STRUCTURAL AND FUNCTIONAL ANALYSIS OF THE H19Y
MUTATED PRIMARY DNA RECOGNITION SUBDOMAIN
OF SLEEPING BEAUTY TRANSPOSASE

by

Gage O. Leighton

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

2017

Approved by:

_____
Dr. Irina Nesmelova

_____
Dr. Donald Jacobs

_____
Dr. Joanna Krueger

ABSTRACT

GAGE O. LEIGHTON. Structural and Functional Analysis of the H19Y Mutated
Primary DNA Recognition Subdomain of Sleeping Beauty Transposase. (Under the
direction of DR. IRINA V. NESMELOVA)

The Sleeping Beauty transposon system, consisting of the transposon DNA and
the transposase enzyme, is a member of the Tc1/mariner family of DNA transposons.
Although it is an important tool in genetic applications and has been adapted for human
gene therapy, its molecular mechanism remains obscure. Here, we use an experimental
biophysics approach to study the molecular mechanism of the Sleeping Beauty
transposon. We investigate the folding of the specific DNA recognition subdomain of the
Sleeping Beauty transposase, the PAI subdomain. We show that only the folded PAI
binds to DNA, however the amount of unfolded conformation is significant at close to
physiologic conditions. We identify amino acid substitutions that result in stabilization of
its folded conformation. Overall, our results provide a mechanistic insight into DNA
recognition by the Sleeping Beauty transposase and suggest modifications to improve its
activity.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

CAT	Catalytic
CD	Circular Dicroism
DR	Direct Repeats
GOI	Gene of Interest
IF	Integral Fluorescence Intensity
IR	Inverted Repeat.
kbp	kiloBase Pairs
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Enhancement
NOESY	Nuclear Overhauser Effect SpectroscopY
ppm	Parts per million
PT	Photon Technology International
pdb	Protein Data Bank
SB	Sleeping Beauty
NaCl	Sodium Chloride
SLS	Static Light Scattering
tDNA	target DNA
TOCSY	TOtal Correlated SpectroscopY
UV	UltraViolet
WT	Wild-Type

Chapter 1: Introduction

Medical diagnosis, and subsequent treatment, has advanced the cure of many diseases such as polio and scurvy. Countless other ailments are becoming eradicated each year, however, much of the diseases that are most concerning to humanity are those that have no potential to be cured (Danaei, et al., 2011) (Gencer, Pelin, Kucukali, Topcuoglu, & Yilmaz, 2011). These may be the cause of viruses such as HIV or HPV or they may be the cause of mutations to a cells genetic code (Mort, Ivanov, Cooper, & Chuzhanova, 2008). This genetic code, forefront to the central dogma of molecular biology, is essential to not only the synthesis of every component within the cell but also reactions to environmental condition. Much like the spelling errors within this document, mutations to the genetic code can result in new words or phrases or in contrast a garbled mess of nonsense. Over time, the accumulation of mutations can lead to the evolution of a species, but when they are acquired quickly and not repaired correctly, they may lead to cell malfunctioning, i.e., cancerous cell (Tomoshige, et al., 2015).

Diseases such as hemophilia; the inability of a patient to clot blood in response to injury, is often the result of "damage" to the cells genetic material through mutations passed from mother to child (Judd, 10. Blood clotting deficiency disorders, 2013). Hemophilia cannot be cured by conventional medicine and, depending on the severity, can be a debilitating lifelong condition whereby the only treatment is frequent transfusions of plasma or replacement of the missing clotting factor.

Another lifelong disease, sickle cell anemia, results from a single mutation to the patient's DNA (Judd, 11.3 Sickle cell disease, 2013). Sickle cell anemia is defined by

abnormally shaped red blood cells, which are poor carriers of oxygen and can clog small blood vessels. It too remains incurable by current medicine. Simply correcting the inflicted individual's DNA sequence would resolve the problem but at present is a challenge seeming impossible to market.

To correct a single mutation there must be a mechanism established to replace the damaged gene without altering the rest of the genome. In addition, this must be done in enough cells within the host so it becomes persistently expressed in each subsequent cell generation. One method of editing the human genome, established within the past 10 years, is the CRISPR/cas9 system (Yamamoto, 2015). It has the potential to delete genes or modify multiple genes at once. CRISPR has shown potential as a useful tool for gene editing countless times (Suzuki, et al., 2014) (Smith, et al., 2014). Although it may currently lead the scientific community with high hopes at an easy solution to gene therapy it has a few disadvantages. For one it works by utilizing the natural repair mechanism of the cell, which, if used incorrectly, can introduce its own mutations or insert multiple copies of the same gene into the genome (Yamamoto, 2015).

