QUANTIFYING THE BINDING AFFINITIES OF THE PAI-H19Y MUTANT OF SLEEPING BEAUTY TRANSPOSASE TO DNA

by

Maria Arango

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

2022

Approved by:

_____
Dr. Irina Nesmelova

_____
Dr. Susan Trammell

_____
Dr. Donald Jacobs

ABSTRACT

MARIA ARANGO. Quantifying the Binding Affinities of the PAI-H19Y Mutant of Sleeping Beauty Transposase to DNA (Under the direction of DR. IRINA NESMELOVA)

The Sleeping Beauty (SB) transposon system is the leading and only DNA transposon-based gene-delivery vehicle that has been adapted for human gene therapy. The SB system consists of a transposon DNA flanked by inverted terminal repeats (ITRs) and a transposase enzyme that catalyzes gene transfer. SB DNA transposition, a cut-and-paste mechanism, occurs when two transposase enzymes bind to ITRs of transposon DNA and form a transpososome, in which the transposon DNA is integrated into the target DNA. The transposase enzyme is comprised of a DNA binding domain containing two subdomains, PAI and RED, and a catalytic domain. During transposition, SB transposase binds to different ITRs of the transposon DNA with different affinities, and this difference must be maintained for efficient transposition; however, the values of binding constants are not known. Here we use the PAI-H19Y mutant of the SB transposase as it is structurally stable at biologically relevant experimental conditions. Using several biophysical methods, such as microscale thermophoresis (MST), and nuclear magnetic resonance (NMR) spectroscopy, we were able to quantify the binding affinities of the PAI-H19Y mutant to transposon DNA.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Irina Nesmelova, whose guidance and invaluable patience has helped me immensely throughout my undergraduate and graduate career. My sincere thanks also go to members of my thesis committee, Dr. Susan Trammell and Dr. Donald Jacobs for their valuable suggestions before this defense. I couldn't have gone on this graduate career without the mentorship and support of Dr. Jacobs, who helped me through many academic roadblocks. I also want to thank the members of the Nesmelov/a lab for helping me whenever the need arose and providing important feedback, most notably Venky Ranjan, who answered the most minor questions I had and trained me when I had no prior wet lab experience. I would be remiss in not mentioning my family and friends. Their belief in me has kept spirits and motivation high during this process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

DR	Direct Repeat
HDR	Half Direct Repeat
HSQC	Heteronuclear Single Quantum Coherence
ITR	Inverted Terminal Repeat
MST	Microscale Thermophoresis
NMR	Nuclear Magnetic Resonance
SB	Sleeping Beauty
TRIC	Temperature-Related Intensity Change

Chapter 1. Introduction

The development of almost all human diseases, to a certain extent, is influenced by genetics. When considering the genetic basis of disease, one might think about single gene disorders such as sickle cell anemia, cystic fibrosis, or certain cancers with a heritable component. Although these genetic disorders are individually rare, they account for approximately 80% of rare diseases according to one estimate, of which there are several thousand, ranging between 5000-8000 [1-2]. It has been estimated that rare diseases affect between 3.5% and 5.9% of the global population, equating to nearly 263-446 million persons globally at any time. This translates to roughly 25-30 million in the United States and 18-30 million in Europe [3]. On an individual scale, many rare genetic disorders can become lifelong debilitating conditions and carry devastating health consequences, including premature death. Unfortunately, only 5% of the several thousand rare diseases currently have an effective treatment [4]. Nonetheless, people living with rare diseases are increasingly benefiting from novel therapeutics, some of which are products of cutting-edge technologies currently emerging in medicine [5].

Gene Therapy is a leading medical approach for treating incurable genetic and acquired disorders. This modality focuses on the genetic modification of cells to produce a therapeutic effect and treat diseases by repairing defective genes in patients by introducing foreign genetic material (gene addition) or modifying a specific site of the genome (gene editing). In vivo gene therapy, in which genes are delivered to a patient's cells, has been developed using either viral or nonviral delivery systems. Viral-based delivery systems employ several virus vectors, including adenovirus, retrovirus, herpes virus, and adeno-associated virus (AAV) [6]. Because viruses are naturally evolved vehicles with good cell uptake and intracellular trafficking machinery that

enable efficient gene transfer, they have been widely employed in nearly 70% of clinical trials [7,8]. However, these viral-based therapies have several drawbacks. Safety is a primary concern as this approach may produce adverse effects, including fatal immune responses and oncogenesis. Moreover, their preparation is time-consuming and expensive. In contrast, nonviral approaches, in which nucleic acid cargo is delivered via polymers, lipids, or inorganic materials, offer a better safety profile and are relatively inexpensive. However, this approach suffers from a low delivery rate to target cell nuclei and an unstable integration rate required for long-term therapeutic transgene expression [9-10].

One nonviral vector approach to gene delivery is to harness the natural ability of DNA transposons to move (transpose) from one genomic location to another. The *Sleeping Beauty* (SB) transposon system, the leading DNA-transposon-based gene delivery vehicle adapted for human gene therapy, has an increased transposition rate compared to other transposons [11]. The SB system was derived from the molecular reconstruction of extinct fossil transposase sequences belonging to the Tc1/mariner class of transposons found in the genomes of salmonid fish [12]. Tc1/mariner-like transposons are members of a superfamily of eukaryotic transposons that transpose via a DNA intermediate by a so-called "cut-and-paste" mechanism. The SB system consists of transposon DNA flanked by inverted terminal repeats (ITRs) and a transposase enzyme that catalyzes gene transfer (transposition). SB DNA transposition occurs when two transposase enzymes bind to the ITRs of the transposon DNA, subsequently forming a transposome, excising (cutting) the transposon, and finally reintegrating (pasting) into a target sequence (Figure 1B). The transposase enzyme comprises a DNA-binding domain containing two subdomains (PAI and RED) and a catalytic domain. The PAI subdomain is SB transposase's

primary DNA recognition subdomain, containing a 3-turn helix turn helix (HTH) motif, as shown in Figure 1A [13].

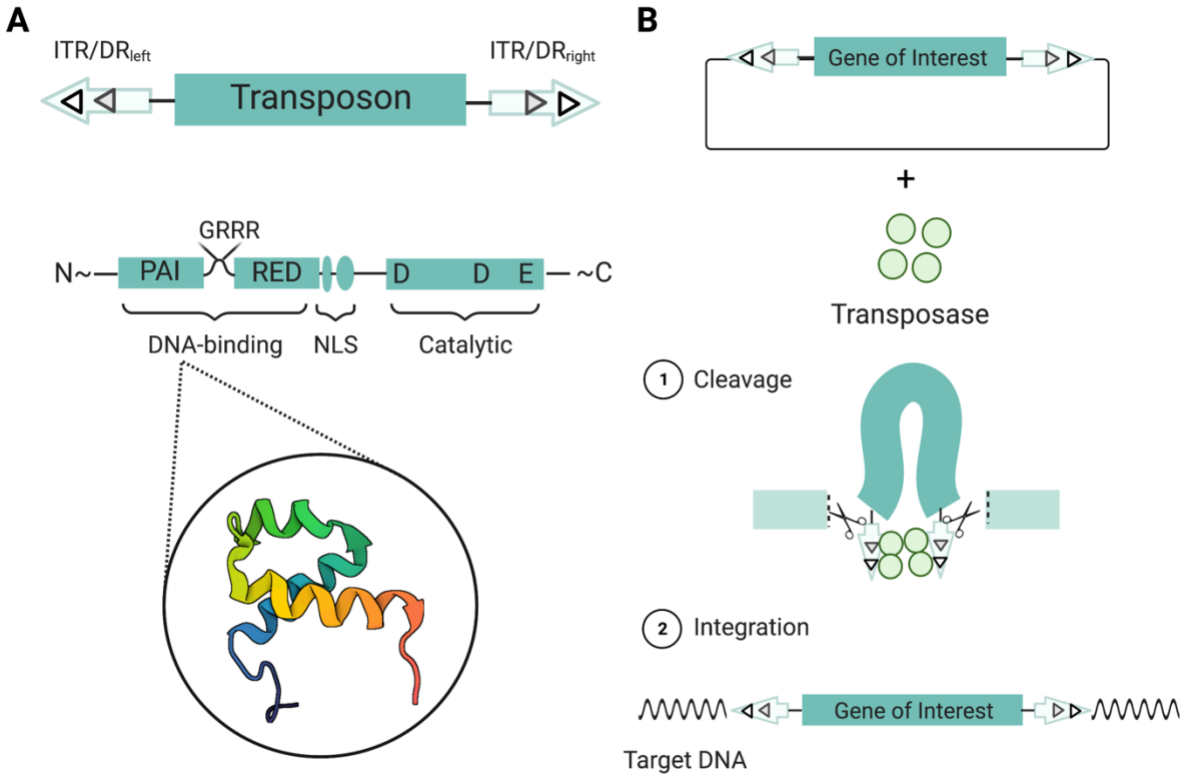

Figure 1. The Sleeping Beauty transposable element and its transposition (A) Components of the element (B) Transposition reaction. Figure created using BioRender illustration software.

In the SB system, ITRs that flank the transposon DNA contain binding sites for the transposase enzyme to recognize and subsequently bind. Each inverted terminal repeat contains two inner and outer direct repeats (DRs). The outer DRs, Lo and Ro, are found at the transposon's left and right termini, respectively, while the inner DRs, Li, and Ri, are located further inside the transposon [13]. The PAI subdomain plays a dominant role in contacting the DRs, associating differently with the inner and outer binding sites of transposon DNA [14]. Although the PAI subdomain is the primary DNA recognition subdomain, it has been shown to bind to DNA only in

its folded conformation. The folded conformation can be achieved under non-physiological environmental conditions, such as low temperature (5 °C), high salt concentration (more than 600 mM NaCl), and pH values greater than 7.0 [15-16]. This problem was addressed in previous research by inducing a structure-stabilizing H19Y mutation in the PAI subdomain, which improves DNA binding activity and transposition [17]. Differences between transposase binding sites are necessary for efficient transposition. However, this requirement is not yet fully understood [18]. To gain a fundamental molecular understanding of this binding difference and to guide the development of more efficient SB transposon-based applications, this work aims to determine the binding sites and affinities of the PAI-H19Y mutant of SB transposase to transposon DNA.

