue) displays a drastically different spectrum with widely dispersed residues, indicative of a folded structure. Therefore, both the CD and NMR results have proven the structural-stabilizing effect of the H19Y mutation.

CHAPTER 3: TOPIC AND MOTIVATION

As shown in Figure 10, one important step during the transposition process is the binding of the transposon DNA to the transposase enzyme and the formation of synaptic complex (transpososome) [6]. As observed from other Tc1/*mariner* transposases, protein-protein interactions between the transposases are required to form the transpososome, and the PAI subdomain is found to mediate protein-protein interactions between the transposase subunits [4]. Previous study has also shown that if protein-protein interactions are inhibited, the DNA transposition is completely abolished [6]. Therefore, it is our interest to understand the protein-protein interaction and identify the condition in which the interaction could occur. Additionally, in our previous research, we collected the [^1H , ^{15}N]-HSQC spectrum for PAI-H19Y subdomain with and without DNA at 0.2 mM protein concentration at pH5, as shown in Figure 5.

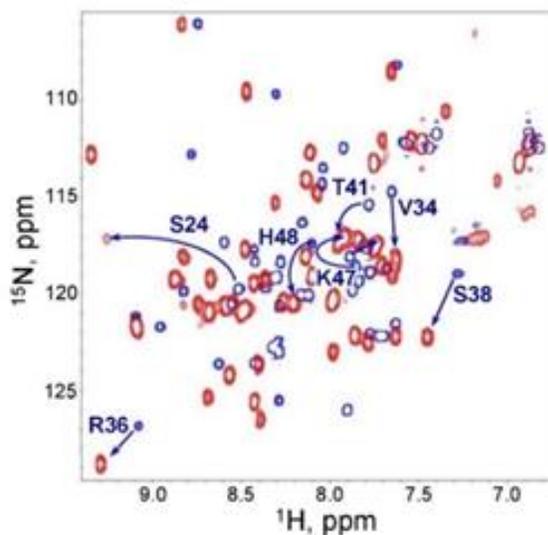


Figure 5. [^1H , ^{15}N]-HSQC spectra of ^{15}N , ^{13}C -labeled H19Y in the absence (blue cross-peaks) and presence (red cross-peaks) of DNA. The red cross-peaks are clearly broader than the blue cross-peaks, suggesting potential protein-protein interaction.

In Figure 5, The blue cross-peaks represents the spectrum without DNA, and the red cross-peaks represents the spectrum after the DNA-H19Y binding. Large number of chemical shifts are present after the binding with DNA, marked with the blue arrows, chemical shifts are clear indications of structural change caused by the binding. We also observed the cross-peak broadening for the DNA-H19Y complex (red) compare to the narrow peaks for H19Y(blue). The NMR peak broadening can be caused by three reasons: the rearrangement of structure of the PAI subdomain in the presences of DNA, the exchange between DNA-bound and unbound states of the PAI subdomain, or the dimerization of the DNA-H19Y complexes, such the complex is formed by two DNA and two PAI-H19Y molecules [7]. This study will help us investigate the possibility of the dimerization of PAI-H19Y mutant and, if detected, determine its binding affinity.

Furthermore, Using the chemical shifts data, we have generated the PAI-DNA complex for the original PAI subdomain using HADDOCK (High Ambiguity Driven biomolecular DOCKing) program version 2.2 [8], in which the helix H3 interacts with the DNA major groove, helix H2 forms contacts with the DNA minor groove, and helix H1 is positioned away from the DNA, available for protein-protein interaction. In this study, if protein-protein interaction is detected using NMR, we should observe chemical shifts on some of the residues that lies on the binding surface, likely the helix H1. The schematic representation of the PAI-DNA complex is shown in Figure 6.

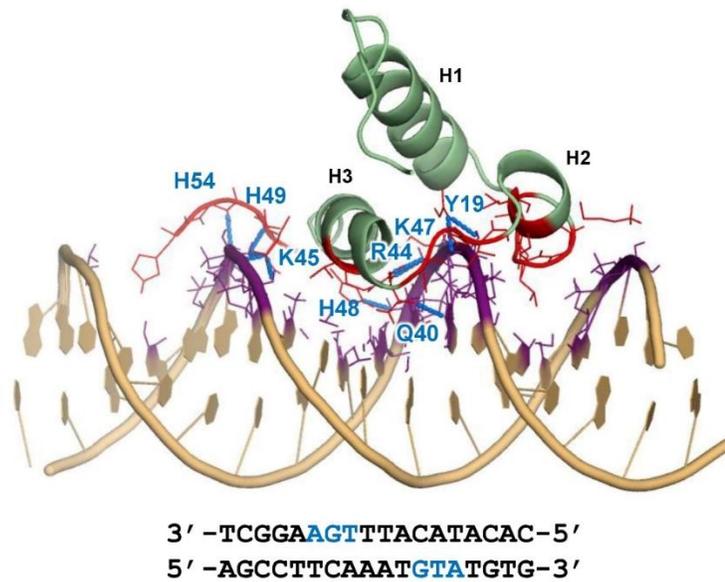


Figure 6. Structural model of PAI subdomain binding with DNA. Helix H1 is positioned away from the DNA, available for protein-protein interaction.

CHAPTER 4: METHODOLOGY

Three techniques are used in this study to examine the binding, microscale thermophoresis (MST), nuclear magnetic resonance (NMR) spectroscopy and crosslinking, each technique has their own advantage and limitation. MST is an invasive method that requires dye as label, but it can measure samples within a wide range of concentration (nM ~ mM) in a short amount of time, a typical binding affinity experiments only takes about 10 minutes. NMR is noninvasive but very time consuming when comes to data collection and analyzation, above all, it can't reach very low concentration (nM) where the signal will be too weak to detect. Crosslinking can be used to check if the protein-protein interaction can occur on certain conditions and provide a general idea about the binding interface, but it will change the natural structure and behavior of the protein by attaching the crosslinker.

4.1. Protein expression, purification and sample preparation

All proteins are expressed and purified using an established protocol [7]. DNA plasmids encoding the original PAI subdomain (PAI-WT) or its K14RK33A and H19YK14RK33A mutants are ordered from GenScript USA Inc. (Piscataway, NJ, USA). The proteins are expressed in BL21-A1 *E. coli* cells. For ¹⁵N labeling, bacterial cells are grown in M9 medium prepared with ¹⁵NH₄Cl as the sole nitrogen sources. The proteins were purified by metal chelating chromatography using a Ni-NTA Agarose (Thermo Fisher Scientific). For DNA binding experiments, the 18-base pair (bp) DNA-core sequence 3'-TCGGAAGTTTACATACAC-5' synthesized by IDT (Integrated DNA Technologies, Inc.) has been used.

4.2. Nuclear magnetic resonance (NMR) spectroscopy

NMR spectroscopy is based on the principle that all nuclei have intrinsic quantum property of spin and angular momentum. When a ground state (spin-down) nucleus is placed in an oscillating magnetic field with oscillation frequency matches its intrinsic resonance frequency, the nucleus absorbs the energy and reach excited state (spin-up). As it returns to the ground state, energy at the same resonance frequency is emitted and picked up by the NMR spectroscopy. Since the resonance frequency of the nucleus in a molecule is changed by the intramolecular magnetic field surrounding it. Analyzing the change of frequencies provides details of a molecule's individual functional groups and its structure. The reason we chose NMR was due to its non-destructive nature, and because the sample can be measured multiple times after changing its buffer condition or protein concentration. Moreover, NMR provides us information on intramolecular interactions, as the structure of the molecule can be measured in solution in their natural state, which is essential when studying protein-protein binding. The structural changes of the molecules are displayed by amino acid chemical shifts. Mapping the chemical shifts on the molecules helps define the binding interface.

We used NMR to collect the [^1H , ^{15}N]-HSQC spectrum of the PAI-H19Y subdomain at varying protein concentration to observe the possible protein-protein interaction. Proton chemical shifts were calibrated with respect to water signal. Sequence-specific resonance assignments have been performed using 3D HNCACB, CBCA(CO)NH, HCCH-TOCSY, ^{15}N -separated TOCSY-HSQC and HSQC-NOESY, and ^{13}C -separated NOESY-HSQC experiments as described in original references [9,19].

Shown in Figure 7 is the pulse sequence for the fundamental HSQC experiment. Thin bars indicate 90° radio-frequency pulses, and thick bars indicate 180° radio-frequency pulses,

whereas y or $-y$ indicates the direction about which the pulse rotates the magnetization. The HSQC experiment employs two blocks for transfer the magnetization from proton (^1H , labeled as I in Fig. 7) to nitrogen (^{15}N , labeled as S in Fig. 7) in an isotopically ^{15}N -enriched protein. The first block, consisting of 90-180-90 sequence applied to ^1H and 180-90 sequence applied to ^{15}N creates antiphase heteronuclear coherence ($2I_xS_z$). Then, the S -spin coherence is frequency labeled during the t_1 period. The second block, consisting of 90-180 pulse sequences applied to both ^1H and ^{15}N converts this coherence back to observable magnetization. Prior to detection, a basic two step phase cycle is used to filter out the undesirable magnetization from protons attached to ^{12}C or ^{14}N , and a four-step cycle is required to completely eliminate the undesirable double quantum terms [20].

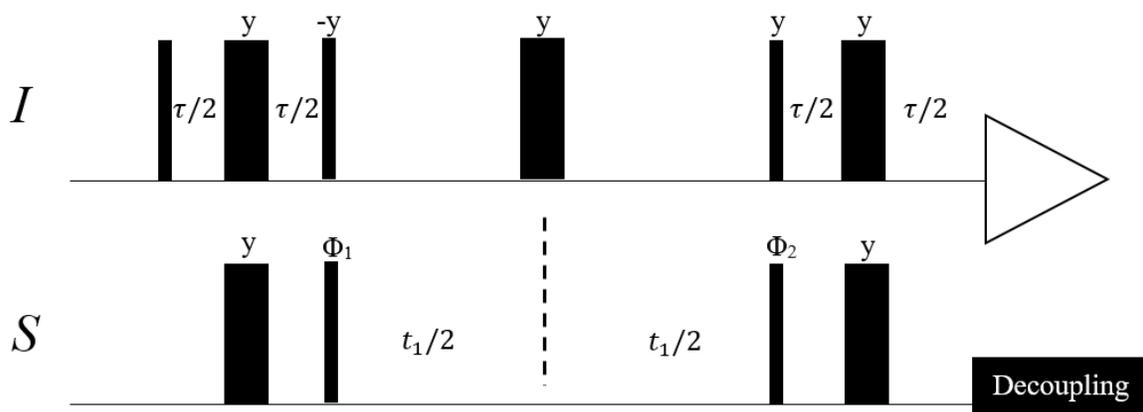


Figure 7. Pulse sequences for the fundamental HSQC sequence.

Throughout the NMR experiments, if we can observe chemical shifts for some of the cross-peaks, it would indicate structural changes and possible protein-protein interaction. Analyzing the chemical shifts with respect to protein concentration helps determine the binding affinity, adding to our knowledge about the PAI-H19Y mutant. If no chemical shifts is observed, it would indicate no binding occurs at chosen experimental conditions.