An alternative approach to deliver the genetic information to a host cell is to utilize a natural ability of DNA transposons to move (transpose) from one location to another (Hackett, Largaespada, Switzer, & Cooper, 2013) (Hudecek, et al., 2017). While several DNA transposons are being developed, the Sleeping Beauty (SB) has the highest activity in mammalian cells and is the most widely used transposon for genetic applications (Geurts AM). Other transposons being developed include Tol2, Frog Prince, and piggyBac, however, the Sleeping Beauty transposon system remains the most active of

the group with respect to vertebrate animal cells (Mátés L, 2009). The SB transposon

system is composed of two components: the DNA transposon, the DNA sequence to be

inserted flanked by inverted terminal repeats (IRs), and the transposase enzyme that

catalyzes the reaction. Inverted terminal repeats contain sequences of DNA that the

transposase enzyme identifies and subsequently binds to. The sequence of the IRs has

been studied extensively and shown to be a critical component in regulation of the

transposition mechanism (Wang, et al., 2017) (Izsvak Z. , et al., 2002) (Cui, Geurts, Liu,

Kaufman, & Hackett, 2002). This system can excise the transposon, containing the gene

of interest (GOI), without damage to the surrounding DNA sequence.

The SB transposase enzyme belongs to a superfamily of enzymes containing a

catalytic motif DDE or DDD (Nesmelova IV) (Yuan & Wessler, 2011). These amino

acids, are highly conserved between all eukaryotes suggesting that they have a common

evolutionary origin. The transposase enzyme contains two structurally and functionally

independent domains, the DNA binding domain and the catalytic (CAT) domain.

Furthermore, the DNA binding domain is comprised of two subdomains, the PAI and

RED subdomains (Figure 1B). It is known that the PAI subdomain, the primary DNA

recognition subdomain, binds to specific regions of the IRs but the RED subdomains

function is still unknown (Izsvak Z. , et al., 2002) (Carpentier, et al., 2014).

The efficiency of the SB transposon system depends on several factors such as

environmental conditions, the length of the GOI, and specific mutations within the

transposase enzyme. The GOI can vary greatly in size with a linear dependence on
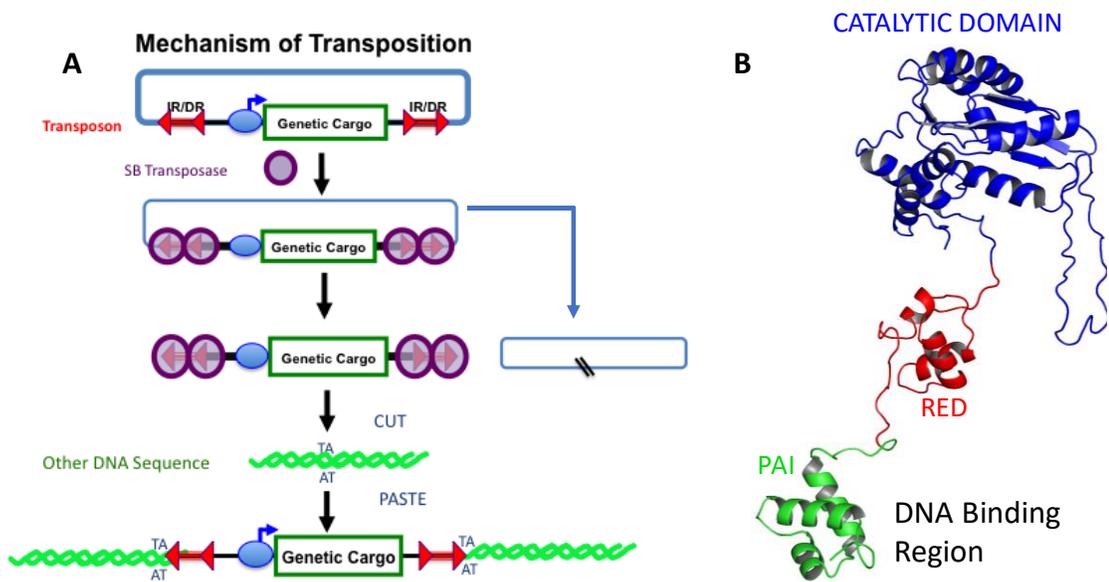
Figure 1. Mechanism of SB-mediated transposition. (A) Top line: A transposon, defined by the inverted sets of red, double arrows (IR/DRs), is shown as contained in another DNA molecule (e.g. a plasmid shown by the blue lines). The transposon in this example harbors an expression cassette comprised of a promoter (blue oval) that can direct transcription of the gene or other DNA sequence labeled "genetic cargo". *Middle lines:* Sleeping Beauty (SB) transposase binds to the IR/DRs as shown and cuts the transposon out of the plasmid (the cut sites are indicated by the two black slashed lines in the remaining plasmid). *Bottom two lines:* Another DNA molecule (green) with a TA sequence can become the recipient of a transposed transposon. In the process, the TA sequence at the insertion site is duplicated. (Hackett & Zerull, 2011) (B) Proposed full length structure of the Sleeping Beauty Transposase. The primary DNA binding domain (PAI) shown in green, the RED subdomain as red, and the catalytic domain in blue. Separately determined structures are published as pdb code 2M8E, 5UNK, and 5CR4, respectively.