Chapter 2. Topic and Motivation

Since it displayed transpositional activity in vertebrates, the synthetic SB transposon system has risen to become a well-established tool that has been utilized in versatile applications ranging from somatic and germline transgenesis to functional genomics and gene therapy [19]. Before the resurrection of SB from fossil DNA sequences within fish genomes, there was no indication that any DNA-based transposon was active in vertebrates [20]. It has been suggested that this results from selective pressure to avoid insertional mutagenesis of essential genes within their hosts' genomes [21]. Although the SB elements were reconstructed to become sufficiently mobilized in vertebrate cells, their limited transpositional activity presented a hindrance in developing efficient SB transposon-based applications. Nonetheless, multiple efforts have substantially optimized the SB transposon system in terms of transpositional efficiency and flexibility, culminating in the current hyperactive transposase, SB100X, demonstrating a 100-fold increase in transposition activity compared to the original (wild-type) transposase [22]. Regardless of current efforts in developing hyperactive variants with greater transpositional efficiencies, it has been shown that transposase activity is cell type dependent, and applications such as gene transfer into medically relevant human cells can significantly benefit from further enhanced variants [23].

Specific structural and mechanistic information is needed to rationalize novel SB variants. Previous structural studies have shown that SB transposase binds differently to the inner and outer DRs of transposon DNA, and this difference must be maintained for efficient transposition [14,16]. Furthermore, extensive studies done on the ITR/DR interactions with the PAI subdomain provide valuable insight for SB transposition, including (i) All four binding sites within the ITR/DR structure are needed for transposition [23], (ii) the paired-like DNA-binding

domain forms tetramers in complex with transposase binding sites (14), (iii) transposase binds more strongly to inner DRs than to outer DRs [16,18] (iv) PAI subdomain binds to the "half-DR" (HDR) sequence motif within the left IR which mediate protein-protein interactions with other transposase subunits [14]. Thus, the PAI subdomain is postulated to have three distinct functions: interaction with the DRs and HDR motif and transposase oligomerization.

One crucial step during the transposition process is the binding of transposase to its binding sites within the ITR/DRs of the transposon DNA and the subsequent formation of the transpososome (synaptic complex). Transpososome assembly is a process by which two transposon ends are paired and held together by transposase enzymes. The necessary factors for transpososome formation include all four transposase-binding sites within the ITRs, the HDR motif, and tetramer-complex transposase. The tetramer complex forms only if all four binding sites are present and in the correct context. The HDR motif is not necessary but is important as it acts as a transpositional enhancer working in conjunction with the PAI subdomain to stabilize complexes formed by the transposase tetramer bound to the ITR/DR. Most importantly, the observation that SB transposase preferentially binds to inner DRs rather than outer DRs suggests that the unequal strengths of transposase binding and positions of DRs within the IRs are required for the ordered assembly of the transpososome, which in turn affects the outcome of the transposition reaction [24].

Chapter 3. Materials and Methods

Two biophysical techniques, microscale thermophoresis (MST) and nuclear magnetic resonance spectroscopy (NMR), were employed to study the binding sites and affinities of transposase to transposon DNA. Both methods help assess biomolecular interactions; however, each technique has advantages and limitations. For example, MST has several benefits, including low sample consumption, fast experimental execution, relatively straightforward data interpretation, and measurement over a large range of binding affinities (nM to mM). However, MST relies on fluorescently labeling one binding partner, which can alter the binding if the label locates to or near the binding site [25]. Moreover, the small, ~500 μm inner diameter of capillaries used in MST can cause the adsorption of sample material onto capillary walls and promote sample aggregation [26]. In comparison, the diameter of tubes used in NMR spectroscopy is 5 mm, avoiding surface adsorption and aggregation issues. In addition, NMR does not require invasive dyes eliminating dye-dependent effects as seen with MST and can provide a structural image of the binding interactions occurring, yielding information on binding sites [27]. However, despite the strengths of NMR, this technique requires significant time and sample demands, and the determination of binding affinities is not straightforward if the system is not in the fast chemical exchange regime on the NMR time scale [27-29].

3.1 Protein Expression, Purification, and Sample Preparation

The PAI-H19Y mutant of SB transposase was purified and expressed using an established protocol [16]. DNA plasmids encoding the H19YK14K33A mutant His-tagged PAI subdomain of SB transposase were ordered from GenScript USA (Piscataway, NJ). Plasmids encoding the PAI subdomain of SB transposase were transformed into competent BL21-A1 E. coli cells. Cells were

grown in Luria Broth (LB) or M9 minimal media in the presence of 100 μg/mL ampicillin at 30 ^{o}C to an OD600 0.6-0.8 and then induced by adding 0.1 M IPTG and 0.2% L-arabinose for 4 h. For NMR Experiments, ^{15}N-labeled samples were prepared by growing bacterial cells in M9 medium containing $^{15}NH_4Cl$, as the sole nitrogen source. The cells were collected by ultracentrifugation and lysed using B-per lysis buffer (Thermo Scientific). The soluble extract containing protein was prepared by centrifugation of cell lysate at 14,000 g for 1 h. The protein was purified by immobilized metal chelating chromatography (IMAC) using Ni-NTA (Thermo Fisher Scientific) or Ni-Penta agarose (Marvelgent Biosciences). After elution from the metal chromatography column, the protein was diafiltrated against 25 mM sodium phosphate buffer at a range of 5.0 to 7.0 pH values. The presence and purity of proteins were monitored by SDS-PAGE electrophoresis using 20% (w/v) polyacrylamide gel. The 18-bp Core-DNA and 32-bp Li and Lo-DNA sequences synthesized by IDT (Integrated DNA Technologies, Inc.) have been used for DNA binding experiments, as shown in Figure 2.

Left Inner (Li) DR of ITR_{left} :	TCCAGTGGG**TCAGAAGTTTACATACAC**TAAGT
Left Outer (Lo) DR of ITR_{left} :	CAGTTGAAG**TCGGAAGTTTACATACAC**TTAAG
Core DR :	**TCGGAAGTTTACATACAC**

Figure 2. DNA sequences for Li, Lo, and Core were used for both MST and NMR Binding experiments. Bold partial sequences in Li and Lo indicate the Core-DNA sequence, the common sequence between the four binding sites on the transposon DNA.

3.2 Microscale Thermophoresis (MST)

Microscale Thermophoresis is an optical fluorescent method that records the strength of the interaction (binding affinity) between two biomolecules by quantifying the variations in the fluorescence signal as a result of an infrared (IR)-laser-induced temperature change. The

fluorescent signal variation corresponds to a ligand's binding to a fluorescently labeled or intrinsically fluorescent target [30]. Two major effects contribute to the variation in the fluorescence signal: Temperature-Related Intensity Change (TRIC) and thermophoresis. TRIC is an effect in which the fluorescence intensity of fluorophores is temperature-dependent. The fluorescence intensity of most fluorophores diminishes with increasing temperature [31]. Most notably, the extent of the temperature dependence is strongly related to the chemical environment of the fluorophore, which can be altered by binding events [32]. On the other hand, molecules display directed movement through temperature gradients, typically from a hot to cold region, an effect called thermophoresis. When molecules are subjected to a temperature gradient, those molecules escape from heat. Mass diffusion drives the flow of molecules along the temperature gradient. A molecule's thermophoretic properties are determined by charge, size, and conformation. Upon binding to an interacting partner, one or more of these parameters may be altered, resulting in a change in its thermophoretic movement. Thermophoresis of fluorescent molecules in a temperature gradient induces a change in concentration, resulting in a quantifiable fluorescence variation [32-33]. Thus, TRIC and thermophoresis are influenced by binding events and contribute to the measured MST signal.

The change in fluorescence of a given sample after IR laser activation is the product of the concentration of molecules (c) and the fluorescence of an individual molecule (F), given by the following equation:

$$\frac{\partial}{\partial T}(cF) = c\frac{\partial F}{\partial T} + F\frac{\partial c}{\partial T} \quad (1)$$

The first term $c\frac{\partial F}{\partial T}$ describes the change in fluorescence due to TRIC, whereas the second term represents the thermophoretic change in concentration [32]. The second term can also be characterized by $F\frac{\partial c}{\partial T} = -S_T c$ [34]. The spatial temperature difference ΔT leads to a depletion of

solvated biomolecules in a higher to a lower temperature region. This movement is quantified by the Soret coefficient S_T:

$$\frac{c_{hot}}{c_{cold}} = \exp\left(-S_T \Delta T\right) \quad (2)$$

where c_{hot} is the molecule concentration in the hot region and c_{cold} is the concentration in the cold region. The Soret coefficient S_T is defined by three parameters, charge, size and conformation which can be expressed by:

$$S_T = \frac{A}{kT}\left(-S_{hyd} + \frac{\beta \sigma_{eff}^2}{4\varepsilon\varepsilon_0 T} \times \lambda_{DH}\right) \quad (3)$$

where A is the molecule's surface area, S_{hyd} is the hydration entropy of the molecule-solution interface (conformation), and σ_{eff} is the effective charge. The depletion (Eqn. 2) depends on the thermophoresis of the ligand-target complex, which is influenced by the three parameters mentioned previously [35]. As shown above, the fluorescence change measured by MST is a bipartite signal that consists of TRIC and thermophoresis. The bipartite signal is a strong advantage of MST measurements, as it allows for the precise analysis of binding events by monitoring the change of fluorescent molecules along a microscopic temperature gradient in μl-volumes.