The NMR experiments were performed at the Molecular Education, Technology, and Research Innovation Center (METRIC) at the North Carolina State University. All NMR spectra are processed using the NMRPipe software and analyzed with programs CARA and NMRView [10,11,12].

4.3. Microscale thermophoresis (MST)

Microscale thermophoresis is based on a phenomenon called thermodiffusion, e.g., the movement of molecules in temperature gradients [13]. Utilizing this effect, MST monitors the change of motion of fluorescent molecules along a microscopic temperature gradient in μl -volumes, allowing for the precise analysis of binding events and determine binding affinities. The temperature gradient ΔT , induced by an infrared (IR) laser, leads to the depletion of the solvated biomolecules in the region of elevated temperature, quantified by the Soret coefficient S_T [14]:

$$c_{hot}/c_{cold} = \exp(-S_T\Delta T) \quad (1)$$

where c is molecule concentration. This depletion depends on the interface between molecule and solvent, and the thermophoresis of a protein differs significantly from the thermophoresis of a protein-ligand complex due to the change in size, charge and energy [15].

To observe the thermophoresis of molecules, we needed to label the molecules with fluorescent dye thus turn them into fluorophores, which has the property of absorbing photons and emitting photons of lower energy in return. When photon is absorbed by the fluorophore, its energy level raises from the ground state to an excited state, it will then return spontaneously to the ground state while emitting a photon. The wavelength of the absorbed and emitted photon is specific to the fluorophore.

We use MST to study protein-protein interaction with an experimental setup shown in Figure 8. As shown on the left, the thermophoresis of the sample is measured in capillaries with a volume of $\sim 10 \mu\text{L}$ each. The thermophoresis is induced by the IR-Laser into the capillary to produce a microscopic temperature gradient spanning $2\text{-}6^\circ\text{C}$ in a volume with a diameter of $50 \sim \mu\text{m}$. The fluorescence within the capillary is excited and detected through the same objective [14].

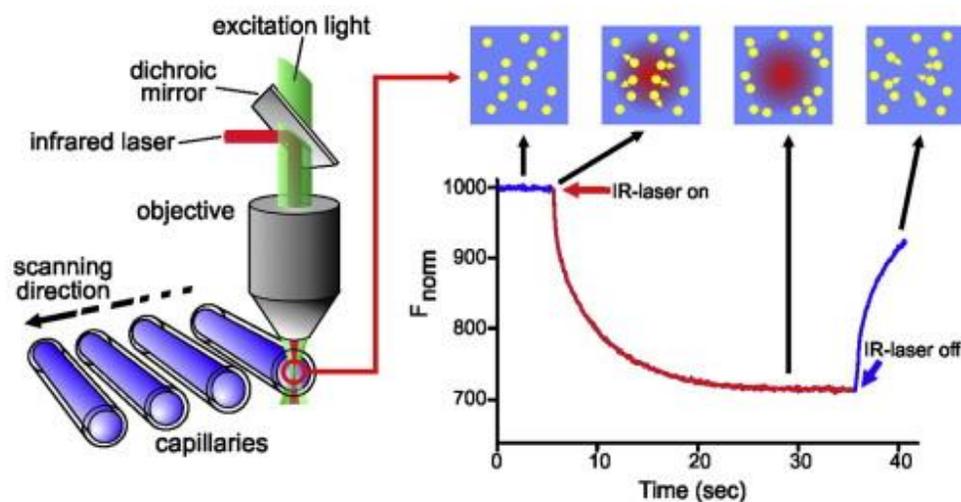


Figure 8. Schematic representation of MST optic [7].

Shown on the right is a typical signal of an MST experiment. Initially, the molecules are homogeneously distributed in the capillary. Within the first second after activation of the IR laser, the “T-Jump” is observed, which corresponds to a rapid change in fluorescence due to the fast temperature change. This fluorescence change is measured for around 30 seconds. After deactivation of the IR-Laser, an inverse “T-Jump” occurs, followed by the redistribution of molecules, which is solely driven by mass diffusion [14].

The fluorescence is induced by labelling the protein with RED fluorescent dye. Two types of dye are used, Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation and Monolith Protein Labeling Kit RED-NHS 2nd Generation.

For deriving binding constant, the concentrations of the labelled protein are to be kept same for all capillaries, and the concentration of the ligand are decreased by a factor of two for each capillary. When the fluorophores within the capillary is excited by the IR-Laser, the fluorescence is emitted and collected by the optics. The MST signal is plotted against the ligand concentration to obtain a dose-response curve, from which the binding affinity is determined.

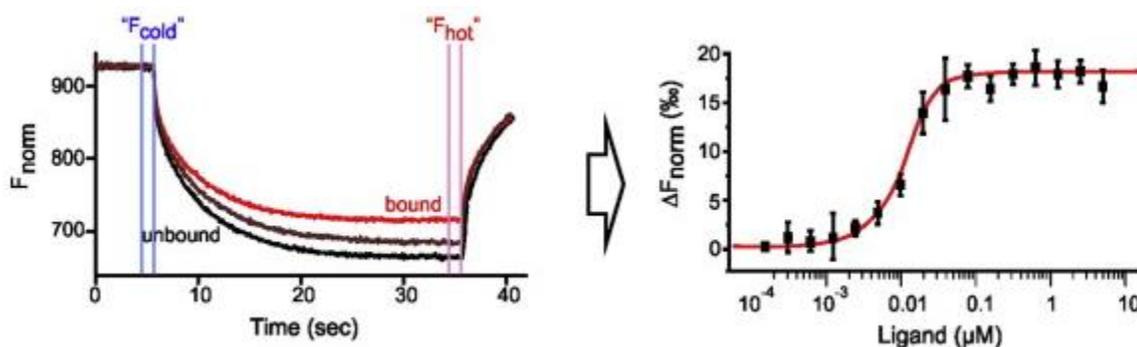


Figure 9. Example of typical MST results and analysis. The traces on the left representing the bounded (red) and unbounded (black) states. The change in thermophoresis is expressed as the change in the normalized fluorescence (ΔF_{norm}), which is defined as F_{hot}/F_{cold} . The signal is plotted against ligand concentration for analysis on the right [7].

Shown in Figure 9 is a typical binding experiment, the MST traces shown on the left indicates the change in fluorescence intensity of one capillary over time, the trace curves up when a fluorescent molecule binds to the non-fluorescence ligand, the black traces are the unbound states the red trace is the bound state. The F_{cold} and F_{hot} are the average fluorescence values between defined areas marked by the blue and red lines, respectively. For analysis, the change in thermophoresis is expressed as the change in the normalized fluorescence (ΔF_{norm}), which is defined as F_{hot}/F_{cold} . As Shown on the right, to calculate the binding affinity, we decrease the concentration of the ligand gradually, thus decrease F_{norm} , which is then plotted against the concentration of the ligand and fitted to derive binding constant [14].

All MST experiments were conducted on Monolith NT.115 at the University of North Carolina at Charlotte.

4.4. Crosslinking

Crosslinking refers to the use of a probe (crosslinker) to link proteins together to check for protein-protein interactions. Crosslinker usually react with specific functional groups on proteins and form a covalent bond between them. In this study, we use crosslinkers with different arm lengths to predict the distances between interacting molecules. By examining which crosslinker successfully obtained the oligomer structure, information may be obtained about the conformational changes that hindered or exposed the binding interface.

Two crosslinkers are used, EDC and BS3. The EDC functions by converting carboxyl group into amine-reactive O-acylisourea intermediate that quickly reacts with an amino group to form an amide bond with release of an isourea by-product. The intermediate is unstable in aqueous solutions and so N-hydroxysuccinimide (NHS) can be added to improve efficiency or create dry-stable (amine-reactive) intermediates, the reaction is illustrated in Figure 10 [16].

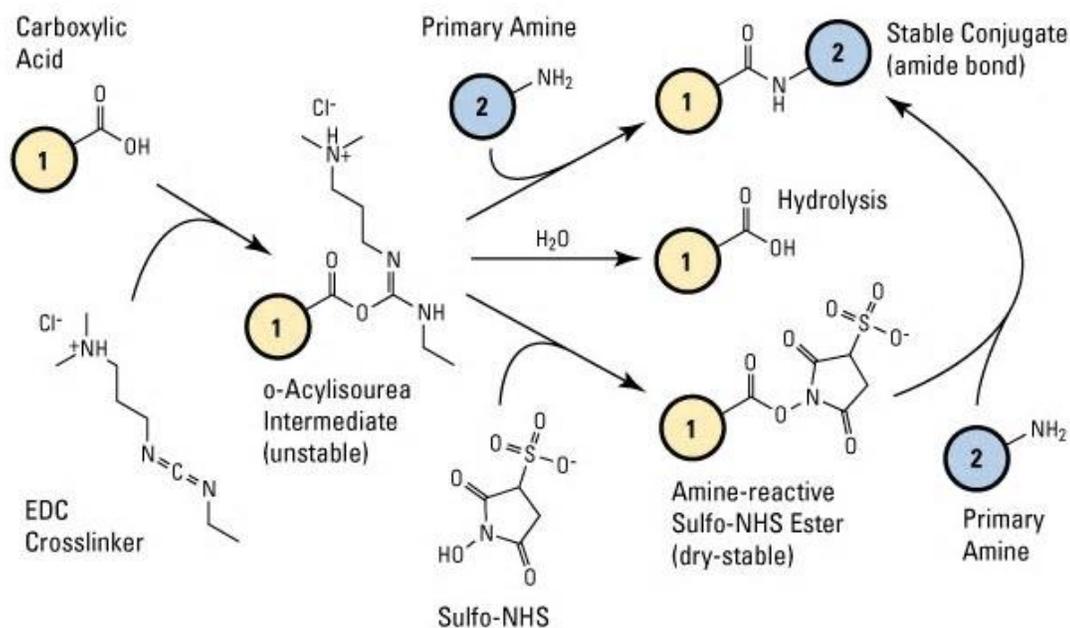
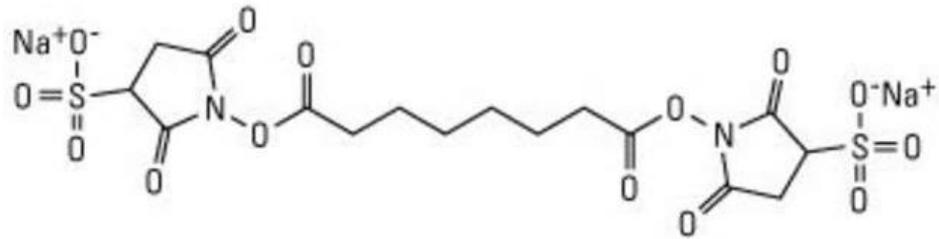


Figure 10. Reaction of EDC crosslinking. Molecules (1) and (2) can be peptides, proteins or any chemicals that have respective carboxylic acid and primary amine groups. The addition of Sulfo-NHS (bottom-most pathway) can increase efficiency and stability [16].

The BS3 crosslinker contains an amine-reactive NHS ester that reacts with primary amines at pH 7-9 to form stable amide bonds, along with release of the NHS leaving group. Proteins generally have several primary amines in the side chain of lysine residues and the N-terminus of each polypeptide that are available as targets for NHS-ester crosslinking reagents [17]. The chemical structure of BS3 is shown in Figure 11.