To determine the binding affinities of protein (transposase) to DNA (transposon), we utilize MST, as shown in the experimental setup in Figure 3. As shown in Figure 3B, thermophoresis is induced and detected in glass capillaries containing sample with a volume of ~4 μl each. An IR laser is guided through an objective into the capillary to produce a microscopic gradient spanning 2-6 ^{o}C in a volume with a diameter of ~50 μm. The applied temperature gradient excites the fluorophores in solution, and their emitted fluorescence is collected through the same objective. In Figure 3C, an MST signal showing the thermophoretic

movement of biomolecules can be observed. Before the IR laser is turned on, molecules are homogeneously distributed, and a stable initial fluorescence is recorded. As soon as the IR laser is turned on, a fast "T-jump" is recorded, corresponding to a rapid fluorescence change due to the quick temperature change. This is followed by thermophoresis, which can be seen as a decrease in fluorescence. After fluorescence reaches a steady state at 30 s, the IR laser is turned on, and an inverse T-jump is observed, followed by a back diffusion of molecules [35]. Back diffusion provides information on the reliability of an experiment by indicating aggregation or oligomerization of molecules if a change in molecule size is found [36]. To detect a fluorescence signal, transposon DNA sequences Li and Lo were fluorescently labeled with Cy5 and titrated in a sequential dilution with unlabeled PAI-H19Y SB Transposase protein.

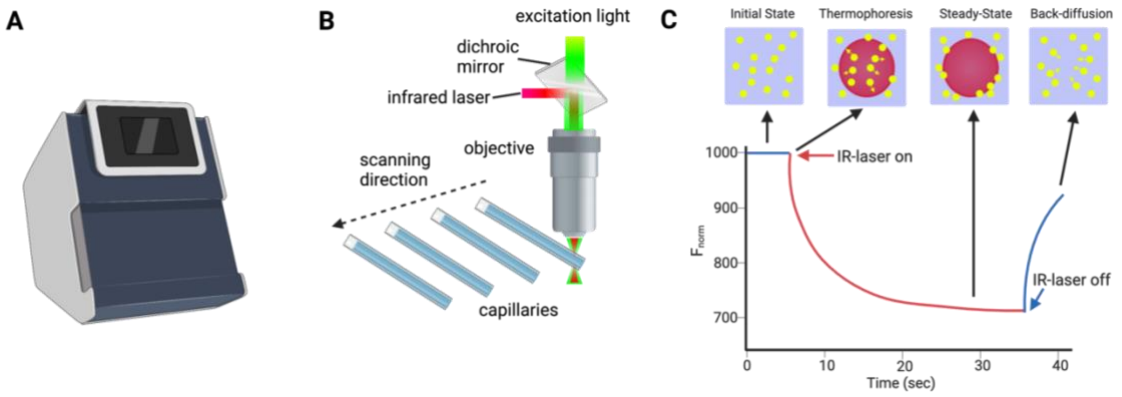

Figure 3. (A) Monolith NT.115 MST device (B) Technical optical setup of the MST device. (C) Example MST time trace captures the movement of fluorescent biomolecules in a temperature gradient. Figure adapted from [37].

To derive the binding affinity from MST signal, a sample solution mounted in 16 capillaries contains a constant concentration of labeled target for all capillaries and decreasing concentration of ligand by a factor of two for each subsequent capillary. This serial dilution is shown through the different traces schematically represented in Figure 4A, where the fluorescent molecule (black trace, "unbound") changes upon the increasing concentration of non-fluorescent

ligand (red trace, "bound), resulting in different traces. From the graph of normalized fluorescence as a function of the ligand concentration, data points are fitted to obtain binding constants K_D. The change in thermophoresis is expressed as the change in the normalized fluorescence ΔF_{norm} defined as $F_{norm}=F_{hot}/F_{cold}$ in which the F values correspond to the values between the defined areas marked by the red and blue cursors in Figure 4A [26,35-37].

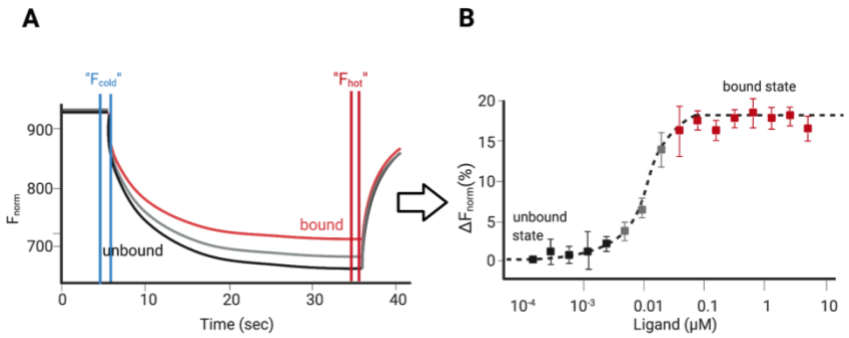

Figure 4. (A) Example results of a typical MST experiment include the fluorescence time traces for 16 capillaries (B) Normalized fluorescence of MST traces as a function of the concentration of the ligand. Figure adapted from [37].

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

3.3 Nuclear Magnetic (NMR) Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a powerful biophysical technique that facilitates the determination of the three-dimensional structure and interactions of proteins and protein complexes in solution at an atomic resolution. NMR is based on the fundamental concept that all nuclei have an intrinsic quantum property known as spin. When sample containing ^{1}H and/or isotopically labeled nuclei ^{13}C or ^{15}N is placed in the external magnetic field, nuclei spins orient parallel (lower energy state) or anti-parallel (higher energy state) to the field. By irradiating the nucleus with radiofrequency (RF) waves with frequencies corresponding

to the difference between lower and higher energy states, a transition from a lower to a higher energy state can be induced. The absorption of RF energy during this transition forms the basis of the NMR method. Because all nuclei are in slightly different chemical environments determined by the protein structure, the transitions between lower and higher energy states will occur at slightly different frequencies [38]. The frequency change can provide details of a molecule's individual functional groups and structure. Moreover, chemical shifts produced by NMR are sensitive probes for structural changes and are used to assess the folded state and monitor the effects of ligand binding on the protein structure [39]. Thus, this method can provide a high-resolution image of the binding interface between the PAI-H19Y mutant of SB transposase to transposon DNA.

For NMR studies, the protein was expressed in cells grown in minimal medium (M9) where ^{15}N-isotopically enriched ammonium chloride was used as a sole source of nitrogen and subsequently purified to obtain pure protein as previously described. [^{1}H-^{15}N]-Heteronuclear Single Quantum Coherence (HSQC) is used, which is a two-dimensional (2D) spectrum with one axis for proton (^{1}H) and the other for a heteronucleus, in this case, ^{15}N. HSQC correlates the amide proton attached to nitrogen in the peptide bond, as all residues (except proline) have an amide proton (Fig. 5). Thus, each residue produces a cross-peak in the HSQC spectra [38].

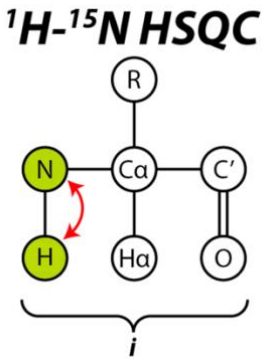

Figure 5. [^{1}H, ^{15}N]-HSQC mechanism scheme for residues in protein

Proton chemical shifts were calibrated with respect to the water signal. Chemical shifts were assigned using existing HSQC peak lists obtained from the Biological Magnetic Resonance Data Bank (BMRB entry 30680) and assigned reference spectra for the PAI-H19Y subdomain to transposon Core-DNA [17]. All NMR experiments were performed at the Molecular Education, Technology, and Research Innovation Center (METRIC) at North Carolina State University. NMR spectra was processed using the NMRPipe software and analyzed with programs SPARKY and POKY [40-42].

3.3.1 NMR Chemical Exchange Regimes

Chemical exchange is a process in NMR when a molecule of interest is exposed to a changing environment that results from it populating two or more states, i.e., bound and unbound. Across a wide range of timescales, NMR spectra are sensitive to chemical exchange. Chemical exchange between two states can be characterized by the exchange rate k_{ex}, which is the sum of forward and backward reaction rates. For a protein P binding reversibly to a ligand L at a single site, a complex PL is formed [43]:

$$P + L \leftrightharpoons PL \tag{3}$$

This simple two-state binding reaction is defined by the rate constants for forward and back reactions of k_{on} and k_{off}. The following equation gives the exchange rate:

$$k_{ex} = k_{on}[L] + k_{off} \tag{4}$$

The forward and backward rates are given by $[P][L]k_{on}$ and $[PL]k_{off,}$ respectively. At equilibrium, the forward and back rates are equal, and the dissociation constant $K_{D,}$ which is the constant which is commonly used to describe the affinity between a protein P and ligand L, is defined as

$$K_D = \frac{k_{off}}{k_{on}} \tag{5}$$

In terms of concentrations of P and L, the dissociation constant can also be defined by

$$K_D = \frac{[P][L]}{[PL]} \tag{6}$$

where [P], [L], and [PL] represent molar concentrations of protein, ligand, and complex, respectively. Moreover, the K_D value has molar units (M) and corresponds to the free ligand concentration at which half the protein population is bound to the ligand [27]. Note that the smaller the K_D value, the more tightly bound the ligand is, or the higher the affinity between protein and ligand.

A protein in a simple two-state reaction (Eqn. 3) will give rise to two species with different resonance frequencies ω_P and ω_{PL}, and a resonance frequency difference of $\Delta\omega = \omega_P - \omega_{PL}$ (Fig. 6). The appearance of the different species of the protein in the NMR spectra varies and is strongly dependent on the concentrations of species (bound PL and unbound P), the K_D value and the exchange rate (Eqn. 4). Chemical exchange regimes, is based on the magnitude of k_{ex} relative to the difference in resonance frequency $\Delta\omega$ (or chemical shift) of a given nucleus in P and PL. The regimes are typically divided into three categories—slow, intermediate, and fast exchange [43].

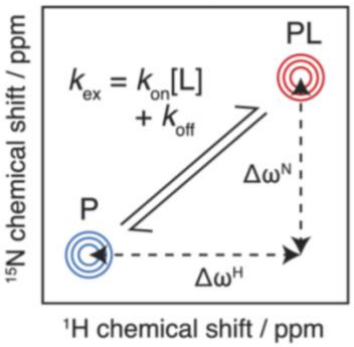

Figure 6. Schematic showing exchange rate and frequency differences for a protein-ligand interaction for a 2D HSQC experiment. Figure adapted from [44].

If the exchange is slow on the NMR time scale ($k_{ex} << \Delta\omega$), that is, when the exchange rate is much slower than the observed difference in resonance frequency, each state and the corresponding resonance frequencies can be individually observed (Fig. 7A-B, top graph).

At the intermediate NMR time scale ($k_{ex} \approx \Delta\omega$), if the lifetimes of the states are very short, the difference in frequency cannot be measured, resulting in a collapse of measured signals, known as coalescence. In other words, there is significant interconversion between bound and unbound states during the NMR detection period, causing signals to broaden and shift simultaneously (Fig. 7A-B, middle two graphs). Moreover, this indicates coupled processes occurring, namely binding interactions and conformational changes, which make K_D values challenging to obtain [27,43].

Finally, at the fast NMR time scale ($k_{ex} >> \Delta\omega$), a single peak will be observed with a population-weighted average position of their respective resonance frequencies given by $\omega_{Obs} = p_P\omega_P + p_{PL}\omega_{PL,}$ where p_P and p_{PL} are the populations of the free and bound protein, respectively (Fig. 7A-B, bottom graph). From this case, obtaining the K_D is straightforward by plotting the peak position as a function of the added ligand concentration.

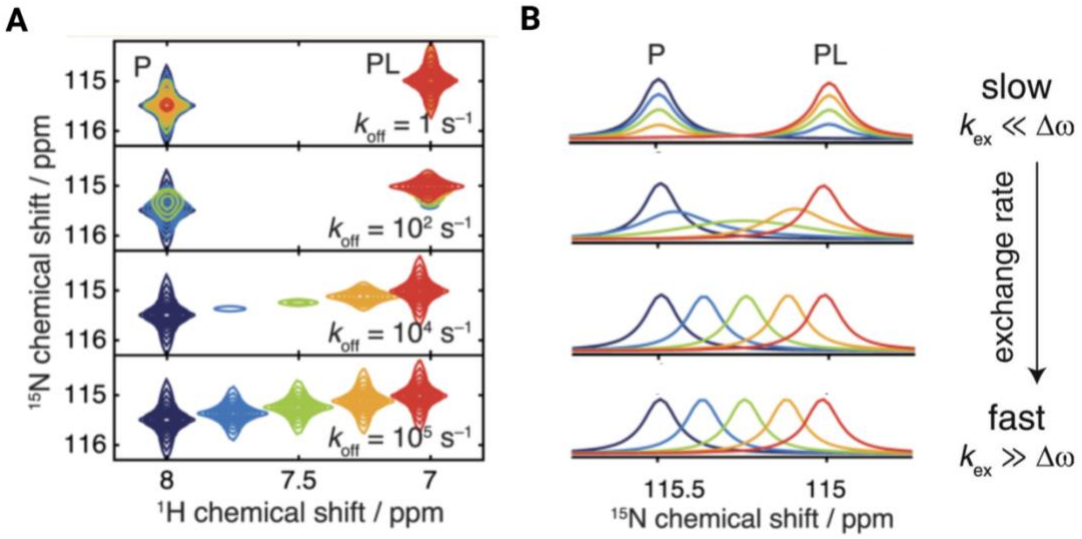

Figure 7. (A) Stimulated [^{1}H, ^{15}N]-HSQC spectra for a protein-ligand interaction illustrating two-dimensional lineshapes that may arise under various exchange regimes as indicated. (B) ^{15}N projections of HSQC spectra are shown in Figure 7A. Figure adapted from [44].

For a simple two-state binding occurring in the fast exchange regime, the K_D can be fitted by the function:

$$\Delta\delta_{obs} = \Delta\delta_{max}\frac{([P]_0 + [L]_0 + K_D) - \sqrt{([P]_0 + [L]_0 + K_D)^2 - 4[P]_0[L]_0}}{2[P]_0}$$

where $\Delta\delta_{obs}$ and $\Delta\delta_{max}$ are the observed/combined and maximum chemical shift changes, respectively. P_0 and L_0 are the total protein and ligand concentrations. Additionally, the K_D value for MST experiments can also be derived with a variation of this equation, not considering the $\Delta\delta_{max}$ [26].

Additionally, it is important to note that only k_{on}, k_{off}, and [L] determine the exchange rate. On the other hand, the dissociation constant K_D will not determine whether a binding reaction will appear in slow, intermediate, or fast exchange. Nonetheless, there is still a tendency for strong binding reactions to be in low exchange and weak binding reactions to be in fast exchange. This is because a small K_D will result in low free ligand concentrations until the protein becomes saturated [43].

Chapter 4. Results

4.1 The association of PAI-H19Y by MST

For protein-DNA binding experiments, the DNA used was left inner and outer direct repeats (DRs) (Fig. 2) on the transposon DNA where the transposase binds. Two DRs were labeled by Cy5, a fluorescent dye that is needed to measure the movement of molecules in temperature gradients, the physical phenomenon on which MST is based. The Monolith 115.NT instrument used for MST experiments is equipped with two LED lasers, blue and red. The red laser has excitation wavelengths of 600-640 nm and emission wavelengths of 660-725 nm. Cy5 is a traditional far-red fluorescent label with excitation ideally suited for 633 or 647 nm laser lines making this fluorophore suitable for labeling and use in MST experiments.

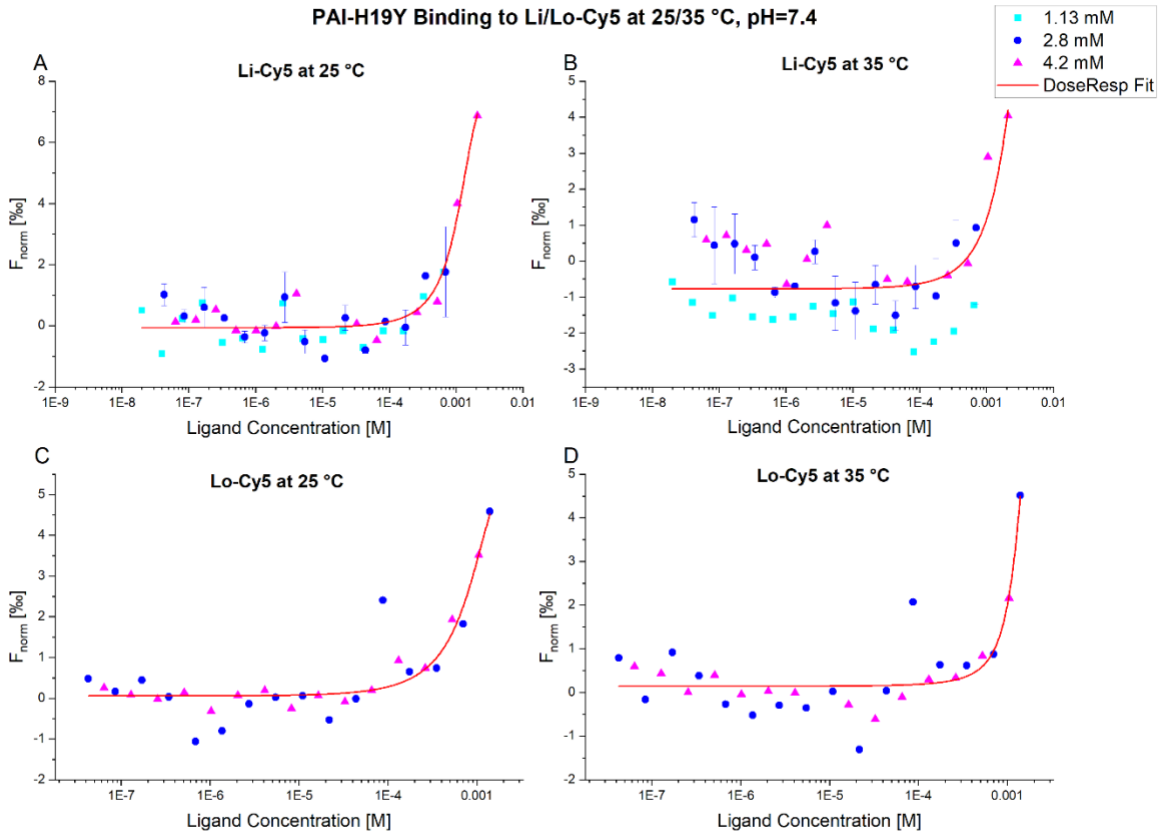

Figure 8. The binding of the PAI-H19Y mutant of the SB transposase to fluorescent-labeled DNA was determined using the Monolith NT. 115 MST instrument.

The four binding affinities were measured using the same buffer, except for varying pH and temperatures. Buffer used was 25 mM sodium phosphate with 150 mM NaCl, 0.2 mg/mL BSA, and 0.05% Tween. The binding affinities of PAI-H19Y protein to the left inner (Li) and left outer (Lo) DRs at temperatures of 25 and 35 °C were quantified as shown in Figure 8A-D. Binding measurements were taken for different maximum H19Y concentrations at 1.13, 2.8, and 4.2 mM, all at a pH of 7.4. For Figures 8A-B, MST experiments were repeated twice for 2.8 mM; thus, average measurements are shown. A serial dilution was prepared to measure binding events in which the H19Y non-fluorescent protein is titrated against a fixed concentration at 50 nM of the fluorescent Cy5-labeled DNA (target). Concentration is chosen based on a pretest experiment to optimize fluorescence signal, and this is a common approach to simplifying K_D determination in biochemical and biophysical assays. Every sequential dilution in each successive capillary is two times, consisting of diluted H19Y protein to diluted buffer and Cy5-labeled DNA. The serial dilution was mounted in 16 capillaries, and with each subsequent capillary, ligand concentration increases, and the proportion of ligand bound to the target increases until saturation. The binding constant is the point at which half the ligand is bound to the target.

Data obtained from MST binding experiments for PAI-H19Y to Li and Lo-DNA was compared to previous MST experiments for PAI-H19Y to Core-DNA. The previous experiment's buffer conditions are nearly the same, with the buffer used being 25 mM sodium phosphate with 0.1% Tween and 150 mM NaCl at pH 7 and temperatures of 25 °C and 35 °C. Figure 9 shows the binding curves for PAI-H19Y to Core-DNA.