BS3
Bis(sulfosuccinimidyl) suberate
MW 572.43
Spacer Arm 11.4 Å

Figure 11. Chemical structure of BS3 crosslinker [17].

Compare to zero-arm length crosslinker (EDC), BS3 have a much longer spacer arm length at 11.4 Angstrom, providing more freedom for the bonding to occur.

After the cross-linking reaction, the sample are loaded onto 16.5% sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) to visualize the binding. If the protein–protein interaction does occur, a band representing the oligomer should appear in the gel above the monomer, as it moves slower in gel. Otherwise, only the band representing the monomer should form, shown in Figure 12 is a schematic representation of a gel result, column on the left is the protein ladder, used as standards to identify the approximate molecular weight of the protein. Middle column represents the sample with binding, displaying two binds for protein monomer and dimer. Right column represents the sample without binding, showing only one bind for protein monomer.

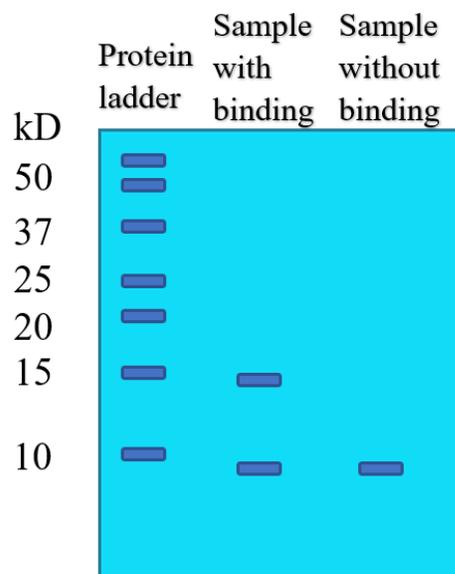


Figure 12. Schematic gel result. Column on the left is the protein ladder, used as standards to identify the approximate molecular weight of the protein. Middle column represents the sample with binding, displaying two binds for protein monomer and dimer. Right column represents the sample without binding, showing only one bind for protein monomer.

CHAPTER 5: RESULTS

5.1. The association of PAI-H19Y by NMR spectroscopy

We used NMR spectroscopy to access whether the PAI-H19Y mutant associates in NMR-accessible concentration range (0.02-1.25 mM). For the NMR experiment, the ^{15}N labeled protein was prepared in 25mM sodium phosphate buffer with 150 mM NaCl at pH 5 with initial protein concentration at 1.25 mM, and was diluted gradually throughout the experiment.

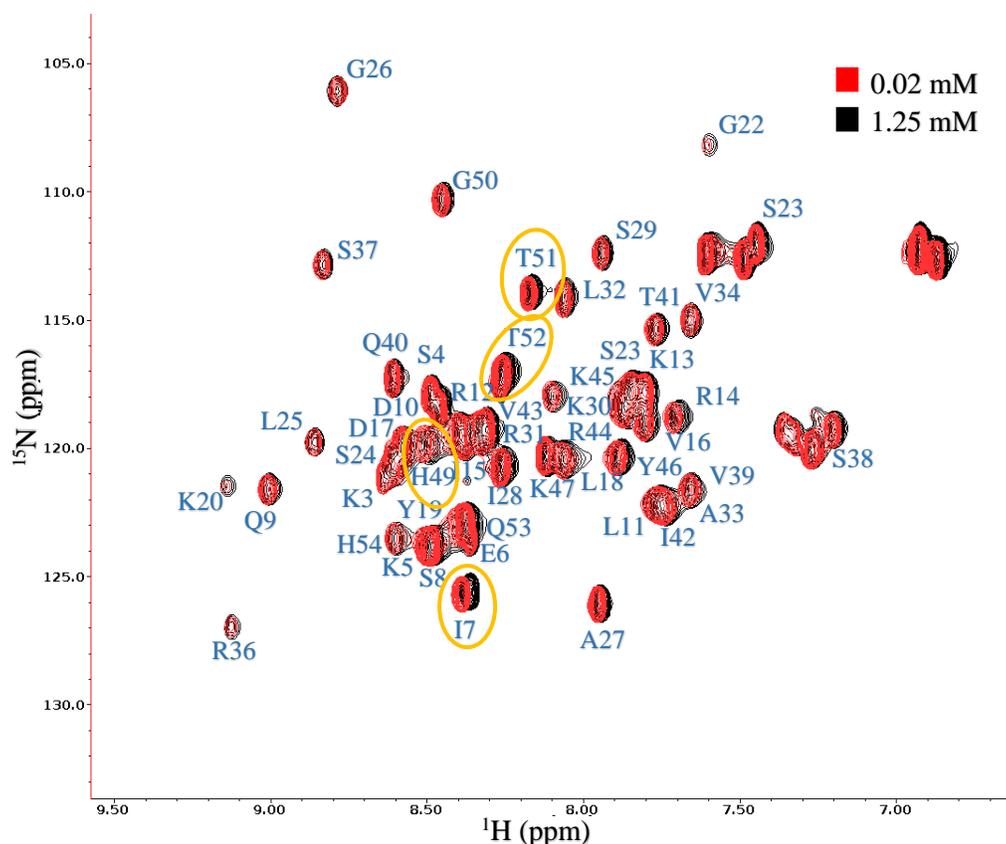


Figure 13. $[^1\text{H}, ^{15}\text{N}]$ -HSQC spectra for ^{15}N -labeled H19Y at concentrations of 1.25 mM and 0.02 mM. Chemical shifts changes can be observed for most residues and are more substantial for some of the residues such as I7, H49, T51 and T52, circled in orange. Buffer condition: 25mM sodium phosphate buffer with 150 mM NaCl at pH 5 at 25 °C.

Figure 13 shows the two of seven ^{15}N - ^1H HSQC spectrum collected at the protein concentrations of 1.25 mM and 0.02 mM, in black and red respectively. NMR signal

assignments to PAI-H19Y amino acids are indicated, the letters are the abbreviation of the amino acids, which are followed by the sequence numbers. As shown in the figure, minor chemical shifts are present for most residues, and some are more substantial than the others, such as I7, H49, T51 and T52, which are circled in orange. The chemical shifts changes are plotted on a 3D-model that exemplifies the structure changes of PAI-H19Y caused by the change of concentration, shown in left panel in Figure 14. The right panel displays the calculated chemical shifts changes in each residues $\Delta\delta$ using formula, $\Delta\delta = \sqrt{(\Delta H)^2 + 0.15 * (\Delta N)^2}$, where ΔH is the change of the proton chemical shifts between samples with the protein concentration of 1.25 mM and 0.02 mM, and ΔN is the respective change of the nitrogen shift.

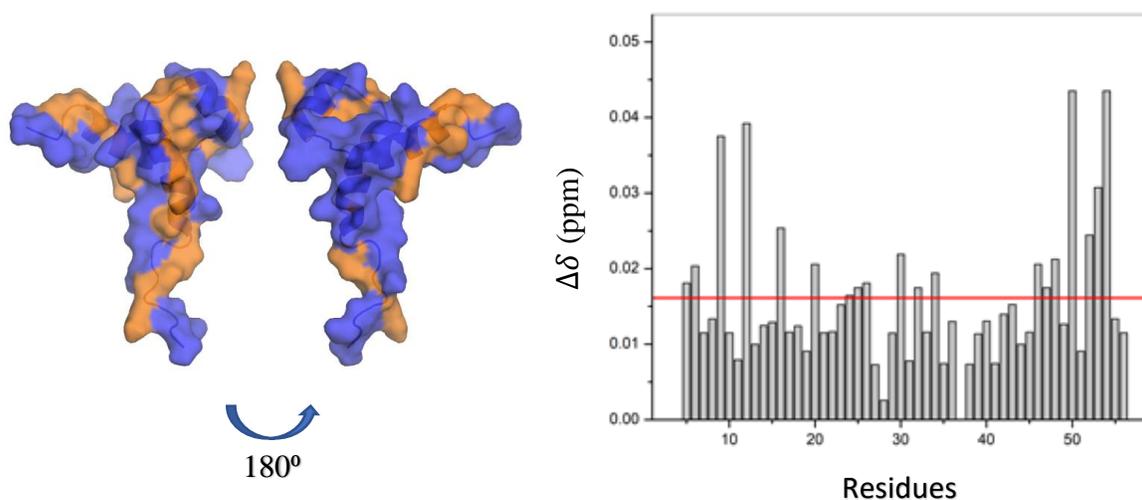


Figure 14. Left: 3D model that reflects conformation changes of PAI-H19Y between 1.25 mM to 0.02 mM at pH 5. The residues with higher than average chemical shifts change are marked in orange. The model is rotated 180 degrees about y-axis. Right: Chemical shifts changes for all residues. Red line is the average chemical shifts change.

We selected four peaks that displayed most significant chemical shifts changes, and plotted their normalized proton chemical shifts change against the protein concentration.

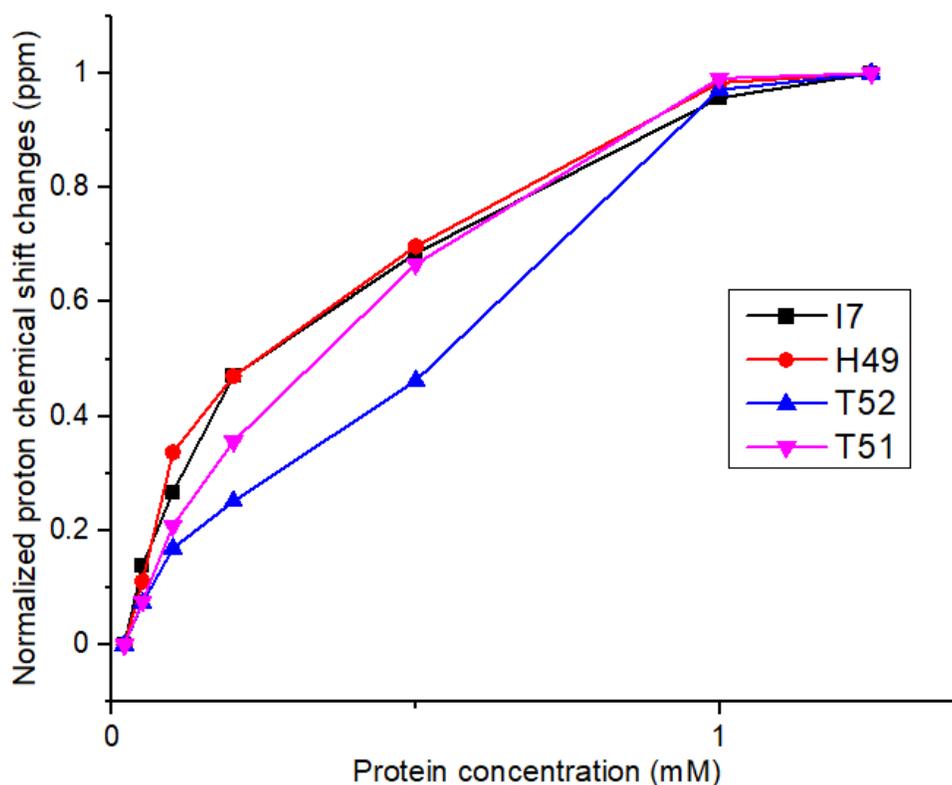


Figure 15. Normalized proton chemical shifts changes at pH 5 is plotted against PAI-H19Y concentration for residues I7, H49, T51 and T52.

From Figure 15, it's clear that the change was more pronounced at lower protein concentration, although the chemical shifts continue to change at higher protein concentrations.

We carried out a similar NMR experiment at pH 7 at protein concentration from 1.25 mM to 0.01 mM. We received a result similar to what we had at pH 5, with minor chemical shifts for most of the residues, and strong chemical shifts for few. The chemical shifts changes are shown on the 3D-model which reflects the conformation changes. The chemical shifts changes $\Delta\delta$ were calculated for all residues. The result is shown in Figure 16.

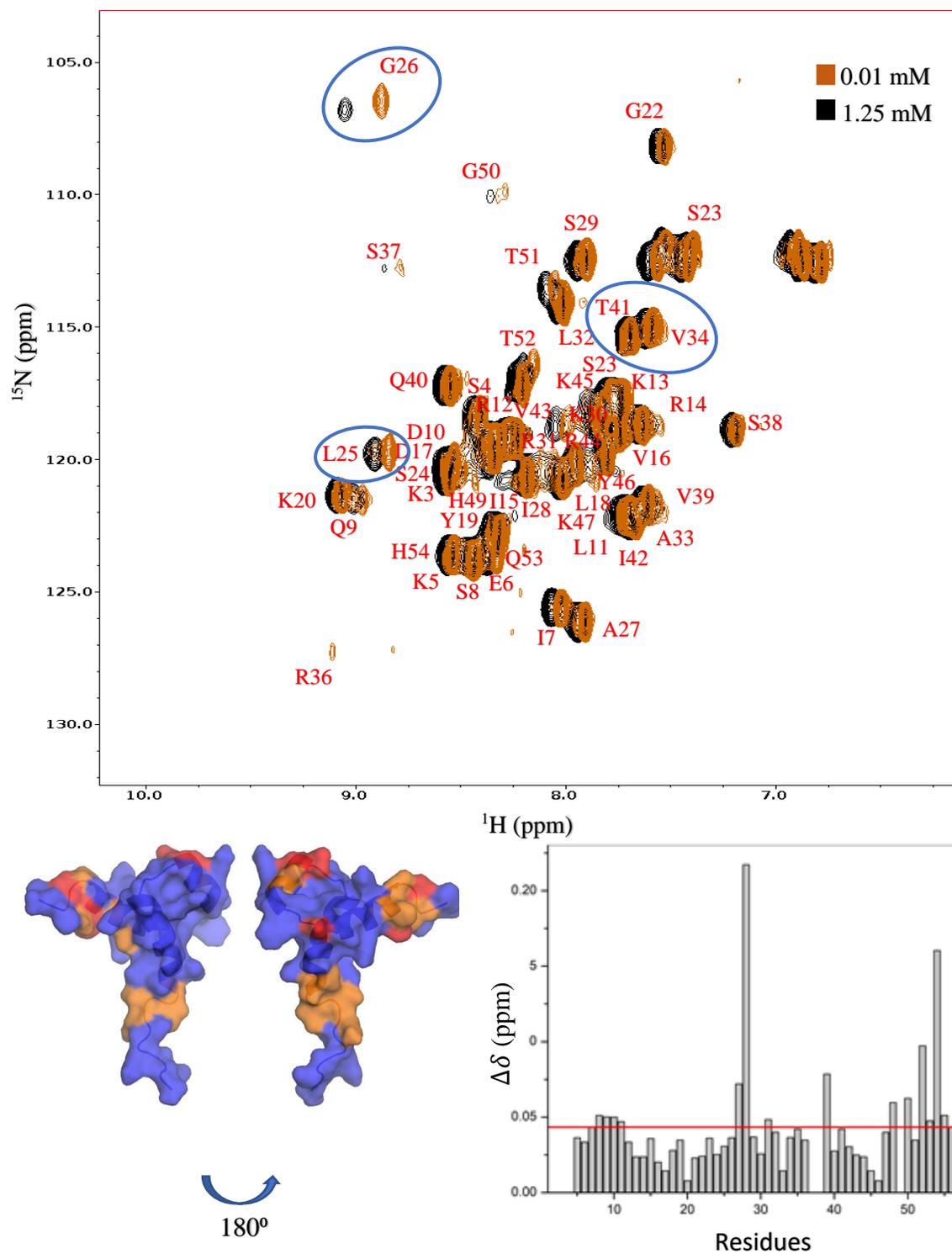


Figure 16. Top: $[^1\text{H}, ^{15}\text{N}]$ -HSQC spectra for ^{15}N -labeled H19Y at concentrations of 1.25 mM and 0.01 mM. Four residues with large chemical shifts changes are circled in blue. Buffer condition: 25mM sodium phosphate buffer with 150 mM NaCl at pH 7 at 25 °C. Bottom left: 3D model that reflect structural changes of PAI-H19Y between 1.25 mM to 0.01 mM at pH 7. The residues with higher than average chemical shifts changes are marked in orange, the residues with outstanding changes are marked in red, such as G26 and T52. The model is

rotated 180 degrees about y-axis. Bottom right: Chemical shifts changes for all residues. Red line is the average chemical shifts.

According to the 3D-model mapping the chemical shifts changes at pH 7, the area with the conformational changes were different than pH 5, in addition to the strong chemical shifts change observed from G26 which was not present at pH 5, suggesting the conformational change at pH 7 were contacting through a different interface. Residues G26, T41, V34 and L25 were picked to display the change in conformation at different protein concentration (circled blue in Figure 16). As shown in Figure 17, the changes were most substantial between 0.2 mM and 0.5 mM protein concentration. As compared to pH 5, the chemical shift changes plateau above 0.6 mM protein concentration.

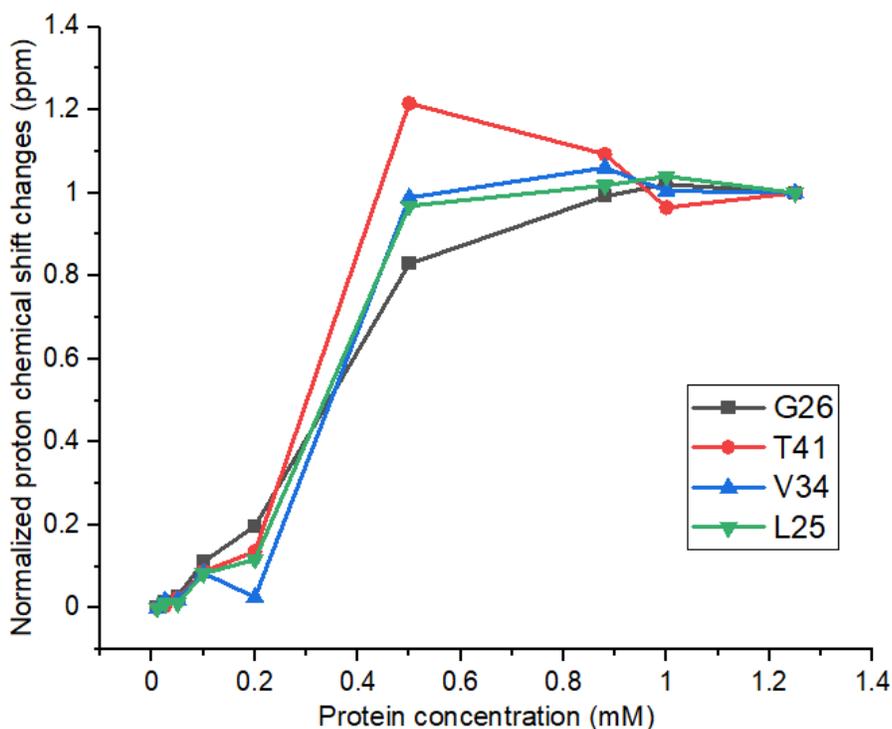


Figure 17. Normalized proton chemical shifts changes at pH 7 is plotted against PAI-H19Y concentration for residues G26, T41, V34 and L25.

5.2. The association of PAI-H19Y by MST

5.2.1. Binding Affinity experiments using Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation

First, we optimized experimental conditions for the MST experiment using Monolith His-Tag Labeling Kit. In order to determine the binding affinity of RED-tris-NTA label to PAI-H19Y mutant, we carried out a titration experiment by adding increasing concentrations of protein (maximum concentration 500 nM PAI-H19Y) to the solution of label dissolved at the concentration of 25 nM in 25 mM sodium phosphate with 150 mM NaCl and 0.05% Tween buffer. The result is presented in Figure 18.

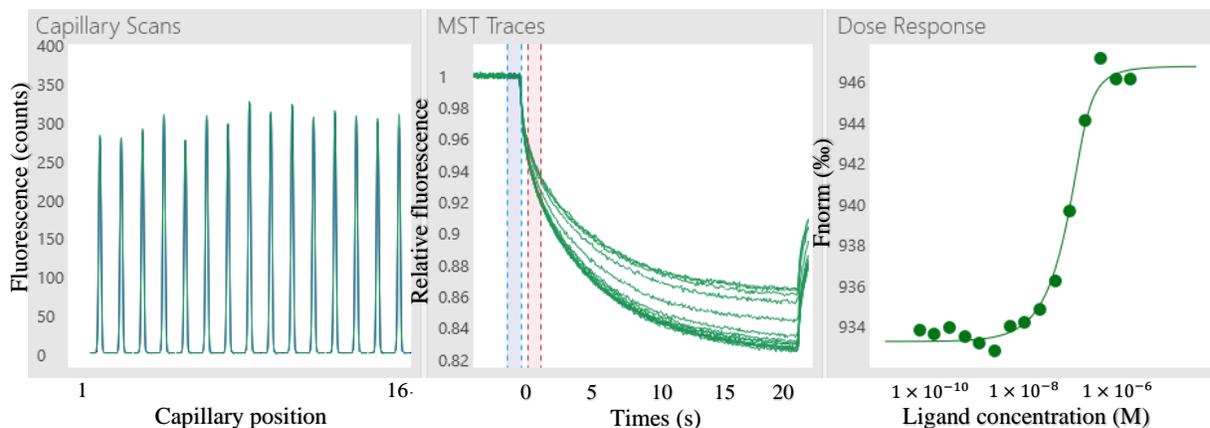


Figure 18. Binding affinity experiment between Monolith His-Tag Labeling Kit and PAI-H19Y, $K_d = 3.4 \pm 1.3$ nM nM.

Left panel shows the Capillary Scans, providing fluorescence intensity and distribution, sharp Gaussian distribution indicated the sample was evenly spread and no adsorption to capillary walls. Middle panel shows MST Traces displaying the change of fluorescence overtime, the bound state is represented by the curved-up traces. The Right panel shows fluorescence at different ligand concentration, the curve formed by the change of fluorescence indicated the binding process. The binding constant (K_d) of RED-tris-NTA

label to protein was found to be equal to 3.4 ± 1.3 nM. Since the affinity is stronger than 10 nM, we used dye to protein ratio of 1:2 for protein labeling as suggest in the protocol [19].