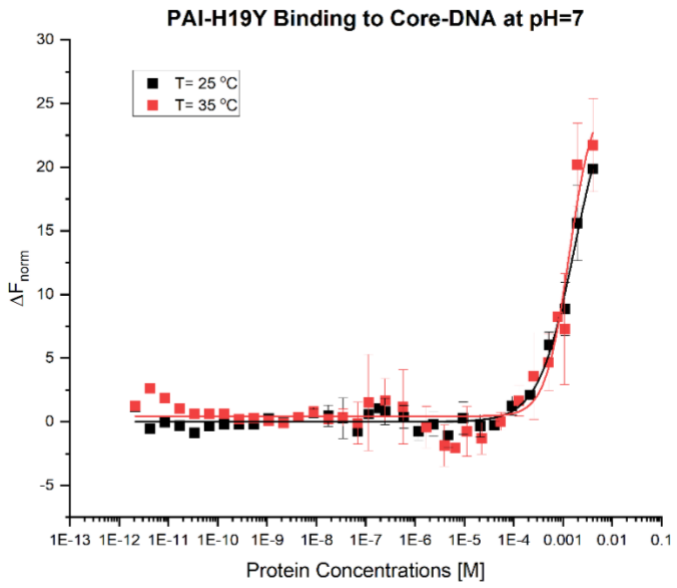

Figure 9. Previous MST binding experiments of PAI-H1Y to fluorescently labeled Core-DNA.

Data analysis was completed using Origin and MST software to obtain the binding constants. For Figure 8, outliers from the capillary scan and dose-response data were omitted in MST software as these values indicated nonspecific ligand adsorption to capillaries. Next, data from each H19Y concentration measurement set was fitted until convergence using the Hill1 function to obtain the background fluorescence value. This value was then subtracted from fluorescence values. Finally, data sets from different H19Y concentrations were combined concerning the same target (Li/Lo), pH, and temperature. Using Origin, binding curves and constants were obtained using Dose-Response fit. For Figure 9, data points were fitted using the Hill1 function. Binding values, K_D, and their corresponding standard error values for Figure 8A-D and Figure 9 are shown in Table 1.

	pH 7.4	
DR DNA	**25 °C**	**35 °C**
Li-Cy5	0.93 ± 0.15 mM	1.88 ± 2.67 mM
Lo-Cy5	0.72 ± 0.28 mM	1.52 ± 0.93 mM
Core-Cy5	2.51 ± 0.45 mM	3.36 ± 1.19 mM

Table 1: Using a data analysis software program, Origin, binding curves, and constants were obtained using Dose-Response and Hill1 fits. K_D values and their corresponding standard error values for Figure 8a-d (Li/Lo-Cy5) and Figure 9 (Core) are shown.

These values indicate low-affinity, millimolar binding of the PAI subdomain to the transposon DNA DRs. Moreover, as expected, there is also weaker binding at a temperature of 35 °C than at 25 °C. Large experimental errors are due to the inability to record the full binding curve due to protein aggregation in the capillary. Although the binding constants suggest a tighter binding occurring between transposase and left outer DR than left inner DR, large experimental errors do not allow for drawing a statistically significant conclusion [18]. Moreover, protein to Core-DNA displays weaker binding than to Li and Lo-DNA.

Preliminary binding experiments done at a pH of 5.2 indicate that there are dye-dependent effects. It could be that at low pH values, the histidine tag on the protein becomes charged and interacts with the fluorescent label (Cy5), which is seen as ligand-induced fluorescence changes through the capillary scans [25]. Recommendations by the MST software program to improve the assay are to change the labeling strategy, increase MST power or reevaluate assay conditions. MST power was increased, to which there was no improvement. Thus, changing the labeling strategy to have the protein fluorescently labeled instead of DNA is the next step in future MST experiments. A different dye can also be used to label DNA, e.g. Atto488 or other hydrophilic dyes. Labeling protein could also allow measurements at a higher ratio of DNA to protein to cover the full range binding curve and decrease standard error values because such measurements were not possible due to aggregates and gel formation at higher protein concentrations (e.g., protein to DNA ratio).

4.2 The association of PAI-H19Y by NMR

We used NMR spectroscopy to assess the binding sites and affinities by analyzing the chemical shifts produced during ligand binding. For NMR experiments, a sample with the constant concentration of 0.2 mM ^{15}N-labeled protein was prepared, to which DNA was added at increasing concentrations. DNA sequences used were the left inner (Li), left outer (Lo), and Core-DNA (Fig. 2). The 18-base pair (bp) Core-DNA sequence is the minimum required DNA sequence for SB transposase binding and the common sequence between the four binding sites on the transposon DNA. The PAI-H19Y protein was prepared and kept in buffer conditions of 25 mM sodium phosphate buffer with 150 mM NaCl at pH 7.4. The protein concentration was chosen at 0.2 mM to achieve a good signal-to-noise ratio. NMR experiments were done at temperatures 7, 25, and 35 °C. Temperatures were chosen to maintain consistency with previous NMR experiments done for PAI-H19Y binding experiments. The highest quality NMR spectrum of pure PAI subdomain with all peaks present was observed at 7 °C, whereas in the presence of DNA, the highest spectrum quality was observed at 35 °C. Therefore, these two temperatures were used to analyze chemical shift changes.

Observed signal broadening in the NMR spectra, collected using a pure PAI subdomain at different temperatures, indicates intermediate conformational exchange on the NMR time scale between folded and unfolded protein states. Moreover, current data and previous NMR studies done on the PAI subdomain of SB show that DNA binding also occurs in the slow to intermediate regime on the NMR time scale, leading to the broadening of the PAI resonances [15-16]. The fact that ligand binding to a protein can produce chemical shifts either by direct interactions at the binding sites or by inducing a conformational change in the protein, makes the differentiation between shift changes due to these coupled processes impossible [27].

Figure 10 shows NMR spectra of protein with and without Lo-DNA at different temperatures of 7 and 35 ^{o}C. The highest quality NMR spectrum of pure protein with all peaks present was observed at 7 °C (Fig. 10A), whereas the highest quality spectra in the presence of DNA was observed at 35 °C (Fig. 10D). The spectral differences between spectra at 7 and 35 °C with no DNA added indicates conformational change occurring, between folded and unfolded protein states. Conformational changes are observed by looking at the range of ^{1}H shifts between spectra. Since most of the peaks in HSQC spectra arise from backbone amides, ^{1}H shifts can be used to evaluate the environment in which the residues are found, serving as indicators of whether the backbone state is folded or unfolded. Significant dispersion of peaks in the ^{1}H scale (Fig. 10A) is indicative of most residues residing in different electronic environments. In this case, the protein is said to be in its folded state. In contrast, an unfolded protein would have most of its backbone amide groups exposed to an identical environment in which peaks lie in a narrow range of ^{1}H shift values. We observe a significant signal broadening at 35 °C (Fig. 10B) for some residues, which is indicative of the intermediate exchange on the NMR time scale between folded and unfolded states. The fully unfolded state is not observed at 35 °C, but one can expect to observe it at higher temperatures, at which the protein is predominantly unfolded, if the unfolding is not followed by protein aggregation.

Coupled processes can be observed by comparing spectra at 35 °C with (Fig. 10D) and without DNA (Fig. 10B). Shift changes between selected residues for analysis indicate binding occurring. Moreover, the increase of signal intensity in Figure 10D suggests the presence of folded protein upon the addition of DNA. Furthermore, the spectral differences between samples with an added DNA concentration of 1 mM at a temperature of 7 ^{o}C (Fig. 10C) compared to 35 ^{o}C show that many signals disappear from the spectra in Figure 10C, due to signal broadening.

This again is indicative of the intermediate chemical exchange on the NMR time scale, most likely due to the formation of higher-order protein-DNA complexes at lower temperatures.