To investigate the affinity of protein-protein interactions, labeled protein (target) kept at a constant concentration was mixed with varying concentrations of unlabeled protein (ligand). A total of 16 different dilutions were made and the samples were loaded into premium capillaries. The target concentration in all 16 capillaries was 30 nM, and the ligand concentration ranged from 1 mM to 30 nM (diluted by half for each capillary). Figure 19 exemplifies a typical titration experiment that we carried out to measure the binding constant of PAI-H19Y mutant association.

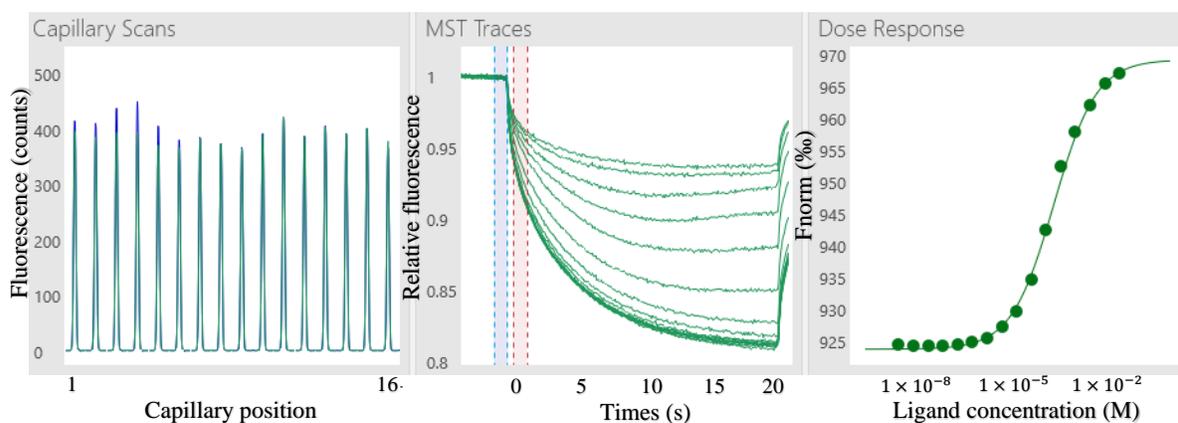


Figure 19. Binding Affinity experiment using His-tag labeling dye, with target concentration 30 nM and ligand concentration from 1 mM to 30 nM, at 25 °C and pH 5. $Kd = 34.3 \pm 9.6$ μ M.

As shown in Figure 19, all Capillary Scans showed sharp Gaussian distribution indicated no adsorption to capillary walls and the molecules were evenly spread. MST traces changed with the concentration of added ligand. The irregularity in some of the traces at high protein concentrations and longer heating times was likely due to protein aggregation. A typical dose-response curve indicated the binding between the target and the ligand was

present. The experiment was repeated at 35 °C and showed a similar result. The binding affinity was determined to be $34.3 \pm 9.6 \mu\text{M}$.

This MST result shows that the binding constant for PAI-H19Y is smaller than what would be expected from NMR experiments (Fig. 15 & 17). Our initial explanation was the possible presence of the free dye in labeled sample, which could bind to unlabeled protein. Therefore, to test this hypothesis, we repeated the MST experiment using an adjusted procedure. After the protein was labeled, the sample was washed extensively through ten runs of buffer exchange to remove the free dye. In each run, the sample was diluted five times then centrifuged using the Amicon-3K filter, a total of 10 million times dilution for the free dye. Our previous procedure only consisted of three runs of buffer exchange (125 times dilution). If the binding was indeed cause by the free dye, we should no longer observe any binding interaction.

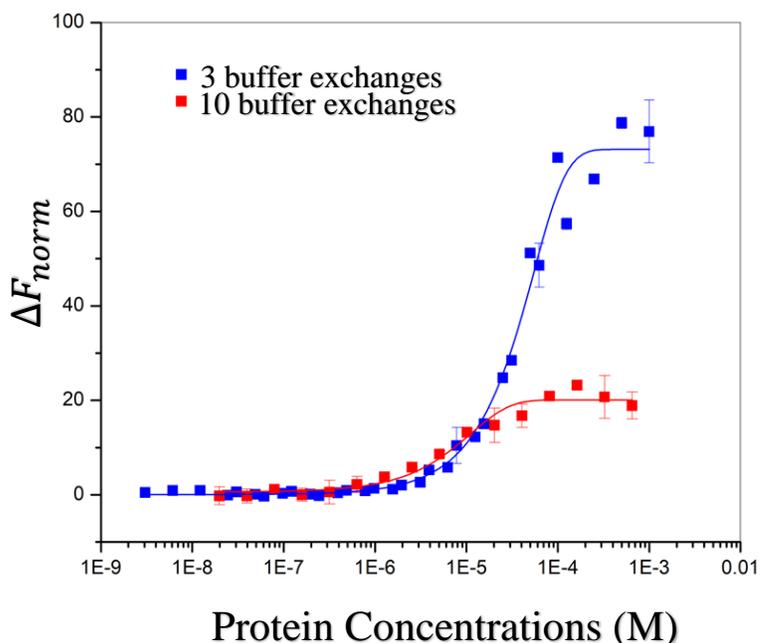


Figure 20. Dose Response results for PAI-H19Y with His-tag labeling dye, with three (blue) and ten (red) runs of buffer exchange, five times dilution for each run. Buffer condition: 25 mM sodium phosphate with 150 mM NaCl and 0.05% Tween at pH 5 at 25 °C and 35 °C.

However, as shown in Figure 20, after extensive wash of 10 million times dilution, the red curve indicated that the binding was still present, and the binding affinity remained the same compare to the result using normal wash (125 times dilution) shown in blue. In contrast, the fluorescence signal for the fully bound state had decreased drastically using the extensive wash, consistent with dye removal from the sample.

To examine if the binding was directly related to the His-tag labeling dye, we repeated the Binding Affinity experiments at pH 7. The comparison between results collected at pH 7 and pH 5 is displayed in Figure 21.

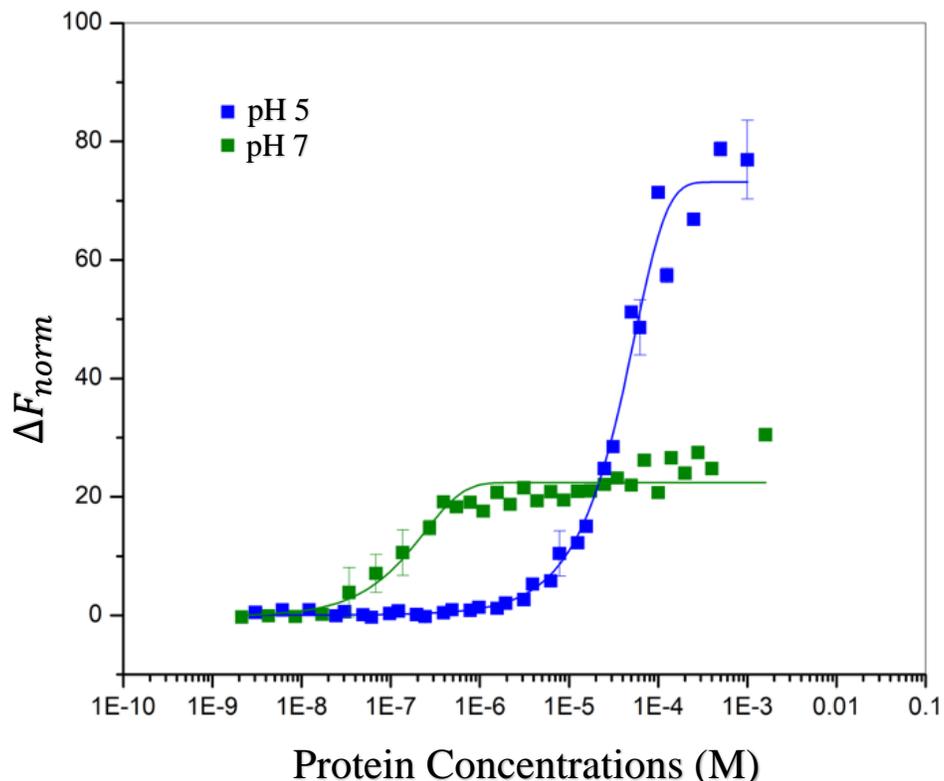


Figure 21. Dose Response results for PAI-H19Y with His-tag labeling dye, at pH 5 (blue) and at pH 7 (green). Buffer condition: 25 mM sodium phosphate with 150 mM NaCl and 0.05% Tween at pH 5 and pH 7 at 25 °C and 35 °C. Temperature change had little effect on the binding affinity compare to the pH change.

As shown in the figure, the protein-protein interaction at pH 7 (green curve) had shifted to a significantly lower concentration compare to pH 5 (blue curve), which means that the binding affinity at pH 7 was significantly stronger than at pH 5. Considering that histidine side-chain was no longer positively charged at pH 7, the binding affinity of dye to His-tag is expected to be stronger at pH 7, supporting the idea that the binding interaction we observed was the binding between dye and histidine. Since the component Ni-NTA has two sites available for binding to histidine [23], the presence of the unlabeled protein at high concentration could have accelerated this process, resulting in the dye induced protein-protein interaction.

In the future, we can test this theory by removing the His-tags from the unlabeled protein before conducting the experiment, the removal could be achieved through a selective protease.

5.2.2. Binding Affinity experiments using Monolith Protein Labeling Kit RED-NHS 2nd Generation

Seeing that His-tag labeling could induce protein-protein interaction, we decided to repeat the MST experiments using a different fluorescent label, Monolith Protein Labeling Kit RED-NHS 2nd Generation. This label reacts with lysine residues instead of His-tags and covalently attaches to protein. The procedure requires the removal of free dye after labeling. The labeling process was achieved with the dye to protein ratio of 3:1 according to the protocol [21]. The free dye was removed through extensive buffer exchange.

The Binding Affinity experiments were conducted at the same buffer conditions as the previous MST experiments. The results are presented in Figure 22.

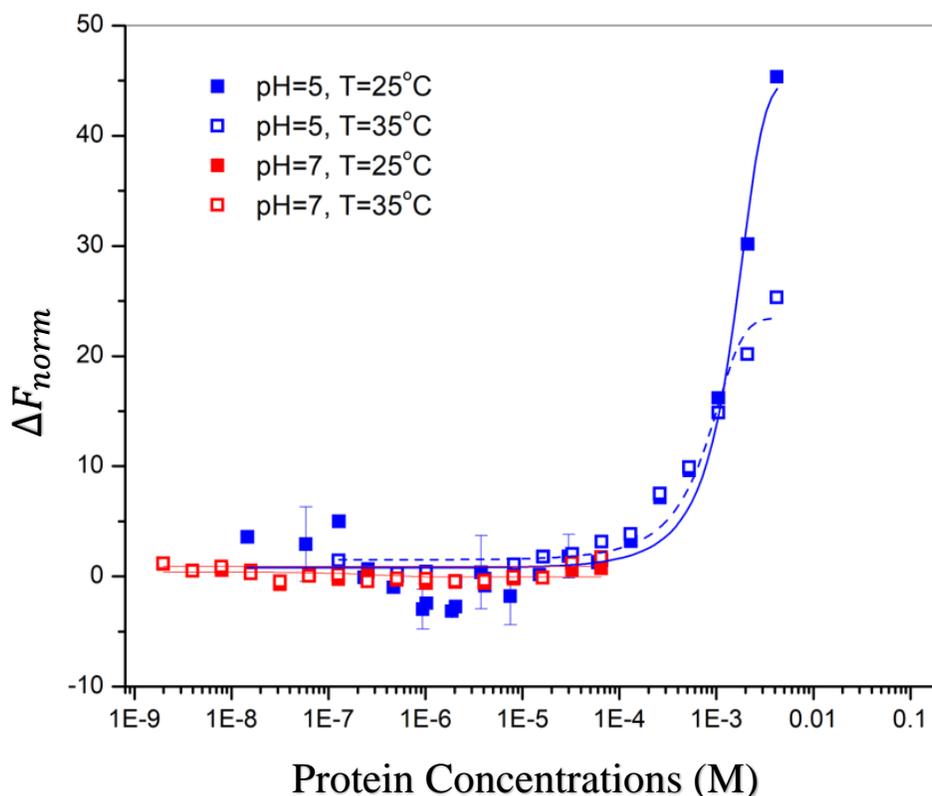


Figure 22. MST results for PAI-H19Y with lysine labeling dye. Buffer condition 25 mM sodium phosphate with 150 mM NaCl and 0.05% Tween at pH 5 and pH 7 at 25 °C and 35 °C.