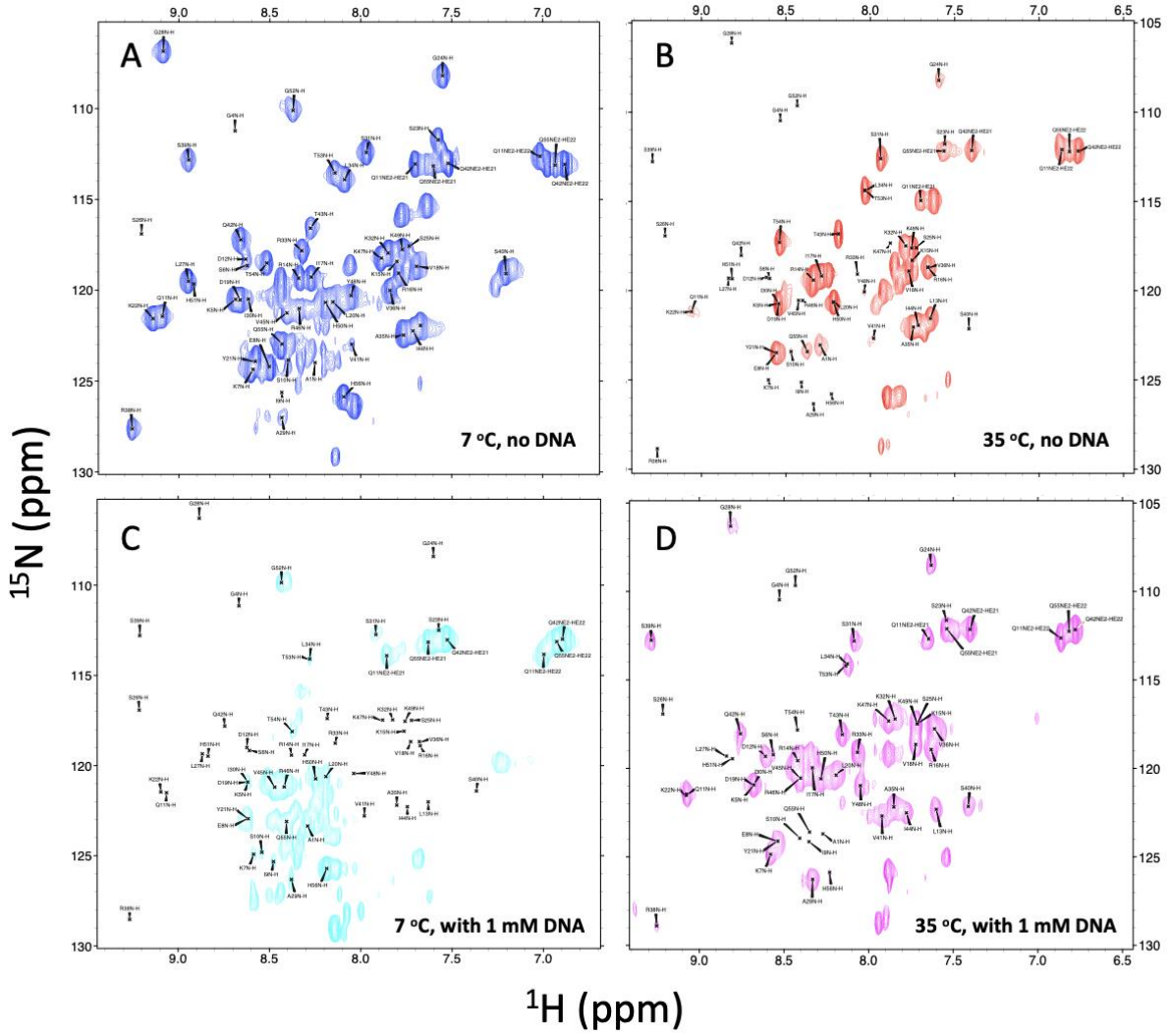

Figure 10: Spectra for PAI-H19Y with and without addition of Lo-DNA collected at temperatures of 7 and 35 °C. (A) Spectra for PAI-H19Y with no DNA were collected at 7 °, (B) 35 °C, with 1 mM Lo-DNA at 7 °C (C) and at 35 °C (D).

To obtain the binding sites and affinities, [^{1}H-^{15}N]-HSQC spectra were collected at an increasing concentration range of 0.04-3 mM of Li, Lo, and Core-DNA sequences. For each respective temperature and DNA sequence, the spectra at varying concentrations were overlayed to map the chemical shifts and to assign residues. Figure 11 shows the overlayed [^{1}H,^{15}N]-HSQC

spectra collected at 35 °C for PAI-H19Y SB transposase to Lo DNA. NMR signal assignments for PAI-H19Y amino acids are indicated; the letters are the abbreviation of the amino acids, followed by the amino acid position in the protein sequence. The Figure shows that severely broadened or overlapped signals are present for many residues. However, several residues, such as Q11, L13, K22, G24, S31, L34, A35, I44, and T53, demonstrate well-separated signals present at both temperatures. These residues were selected for the detailed analysis.

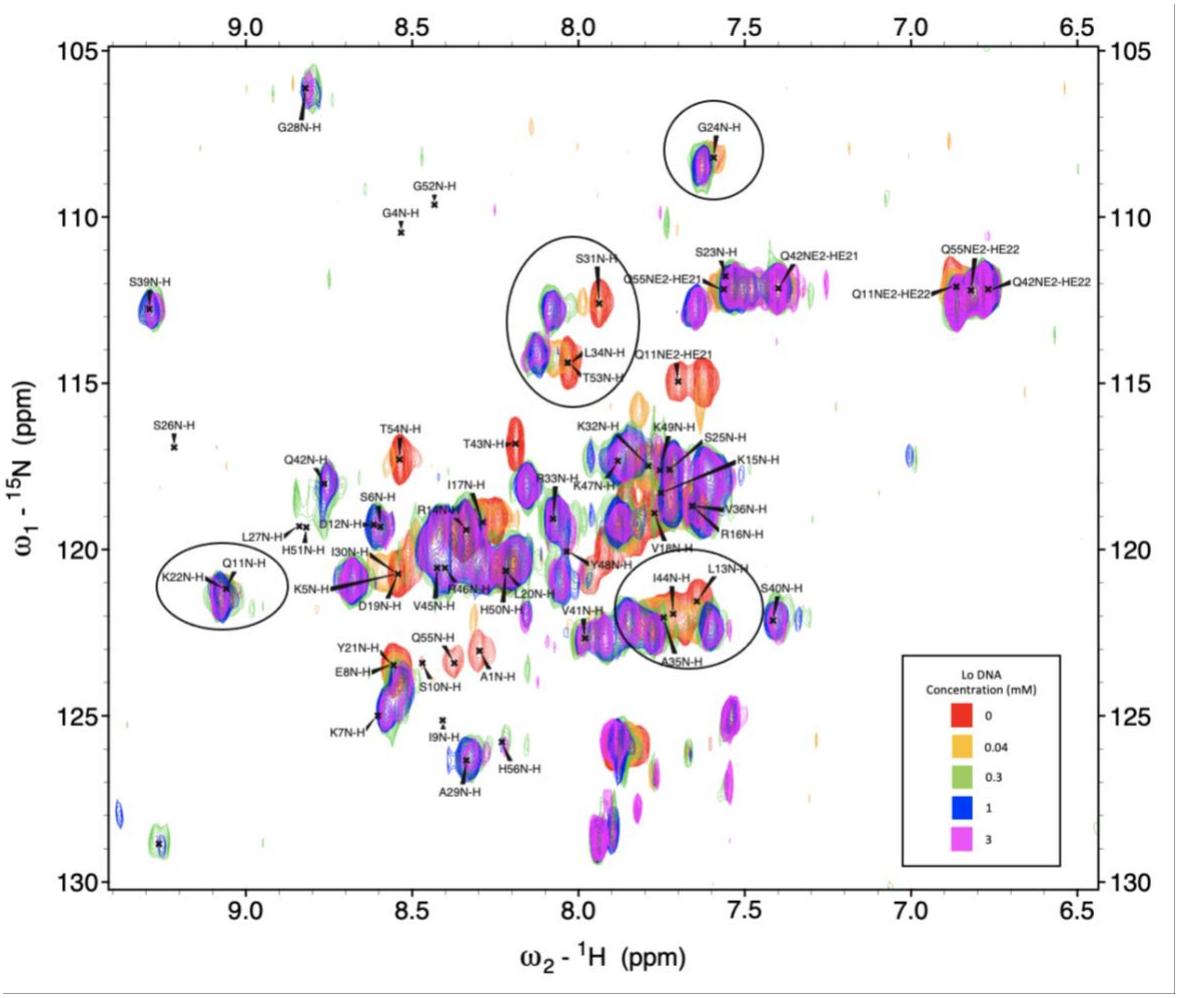

Figure 11. [^{1}H,^{15}N]-HSQC spectra for ^{15}N-labeled PAI-H19Y at a range of 0.04-0.3 mM of Lo DNA concentration. Residues selected for analysis, such as Q11, L13, K22, G24, S31, L34, A35, I44, and T53 are circled in black. Buffer condition: 25 mM sodium phosphate buffer with 150 mM NaCl at pH 7.4 at 35 °C.

Although chemical shifts are not indicative of actual binding sites or affinities, they can still be used to obtain an apparent binding constant value, K_D. For each of the nine residues selected for

the analysis and its corresponding spectra for Li, Lo and Core-DNA titrations, the chemical shift differences between the bound and unbound states were obtained for each titration point. The binding curves for each residue and its corresponding DNA sequence were plotted as normalized ΔH, ΔN, and Δppm shifts as a function of DNA concentration, respectively. The Δppm shifts in each residue were calculated using the formula $\Delta\delta = \sqrt{(\Delta H)^2 + 0.15 * (\Delta N)^2}$ where ΔH is the change of the proton chemical shifts between samples without DNA and varying DNA concentrations, and ΔN is the respective change of the nitrogen shift. To obtain the apparent binding affinities, the experimental Δppm values were fitted on Origin using the Logistic function, as seen for residues S31 and T53 shown in Figure 12A-B.

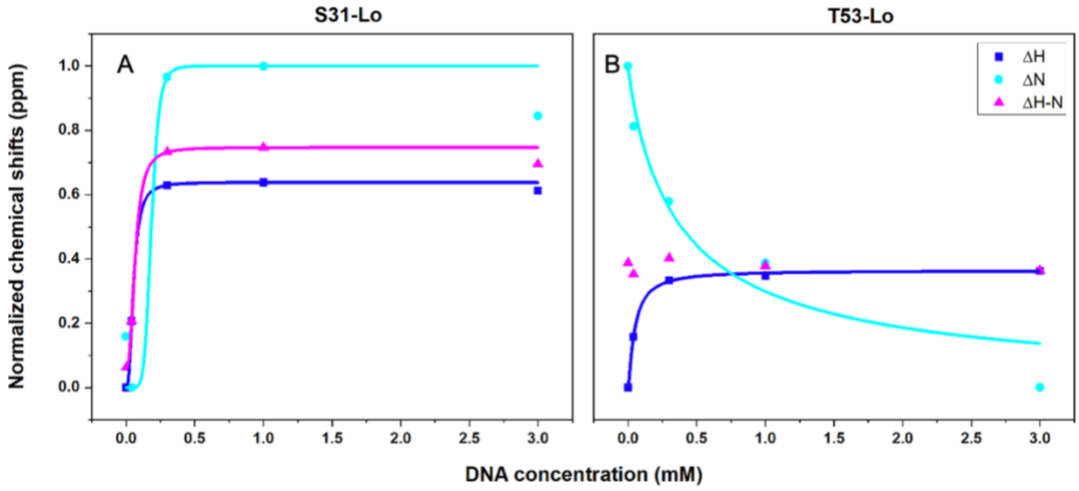

Figure 12. Binding curves for residues S31 (Figure 12A) and T53 (Figure 12B) were obtained from [^{1}H,^{15}N]-HSQC as shown in Figure 11. Data points are plotted as normalized ΔH, ΔN, and Δppm shifts as a function of DNA concentration, respectively. Curves are fitted using the Logistic Function on Origin software to obtain K_D values.

For the DNA sequences (Li and Lo) and its' corresponding [^{1}H-^{15}N]-HSQC spectra, there was overlap present for one residue, T53, as shown in Figure 13. From the spectra, T53 was indistinguishable from L34, and both peaks were assigned right next to each other. Because of this, the resulting Δppm shifts in Figure 12B could not be fitted to obtain a binding curve. However, there was a noticeable shift for the eight residues, including S31, and curves could be

easily fitted with the Logistic function. The K_D values for the selected residues for each Li, Lo, and Core spectra, excluding T53 for Li and Lo, were averaged to obtain a final apparent K_D value.

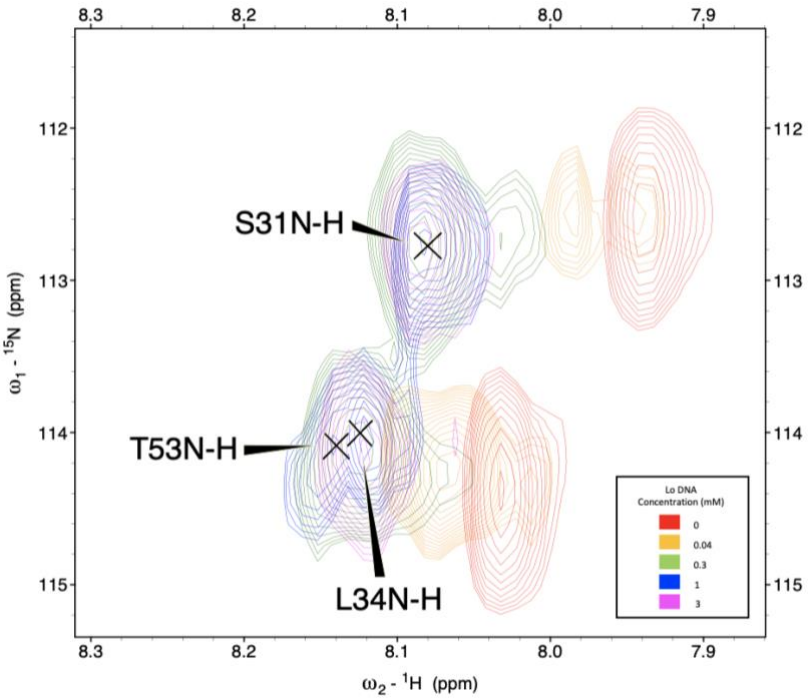

Figure 13. Zoomed in overlayed [^{1}H,^{15}N]-HSQC spectra from Figure 11 displaying residues chemical shifts for residues S31, L34, and T53.

Figure 14 shows the normalized Δppm shifts for all analyzed residues for each DNA sequence plotted as a function of DNA concentration. Figures 14A-C show the chemical shift changes for nearly all binding curves plateau at 1.0 mM DNA concentration, with the most

substantial changes occurring between DNA concentrations of 0-0.3 mM.

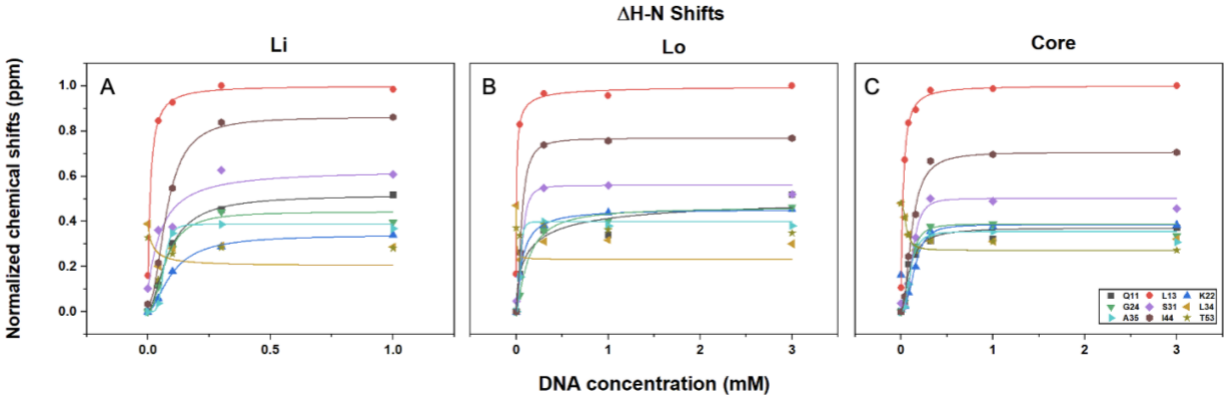

Figure 14. Binding curves for each DNA sequence. Data points are plotted as normalized Δppm shifts as a function of DNA concentration.

Table 2 shows the final K_D values for each DNA sequence binding to the PAI-H19Y mutant of SB transposase. Noting that averaging eight residues for Li and Lo compared to nine residues for Core-DNA could skew the final Core K_D values, we verified that the K_D value stayed relatively with or without excluding T53 from the analysis. Both Li and Lo K_D values are essentially the same at 60.77 ± 8.48 and 62.35 ± 18.31 μM, respectively, and the Core K_D value is weaker at 93.68 ± 11.46 μM. Note that higher dissociation constant values indicate weaker binding.

Average apparent K_D values for each DR DNA sequence are plotted as a bar graph in Figure 15. A one-way ANOVA was performed on the data set to determine if there is any difference between the several K_D means. The test returned a p-value greater than 0.05, at 0.18. However, from the Figure, it is evident that there is a noticeable difference between average K_D values for Core and the other two DNA sequences, Li and Lo. Additionally, it is observed that average K_D values and corresponding errors overlap between Li and Lo, indicating no notable difference between the two DNA constructs.

	DR DNA (μM)		
Residues	**Li**	**Lo**	**Core**
Q11	83.83 ± 2.67	124.73 ± 60.44	91.77 ± 14.03
L13	12.29 ± 3.64	6.01 ± 2.92	25.33 ± 19.8
K22	98.86 ± 4.27	65.33 ± 2.24	156.68 ± 32.72
G24	73.26 ± 7.61	130.93 ± 17.87	106.22 ± 8.62
S31	55.8 ± 17.78	66.9 ± 15.04	129.07 ± 8.12
L34	19.22 ± 0.027	3.05 ± 13.13	57.7 ± 13.81
A35	65.26 ± 28.6	45.24 ± 33.82	102.68 ± 7.46
I44	77.67 ± 2.05	56.64 ± 1.04	116.01 ± 5.96
T53	N/A	N/A	57.71 ± 10.45
Average K_D	**60.77 ± 8.48**	**62.35 ± 18.31**	**93.68 ± 11.46**

Table 2: Binding constants and their corresponding standard error values for each residue and its corresponding DR DNA sequence.

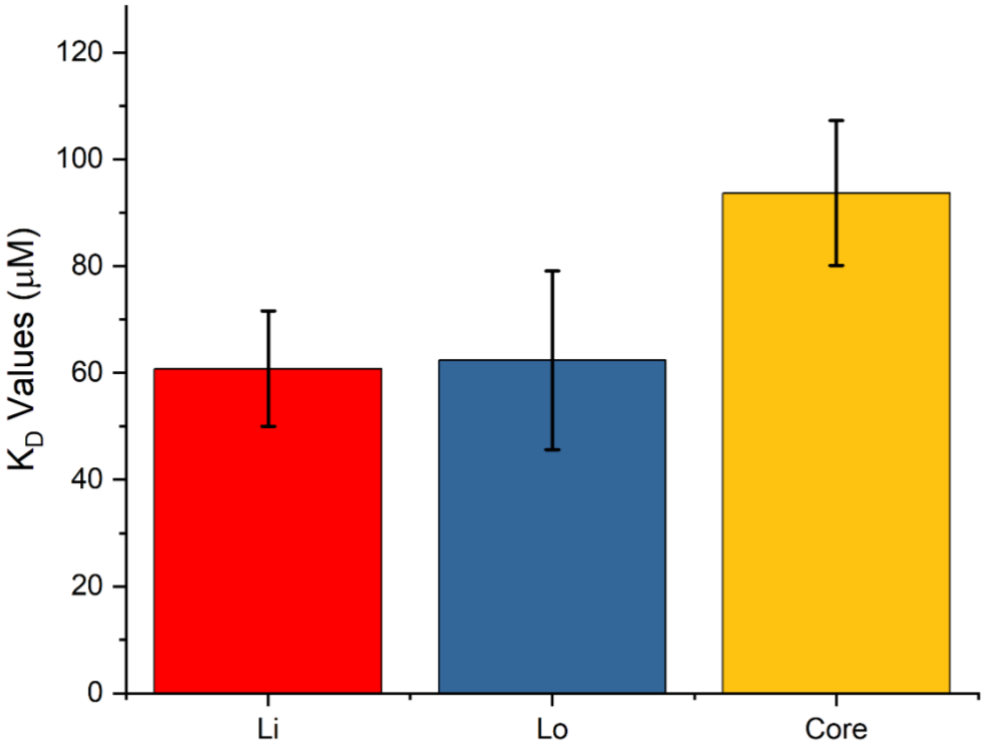

Figure 15: Bar graph for the average K_D values for PAI-H19Y SB transposase binding to Li, Lo, and Core-DNA sequences. Both Li and Lo had comparable values and standard errors with 60.77 ± 8.48 and 62.35 ± 18.31 μM, respectively. Transposase to Core-DNA has weaker binding at 93.68 ± 11.46 μM. Nonetheless, all K_D values are in the micromolar range.

Chapter 5. Discussion

Binding experiments conducted with MST and NMR spectroscopy to determine the association of PAI-H19Y mutant of SB transposase to transposon DNA resulted in different K_D values. Binding affinities obtained through MST showed close to millimolar values. Moreover, these values were in agreement with previous studies that utilized MST to determine the affinity of PAI-H19Y to Core-DNA [45]. In contrast, values obtained by NMR were smaller. We attribute such a difference to the coupling of DNA binding and folding observed at the intermediate exchange regime on the NMR time scale. It is possible that such a coupling displays itself differently during the MST measurement due to the different physical basis of MST experiments and the fact that we label and observe signal from the DNA as compared to NMR where we directly observe the signal from protein. In the case of MST, DNA binds to already folded protein, and hence, the measured binding constant reflects protein-DNA interactions only. In this regard, both methodologies provide a different quantitative measurement of binding interactions occurring. Regardless of these contrasting results, values displayed in Table 1 suggest that tighter binding occurs between transposase and outer DRs or inner DRs than the Core-DNA. Binding values obtained via NMR also indicate weaker binding between transposase and Core-DNA. These data suggest that although the Core-DNA sequence is the minimal DNA sequence required for transposase binding, the additional DNA sequences of Li and Lo enhance its DNA-binding capability.