Figure 22 shows that the curves for both pH values of 5 and 7 acquired at both temperatures of 25 °C and 35 °C demonstrate a similar trend. As seen in all curves, no binding is observed in the range of concentrations where we observed binding using the His-tag label. The binding is observed at higher protein concentration compare to the His-tag labeled result. The binding affinity at 25 °C was determined to be 2.7 ± 0.7 mM, and the binding affinity at 35 °C was determined to be 1.2 ± 0.4 mM. This result suggested that there could be binding interaction at higher concentration, but not at μ M scale as we saw before. Therefore, the MST result for lysine labeled protein is in a better agreement with the NMR result, also showing that the binding constant can be expected in sub-millimolar range

5.3 The association of PAI-H19Y by crosslinking

To further examine the protein-protein interaction, we used chemical cross-linking. The first crosslinker used was EDC (Fig. 10). Before adding the crosslinker, the sample was prepared in two different buffers, the MES buffer (10 mM MES, 50 mM NaCl, pH 5) and the sodium phosphate buffer (25 mM sodium phosphate, pH 5), each having three protein concentrations, 0.2 mM, 0.1 mM and 0.05 mM (six samples in total). Freshly made EDC was added into the samples reaching 2 mM final concentration, and the mixture was incubated for 2 hours at room temperature. The reaction was stopped by adding β -mercaptoethanol reaching 20mM final concentration. After centrifugation, the supernatant was loaded onto 16.5% SDS-PAGE gel (Fig. 23).

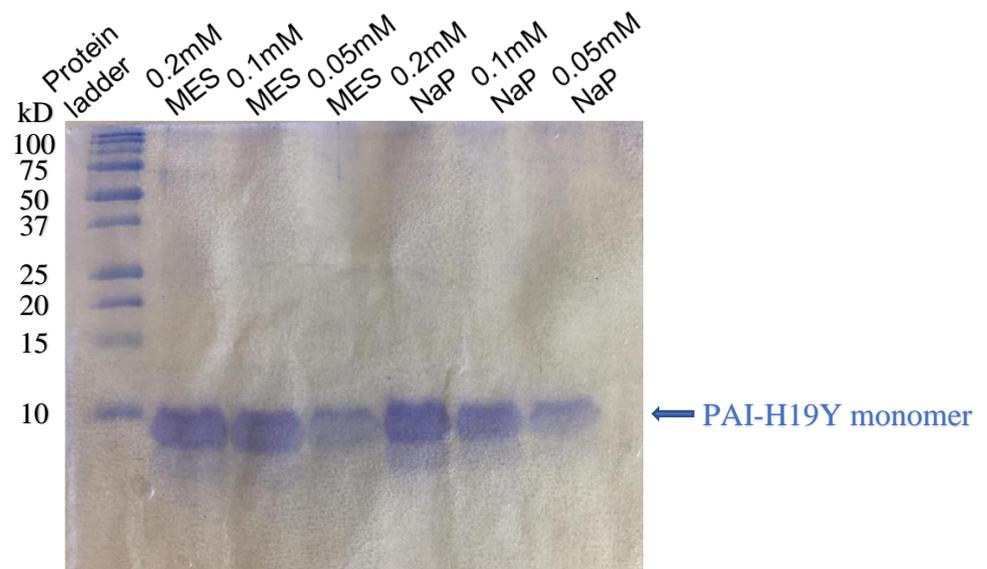


Figure 23. SDS-PAGE gel result for EDC crosslinking, with EDC concentration of 2 mM, protein concentrations of 0.2 mM, 0.1 mM and 0.05 mM, in MES (10 mM MES, 50 mM NaCl, pH 5) and sodium phosphate buffer (25 mM sodium phosphate, pH 5).

Shown in Figure 23 is the SDS-PAGE gel result for EDC crosslinking experiment, the columns from left to right are: Protein ladder, 0.2 mM protein in MES buffer, 0.1 mM

protein in MES buffer, 0.05 mM protein in MES buffer, 0.2 mM protein in sodium phosphate buffer, 0.1 mM protein in sodium phosphate buffer, 0.05 mM protein in sodium phosphate buffer. As clearly displayed in the figure, only one band is seen for all six samples, indicating that only the protein monomer is present.

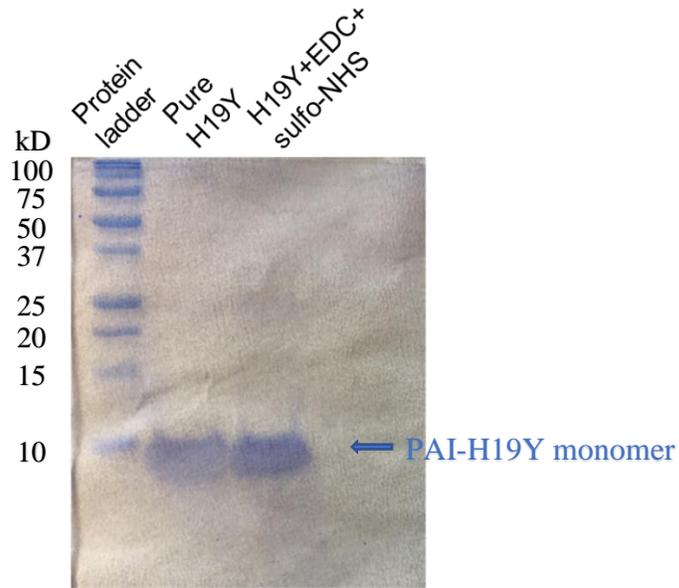


Figure 24. SDS-PAGE gel result for EDC plus sulfo-NHS crosslinking, with EDC concentration of 1 mM, sulfo-NHS concentration of 2.5 mM, PAI-H19Y concentration of 1 mg/mL (0.14 mM), in the mixer of activation buffer (0.1 M MES, 0.5 M NaCl, pH 6.0) and coupling buffer (100 mM sodium phosphate, 150mM NaCl, pH 7.2).

EDC can also be used together with sulfo-NHS for its stabilizing effect (Fig. 10). The protein was prepared in two separate buffers at 1 mg/mL: activation buffer (0.1 M MES, 0.5 M NaCl, pH 6.0) and coupling buffer (100 mM sodium phosphate, 150mM NaCl, pH 7.2). The crosslinking starts by adding EDC and sulfo-NHS into the activation buffer protein to a final concentration of 2 mM EDC and 5 mM sulfo-NHS. After 15 minutes, the reaction was quenched using 20 mM 2-mercaptoethanol. The couple buffer protein was then added to the activated protein and reacted for 2 hours. The reaction was quenched by adding 1 M Tris·HCl to a final concentration of 50 mM Tris. The SDS-PAGE gel result for this

experiment is shown in Figure 24. The columns from left to right are: protein ladder, protein without crosslinker and protein with EDC and sulfo-NHS. Similar to the crosslinking with EDC only, the band for protein monomer is present, but no higher order oligomers are seen. Considering the EDC crosslinker has zero spacer arm lengths and reacts with carboxylic acid and primary amine groups, it may not cross-link PAI-H19Y molecules as the interaction interface may not contain the necessary residues, it may also require a higher protein concentration.

The second crosslinker we used was BS3 (Fig. 11). The buffer was 25 mM sodium phosphate buffer at pH 7. Three samples were prepared with protein concentration at 0.05 mM and BS3 concentration at 1 mM, since higher protein concentration had resulted in heavy precipitation. The reaction time for the three samples were different at 0.5 hours, 2 hours and 12 hours respectively. The reaction was quenched using 50 mM Tris. After centrifugation, the supernatant was loaded onto 16.5% SDS-PAGE gel (Fig. 25).

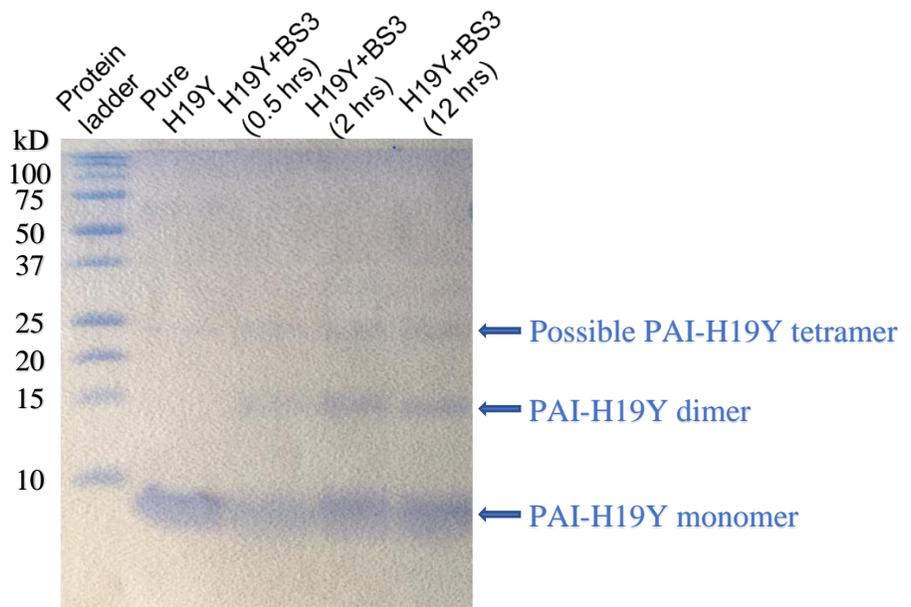


Figure 25. SDS-PAGE gel result for BS3 crosslinking, with BS3 concentration of 1 mM, protein concentration of 0.05 mM, in 25 sodium phosphate buffer at pH 7.

The five columns from left to right are: protein ladder, PAI-H19Y without crosslinker, PAI-H19Y with BS3 for 0.5 hours, PAI-H19Y with BS3 for 2 hours, PAI-H19Y with BS3 for 12 hours. As shown in the figure, without BS3 crosslinker, only monomer band was apparent, after reacting with the crosslinker, band representing the protein dimer had formed above the monomers. Another faint band can also be seen above the dimer, indicating possible protein tetramer. In addition to the longer spacer arm lengths, the oligomerization induced by BS3 crosslinker was likely due to the high lysine count in PAI-H19Y, with seven lysine residues in 61 amino acids (11.5%). Although we observed the presence of PAI-H19Y dimers and possibly tetramers, their relative amount is much smaller than that of PAI-H19Y monomer, suggesting that only small fraction of protein dimerizes, and significant amount of crosslinker and time is required to induce dimerization (crosslinker to protein ratio of 1:20 was used in this experiment).