Chapter 6. Conclusion

Collectively, our results indicate that there is binding occurring in the sub-millimolar (MST data) range between the PAI-H19Y subdomain of the Sleeping Beauty transposase and transposon DNA. The binding to the Core-DNA is weaker than the binding to Li and Lo DNA constructs. NMR data show that the binding is coupled with the PAI subdomain folding. This is reflected in smaller apparent dissociation constants in the micromolar range.

To obtain the full binding curve in the MST experiment, changing the labeling strategy could be explored. Furthermore, ^{31}P NMR can be used as a complementary method for probing DNA interactions. This study relies on the fact that DNA contains phosphorus, but protein does not. Hence the binding affinity could be measured by observing the binding process from the "DNA point of view". As a next step, considering previously stated recommendations for techniques used, I propose further investigating protein-DNA interactions between the PAI-H19Y subdomain of the Sleeping Beauty transposase to the inner and outer DRs on the right ITR of transposon DNA.

References

1. Jackson, M., Marks, L., May, G. H., & Wilson, J. (2018). The genetic basis of disease. *Essays in Biochemistry*, *62*(5), 643–723.

2. Haendel, M., Vasilevsky, N., Unni, D., Bologa, C., Harris, N., Rehm, H., Hamosh, A., Baynam, G., Groza, T., McMurry, J., Dawkins, H., Rath, A., Thaxon, C., Bocci, G., Joachimiak, M. P., Köhler, S., Robinson, P. N., Mungall, C., & Oprea, T. I. (2019). How many rare diseases are there? *Nature Reviews Drug Discovery*, *19*(2), 77–78.

3. Nguengang Wakap, S., Lambert, D. M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., le Cam, Y., & Rath, A. (2019). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *European Journal of Human Genetics*, *28*(2), 165–173.

4. Kaufmann, P., Pariser, A. R., & Austin, C. (2018). From scientific discovery to treatments for rare diseases – the view from the National Center for Advancing Translational Sciences – Office of Rare Diseases Research. *Orphanet Journal of Rare Diseases*, *13*(1).

5. Li, H., Yang, Y., Hong, W., Huang, M., Wu, M., & Zhao, X. (2020). Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances, and prospects. *Signal Transduction and Targeted Therapy*, *5*(1).

6. Ghosh, S., Brown, A. M., Jenkins, C., & Campbell, K. (2020). Viral Vector Systems for Gene Therapy: A Comprehensive Literature Review of Progress and Biosafety Challenges. *Applied Biosafety*, *25*(1), 7–18.

7. Al-Dosari, M. S., & Gao, X. (2009). Nonviral Gene Delivery: Principle, Limitations, and Recent Progress. *The AAPS Journal*, *11*(4).

8. Ginn, S. L., Amaya, A. K., Alexander, I. E., Edelstein, M., & Abedi, M. R. (2018). Gene therapy clinical trials worldwide to 2017: An update. *The Journal of Gene Medicine*, *20*(5), e3015.

9. Liu, H., & Visner, G. A. (2007). Applications of *Sleeping Beauty* transposons for nonviral gene therapy. *IUBMB Life*, *59*(6), 374–379.

10. Hackett, P. B., Largaespada, D. A., & Cooper, L. J. (2010). A Transposon and Transposase System for Human Application. *Molecular Therapy*, *18*(4), 674–683.

11. Aronovich, E. L., McIvor, R. S., & Hackett, P. B. (2011). The Sleeping Beauty transposon system: a nonviral vector for gene therapy. *Human Molecular Genetics*, *20*(R1), R14–R20.

12. Ivics, Z., Hackett, P. B., Plasterk, R. H., & Izsvák, Z. (1997). Molecular Reconstruction of *Sleeping Beauty*, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. *Cell*, *91*(4), 501–510.

13. Ochmann, M. T., & Ivics, Z. (2021). Jumping Ahead with *Sleeping Beauty*: Mechanistic Insights into Cut-and-Paste Transposition. *Viruses*, *13*(1), 76.

14. 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. *Journal of Biological Chemistry*, *277*(37), 34581–34588.

15. 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.

16. Carpentier, C. E., Schreifels, J. M., Aronovich, E. L., Carlson, D. F., Hackett, P. B., & Nesmelova, I. V. (2013). NMR structural analysis of *Sleeping Beauty* Transposase binding to DNA. *Protein Science*, *23*(1), 23–33.

17. Nesmelova, I; Leighton, G.; Yan, C.; Lustig, J.; Corona, R.; Guo, J.; Ivics, Z. (2020). H19Y mutation in the primary DNA-recognition subdomain of the *Sleeping Beauty* transposase improves structural stability, transposon DNA-binding and transposition [Unpublished manuscript].

18. Cui, Z., Geurts, A. M., Liu, G., Kaufman, C. D., & Hackett, P. B. (2002). Structure–Function Analysis of the Inverted Terminal Repeats of the *Sleeping Beauty* Transposon. *Journal of Molecular Biology*, *318*(5), 1221–1235.

19. Ivics, Z., & Izsvák, Z. (2010). The expanding universe of transposon technologies for gene and cell engineering. *Mobile DNA*, *1*(1), 25.

20. Ivics, Z., & Izsvák, Z. (2015). *Sleeping Beauty* Transposition. *Microbiology Spectrum*, *3*(2).

21. Lohe, A. R., Moriyama, E. N., Lidholm, D. A., & Hartl, D. L. (1995). Horizontal transmission, vertical inactivation, and stochastic loss of mariner-like transposable elements. *Molecular Biology and Evolution*, *12*(1), 62–72.

22. Mátés, L., Chuah, M. K. L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D. P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., VandenDriessche, T., Ivics, Z., & Izsvák, Z. (2009). Molecular evolution of a novel hyperactive *Sleeping Beauty* transposase enables robust, stable gene transfer in vertebrates. *Nature Genetics*, *41*(6), 753–761.

23. Izsvák, Z., Ivics, Z., & Plasterk, R. H. (2000). *Sleeping Beauty*, a wide host-range transposon vector for genetic transformation in vertebrates. *Journal of Molecular Biology*, *302*(1), 93–102.

24. Ivics, Z., & Izsvák, Z. (2015b). *Sleeping Beauty* Transposition. *Microbiology Spectrum*, *3*(2).

25. López-Méndez, B., Uebel, S., Lundgren, L. P., & Sedivy, A. (2021). Microscale Thermophoresis and additional effects measured in NanoTemper Monolith instruments. *European Biophysics Journal*, *50*(3–4), 653–660.

26. 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*, *59*(3), 301–315.

27. Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *73*, 1–16.

28. Kleckner, I. R., & Foster, M. P. (2011). An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et biophysica acta*, *1814*(8), 942–968.

29. Scheuermann, T. H., Padrick, S. B., Gardner, K. H., & Brautigam, C. A. (2016). On the acquisition and analysis of microscale thermophoresis data. *Analytical biochemistry*, *496*, 79–93.

30. Romain, M., Thiroux, B., Tardy, M., Quesnel, B., & Thuru, X. (2020). Measurement of Protein-Protein Interactions through Microscale Thermophoresis (MST). *BIO-PROTOCOL*, *10*(7).

31. Deepa, H. R., Thipperudrappa, J., & Suresh Kumar, H. M. (2020). Effect of temperature on fluorescence quenching and emission characteristics of laser dyes. *Journal of Physics: Conference Series*, *1473*(1), 012046.

32. Gupta, A. J., Duhr, S., & Baaske, P. (2018). Microscale Thermophoresis (MST). *Encyclopedia of Biophysics*, 1–5.

33. el Deeb, S., Al-Harrasi, A., Khan, A., Al-Broumi, M., Al-Thani, G., Alomairi, M., Elumalai, P., Sayed, R. A., & Ibrahim, A. E. (2022). Microscale thermophoresis as a powerful growing analytical technique for the investigation of biomolecular interaction and the determination of binding parameters. *Methods and Applications in Fluorescence*, *10*(4), 042001.

34. Nanotemper Technologies. User Manual Monolith NT.115. https://www2.helsinki.fi/sites/default/files/atoms/files/manual_nt115.pdf

35. Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P., & Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*, *1077*, 101–113.

36. Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P., & Duhr, S. (2011). Molecular Interaction Studies Using Microscale Thermophoresis. *ASSAY and Drug Development Technologies*, *9*(4), 342–353.

37. Entzian, C., & Schubert, T. (2016). Studying small molecule–aptamer interactions using MicroScale Thermophoresis (MST). *Methods*, *97*, 27–34.

38. Rahman, A., Choudhary, M. I., & Wahab, A. (2016). Chapter 1 - The Basics of Modern NMR Spectroscopy. In *Solving Problems with NMR Spectroscopy* (2nd ed., pp. 1–34). Academic Press.

39. Cavalli, A., Salvatella, X., Dobson, C. M., & Vendruscolo, M. (2007). Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences*, *104*(23), 9615–9620.

40. Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR*, *6*(3).

41. Lee, W., Tonelli, M., & Markley, J. L. (2014). NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, *31*(8), 1325–1327.

42. Lee, W., Rahimi, M., Lee, Y., & Chiu, A. (2021). POKY: a software suite for multidimensional NMR and 3D structure calculation of biomolecules. *Bioinformatics*, *37*(18), 3041–3042.

43. Teilum, K., Kunze, M. B., Erlendsson, S., & Kragelund, B. B. (2017). (S)Pinning down protein interactions by NMR. *Protein science : a publication of the Protein Society*, *26*(3), 436–451.

45. Yan, C. (2020). *Protein-protein Interaction of Pai-H19y Mutant of the Sleeping Beauty Transposase* (Order No. 28086528). Available from Dissertations & Theses @ University of North Carolina Charlotte; ProQuest Dissertations & Theses Global. (2444387126).