5.4. DNA-binding affinity to PAI-H19Y determined by MST

As mentioned in the Motivation section, part of our goal was to explain the NMR peak broadening happened during the NMR experiment (Fig. 5). Therefore, we proceed to DNA-binding experiment attempting to determine the binding constant of DNA to H19Y.

5.4.1. Optimization of experimental condition for DNA-binding with PAI-H19Y by MST

During the MST experiment, one important parameter to monitor is the Capillary Scans, which provides information about how well the sample is distributed within the capillaries and the intensity of the fluorescence. The labeled DNA used was 18 base pair (bp) DNA-core sequence 3'-TCGGAAGTTTACATACAC-5' synthesized by IDT (Integrated DNA Technologies, Inc.), which is the minimum required DNA sequence for SB transposase binding and the common sequence between the four transposase binding sites on the

transposon DNA [22]. We firstly performed the Pretest experiment, where the labeled DNA was loaded into the capillaries without the protein, the result is displayed in Figure 26.

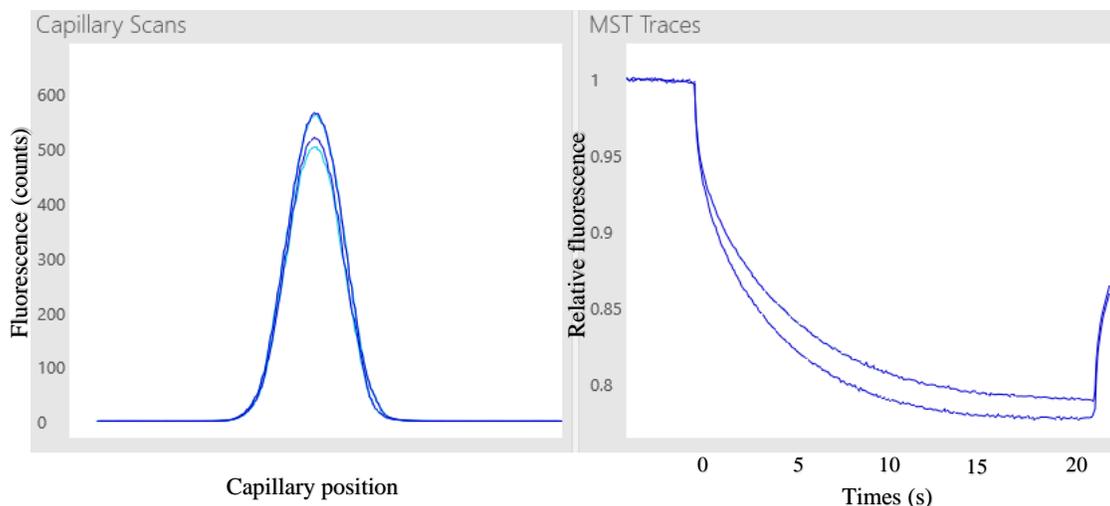


Figure 26. MST result for Pretest with DNA at 0.4 nM, in 25 mM sodium phosphate with 0.05% Tween at pH 5 and 25 °C.

Shown on the left in Figure 26 is the Capillary Scans, where two capillaries both loaded with labeled DNA were scanned two times each and with scans overlapped. This result showed that the intensity of the fluorescence was around 550 counts, and the fluorescent molecules were evenly distributed inside the capillaries, judged by the sharp Gaussian distribution. The fluorescence intensity shows the concentration of the fluorescent target, as shown in the figure, there were slight variations between the two capillaries, it was likely due to the sample not being completely homogeneous, but as long as the deviations were within 20% and the magnitudes are between 200-2000 counts, it shouldn't affect the determination of the binding constant. The intensity variation can also be seen in MST Traces on the right.

We then performed the Binding Check experiment, where four capillaries with DNA only and four capillaries with DNA-protein mixture were scanned. The result is shown in Figure 27.

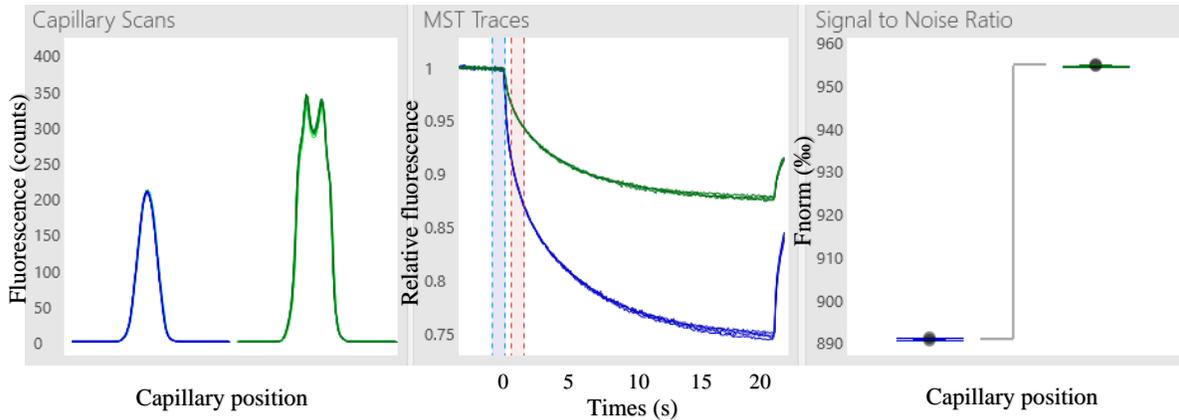


Figure 27. MST result for Binding Check for DNA (blue) and PAI-H19Y (green), with DNA concentration of 30 nM, protein concentration of 50 nM, in 25 mM sodium phosphate with 0.05% Tween at pH 5 and 25 °C.

In Figure 27, blue lines represent the DNA scans and green lines represent the DNA-protein scans. One clear discrepancy in Capillary Scans is the shape of the DNA-protein scans, instead of a sharp Gaussian distribution, it has a pincer like shape with two peaks on the sides, indicating a very strong fluorescence adsorption to capillary walls, this should be prevented since proteins might show altered functionality due to surface effects. The MST Traces result in the middle clearly shows the bounded states for the DNA-protein sample, judged by the up curving green traces, and the large Signal to Noise ratio on the right is a strong indication of the DNA-protein binding, proving the effectiveness of the Binding Check experiments.

Our first attempt to solve the adsorption problem was switching to the Premium Capillaries, compare to the Standard Capillaries we had been using, Premium Capillaries are

covalently coated with a dense brush of a specially designed polymer to avoid molecule adsorption [14]. The result is shown in Figure 28.

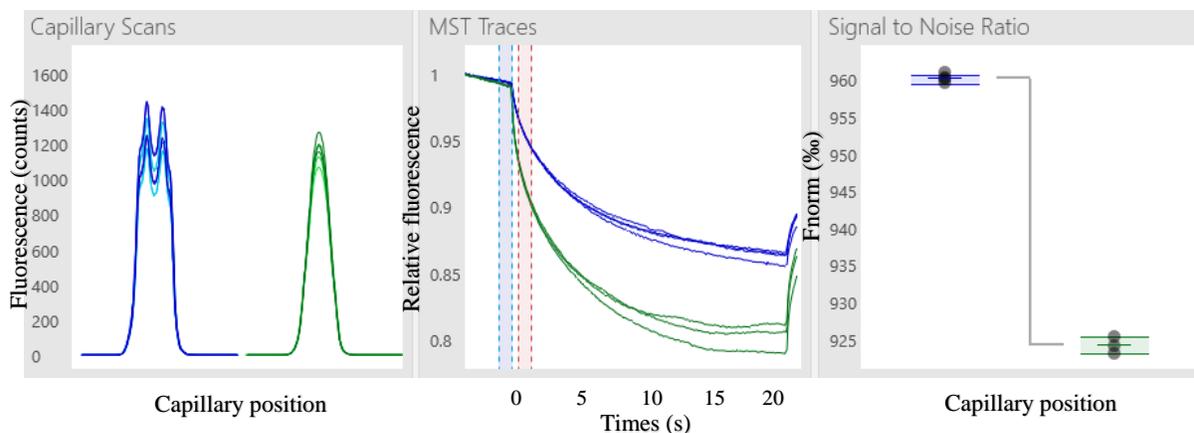


Figure 28. MST result for Binding Check for DNA (blue) and PAI-H19Y (green) using Premium Capillaries, with DNA concentration of 30 nM, protein concentration of 50 nM, in 25 mM sodium phosphate with 0.05% Tween at pH 5 and 25 °C.

Strangely, after we switched to the Premium Capillaries, instead of the protein, the DNA (blue) had become strongly adsorbed to the capillary walls as shown in Figure 28. Our solution was adding detergents to reduce surface tension and increase miscibility, We found that the presence of 0.05% Tween and 150 mM NaCl was enough to significantly reduce sample adsorption to capillary walls. The effectiveness of the detergent can be seen in Figure 29, where the Capillary Scans for both DNA and DNA-protein have obtained the shape of Gaussian distribution. All MST DNA-binding experiments were performed at this buffer condition.

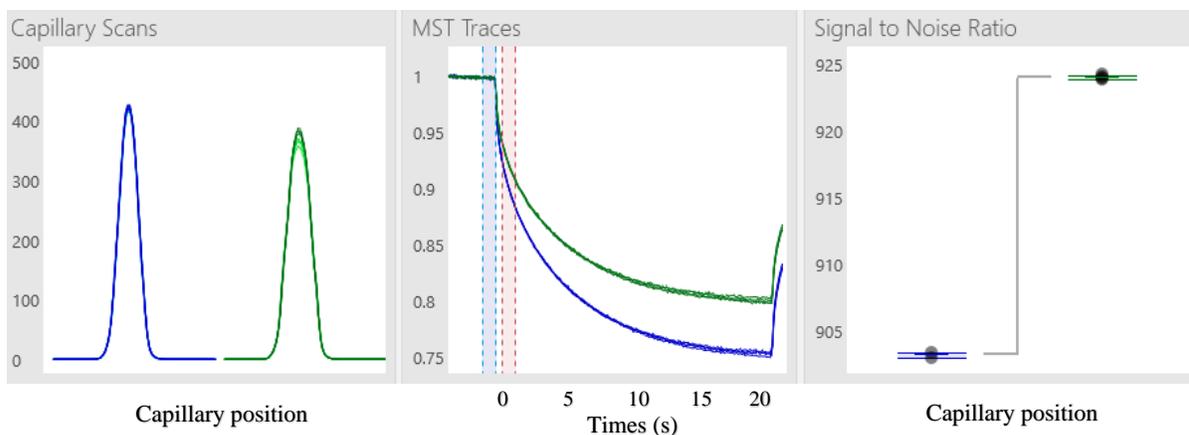


Figure 29. MST result for Binding Check for DNA (blue) and PAI-H19Y (green) with DNA concentration of 30 nM, protein concentration of 50 nM, in 25 mM sodium phosphate with 0.05% Tween and 150 mM NaCl at pH 5 and 25 °C.

5.4.2. Binding Affinity result for DNA-binding with PAI-H19Y

We performed the Binding Affinity experiments using the same experimental condition as we used in the NMR DNA-binding experiment (25 mM sodium phosphate buffer with 150 mM NaCl at pH 5 and pH 7) in the addition of 0.01% Tween as detergent. All Dose Response results are plotted onto a single graph shown in Figure 30.

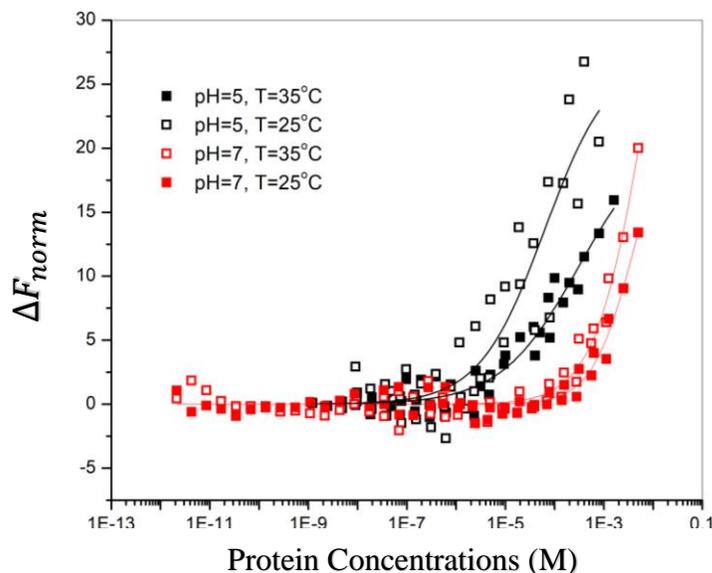


Figure 30. MST results for DNA-binding experiments with PAI-H19Y. Buffer condition: 25 mM sodium phosphate buffer with 0.1% Tween and 150 mM NaCl at pH 5 and pH 7 at 25 °C and 35 °C.

As shown in Figure 26, the binding of DNA to protein is observed at all conditions. Although, even at the highest ligand concentration of 1.6 mM, the plateau where all PAI-H19Y would be in bound states with DNA is not reached. Therefore, we can only estimate that the binding constant for DNA-protein is in sub-millimolar range. This result also proved the stabilizing effect of the H19Y mutation, judged by the stronger binding affinity showed at pH 5 for PAI-H19Y. Since without the mutation, PAI-WT can only bind to DNA at pH 7 and is unfolded at pH 5.

CHAPTER 6: DISCUSSION

MST experiments had been the most time consuming partially due to our unfamiliarity with the method, along with the trials and errors needed to identify the proper experimental condition. Using His-tag labeling dye, we observed some form of binding interaction during Binding Affinity experiment and determined the binding constant to be $34.3 \pm 9.6 \mu\text{M}$. (Fig. 19). After extensively washing the sample, we still observed the binding with same binding constant, thus eliminated the possibility that the binding was caused by the free dye. We then repeated the MST experiments at pH 7, seeing the increase in binding affinity of protein-protein interaction, we concluded that the protein-protein interaction was induced by the dye, since binding affinity of dye to His-tag is stronger at pH 7, and one dye molecule can bind to two or more His-tags. Therefore, we performed the MST experiment again but using lysine labeling dye. The result was in a better agreement with NMR result, showing the presence of protein-protein interactions at sub-millimolar concentrations.

We then used crosslinking as a different approach to confirm that the protein-protein interaction may occur. We firstly used EDC crosslinker, which forms a zero-length bound between two carboxylic acid or primary amine groups. After reaction with crosslinker, the sample was then loaded onto the SDS-PAGE gel. Judged by the presence of protein monomer band only, the gel result showed no protein oligomerization (Fig.23). Knowing that sulfo-NHS can improve efficiency or create dry-stable (amine-reactive) intermediates, we also performed EDC plus sulfo-NHS crosslinking experiment, but similar to EDC only experiment, we saw no protein oligomerization (Fig. 24). Thus, concluding that EDC does not induce protein-protein interaction at 0.2 mM, 0.1 mM and 0.05 mM protein

concentration. We then used BS3 crosslinker, which reacts with the primary amines on the side chain of lysine residues and the N-terminus of polypeptide chains. BS3 has a 11.4 Angstrom spacer arm length, providing more freedom for the bound to form. As shown in Figure 25, the gel result did show oligomerization of the protein, with dimer and tetramer band present, which was likely due to the high lysine count in PAI-H19Y. The faint band of the dimer and tetramer suggested that only small fraction of protein oligomerize, and significant amount of crosslinker and time is required to induce oligomerization.

Finally, we proceed to the DNA-binding experiments, determine the protein-DNA binding affinity. Through binding affinity experiments, we observed the signal increase started around 1 mM protein concentration and continued passing 1.6 mM. Therefore, we can only estimate that the binding constant for DNA-protein is in sub-millimolar range.

Collectively, our results suggest that there is a PAI-H19Y association that starts at sub-millimolar protein concentrations. To determine the exact binding constant of PAI-H19Y association, future NMR experiments at 35 °C and His-tag labeled MST experiments could be helpful. The MST experiments should be carried out by adding His-tag free protein to eliminate the possibility of binding multiple tags by one label. Additionally, the experiments should be carried out at higher protein concentrations as suggested by the lysine labeled MST experiments and crosslinking experiments.

CHAPTER 7: CONCLUSION

In this study, we examined the protein-protein interaction of the PAI-H19Y subdomain of the Sleeping Beauty transposase using three techniques, microscale thermophoresis (MST), nuclear magnetic resonance (NMR) spectroscopy and crosslinking. His-tag labeled MST results showed binding interaction with binding constant of 34.3 ± 9.6 μM . However, the binding was induced by the His-tag labeling dye. Lysine labeled MST results and NMR results all showed the presence of protein-protein interactions at sub-millimolar concentrations. EDC and EDC plus sulfo-NHS crosslinking experiment did not induce protein-protein interaction at 0.2 mM, 0.1 mM and 0.05 mM protein concentration. BS3 crosslinker did induce oligomerization, resulted in both protein dimer and tetramer. Although only small fraction of protein oligomerized, and significant amount of crosslinker and time is required to induce oligomerization. MST DNA-binding results suggested that the binding constant for DNA-protein is also in sub-millimolar range. Future NMR experiments at 35 °C and His-tag labeled MST experiments are needed to determine the exact binding constant of PAI-H19Y association.

REFERENCES

1. Jones, C. H., Chen, C. K., Ravikrishnan, A., Rane, S., & Pfeifer, B. A. (2013). Overcoming nonviral gene delivery barriers: perspective and future. *Molecular pharmaceutics*, *10*(11), 4082–4098.
2. Hackett, P.B., Largaespada, D.A. and Cooper, L.J. (2010) A transposon and transposase system for human application. *Mol Ther*, **18**, 674-683.
3. Aronovich, E. L., McIvor, R. S., & Hackett, P. B. (2011). The Sleeping Beauty transposon system: a non-viral vector for gene therapy. *Human molecular genetics*, *20*(R1), R14–R20.
4. Nesmelova, I. V., & Hackett, P. B. (2010). DDE transposases: Structural similarity and diversity. *Advanced drug delivery reviews*, *62*(12), 1187–1195.
5. Leighton, G. O., Konnova, T. A., Idiyatullin, B., Hurr, S. H., Zuev, Y. F., & Nesmelova, I. V. (2014). The folding of the specific DNA recognition subdomain of the sleeping beauty transposase is temperature-dependent and is required for its binding to the transposon DNA. *PloS one*, *9*(11), e112114.
6. Izsvák, Z., Khare, D., Behlke, J., Heinemann, U., Plasterk, R. H., & Ivics, Z. (2002). Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in Sleeping Beauty transposition. *The Journal of biological chemistry*, *277*(37), 34581–34588.
7. Carpentier, C.E., Schreifels, J.M., Aronovich, E.L., Carlson, D.F., Hackett, P.B. and Nesmelova, I.V. (2014) NMR structural analysis of Sleeping Beauty transposase binding to DNA. *Protein Sci*, **23**, 23-33.

8. van Zundert, G.C., Rodrigues, J.P., Trellet, M., Schmitz, C., Kastiris, P.L., Karaca, E., Melquiond, A.S., van Dijk, M., de Vries, S.J. and Bonvin, A.M. (2016) The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J Mol Biol*, **428**, 720-725.
9. Muhandiram, D.R. and Kay, L.E. (1994) Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. Magn. Res.*, **B103**, 203-216.
10. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277-293.
11. Keller, R.L.J. (2004) Keller, R.L.J. (2004). The computer aided resonance assignment tutorial. (Cantina Verlag).
12. Johnson, B.A. and Blevins, R.A. (1994) NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603-614.
13. Duhr, S., & Braun, D. (2006). Thermophoretic depletion follows Boltzmann distribution. *Physical review letters*, *96*(16), 168301.
14. Jerabek, W., Moran & André, Timon & Wanner, Randy & Resch, Heide & Duhr, Stefan & Baaske, Philipp & Breitsprecher, Dennis. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*. 10.1016/j.molstruc.2014.03.009.
15. Seidel, S. A., Dijkman, P. M., Lea, W. A., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J. S., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F. A., Ladbury, J. E., Schreiber, G., Watts, A., Braun, D., & Duhr, S. (2013).

- Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods (San Diego, Calif.)*, 59(3), 301–315.
16. Grabarek, Z. and Gergely, J. (1990). Zero-length crosslinking procedure with the use of active esters. *Anal Biochem* 185:131-5.
 17. ThermoFisher Scientific. (2018). DSS and BS³ Crosslinkers. Pub. Part No. 2160418. Retrieved from ThermoFisher website.
 18. Live, D.H., Davis, D.G., Agosta, W.C. and Cowburn, D. (1984) Long-Range Hydrogen-Bond Mediated Effects in Peptides - N-15 Nmr-Study of Gramicidin-S in Water and Organic-Solvents. *J Am Chem Soc*, **106**, 1939-1941.
 19. NanoTemper Technologies. (2019). Monolith His-Tag Labeling Kit RED-tris-NTA 2nd Generation. *Cat# MO-L018*.
 20. Mandal, Pravat & Majumdar, Ananya. (2004). A comprehensive discussion of HSQC and HMQC pulse sequences. *Concepts in Magnetic Resonance Part A*. 20A. 1 - 23. 10.1002/cmr.a.10095.
 21. NanoTemper Technologies. (2019). Monolith Protein Labeling Kit RED-NHS 2nd Generation. *Cat# MO-L011*.
 22. Cui, Z., Geurts, A.M., Liu, G., Kaufman, C.D. and Hackett, P.B. (2002) Structure-function analysis of the inverted terminal repeats of the sleeping beauty transposon. *J Mol Biol*, **318**, 1221-1235.
 23. Khan F, He M, Taussig MJ.(2006) Double-hexahistidine tag with high-affinity binding for protein immobilization, purification, and detection on ni-nitrilotriacetic acid surfaces. *Anal Chem*. 78(9):3072-3079.