transposon length allowing the successful transposition of genes up to 8 kilobase pairs

(kbp) (Geurts AM) (Izsvak, Hackett, Cooper, & Ivics, 2010). A mutant of the transposase

enzyme, SB100X, was engineered using molecular evolution to have 100-fold

enhancement in efficiency when compared to the original (wild-type) transposase (Mátés

L, 2009). This mutant contains a total of 9 mutations; K14R and K33A within the PAI

subdomain, K115H in between the DNA binding domain and the catalytic domain, and

RKEN214-217DAVQ, M243H, and T314N located in the CAT domain. It has been

proposed that the position of the mutations in the CAT domain contribute to the overall

increase in efficiency in a few ways. The RKEN214-217DAVQ mutations are located on

the target DNA (tDNA) binding region of the protein and work to effectively better shape

4

and position the protein relative to DNA. The M243H mutation is located next to the second catalytic residue D244 and is responsible for positioning and stabilizing this residue into the active site by formation of a parallel-displaced pi-stack with H249 (Voigt, et al., 2016). The last mutation within the CAT domain, T314N, is positioned on the exterior of the protein and likely increases the solubility of the protein, which is important to efficient transposition (Mátés L, 2009).

The K33A mutation has been shown to increase the activity of SB transposase by 400% (Yant, Park, Huang, Mikkelsen, & Kay, 2004). However, both mutations, K14R and K33A, are required (Mátés L, 2009). These mutations are in the PAI subdomain, suggesting that the primary DNA-recognition subdomain of the SB transposase is not optimal and can be further modified for improved activity. The goal of this work is to investigate physicochemical properties of the PAI subdomain in order to identify the requirements for its optimal functioning.

Chapter 2: Materials and Methods

2.1 - Protein expression and purification.

    The production of protein sample for experiments involved expression in bacterial host, purification using affinity chromatography techniques, and transfer to buffer at the concentration suitable for a given experiment. These steps are listed in detail below.

    DNA plasmid encoding the His-tagged PAI subdomain of SB transposase was ordered from GenScript USA (Piscataway, NJ). The DNA encoding the full-length DBD of SB transposase was cloned into pET 21a(+) vector at the NheI and XhoI restriction sites. Plasmids encoding full-length DBD or PAI subdomain of SB transposase were transformed into competent BL21-A1 Escherichia coli cells. Cells were grown in LB or M9 minimal media with ampicillin at 30°C to an OD600 of 0.4–0.6, and then induced by adding 0.1M IPTG and 0.2% l-arabinose for 4 h. The cells were collected by centrifugation and lysed using B-Per lysis buffer. Soluble extract containing DBD or PAI subdomain of SB transposase was prepared by centrifugation of cell lysate at 14,000g for 1 h. The protein was purified using a TALON Ni Affinity Resin (ClonTech, Mountain View, CA). After elution from Ni-affinity column, the proteins were dialyzed against 25 mM aqueous sodium phosphate buffer at a range of 5.0 to 7.0 pH values. The presence and purity of proteins were monitored using 15% (w/v) polyacrylamide gel electrophoresis. The details of this procedure were previously published (Carpentier, et al., 2014) (Leighton, et al., 2014).

2.2 - Circular Dichroism Spectroscopy

Circular dicroism (CD) spectroscopy was used for the assessment of protein secondary structure. The advantages of this method include low sample consumption, fast experiment execution, and relatively straightforward data interpretation. The method registers the difference in absorption of right or left circularly polarized light by chiral chromophores. The wavelengths of far ultraviolet (UV) light are useful for analysis of protein secondary structure which is sensitive to its environment. Therefore, it is useful to observe the change in CD spectra upon changes in temperature, salinity, and/or pH. We selected to use CD spectroscopy in our work for the assessment of alpha-helical conformation in the PAI subdomain dependent on experimental conditions such as solution pH, presence of sodium chloride (NaCl), temperature, and modifications to protein amino acid composition. The detailed procedures are described below.

CD measurements were performed on a Jasco-715 spectropolarimeter, equipped with a Peltier temperature control system, using a quartz glass cell with a pathlength, l, of 1 mm. Far-UV CD spectra were recorded in the range of 190 to 250 nm at 22°C. The corresponding buffer baseline was subtracted from each spectrum. Spectra were recorded using a 50 nm/min scan rate with a 2 s response and a 1 nm bandwidth. Reported spectra are averages of 4-8 scans and are expressed as mean-residue molar ellipticity, $[\theta]$, calculated using the relation

$$[\theta] = \frac{M_o \theta_\lambda}{100 \cdot C \cdot l} \quad . \tag{1}$$

Here, M_0 is the mean residue molar mass, θ_λ, is the measured ellipticity in degrees, and C is the total concentration of protein. The value of M_0 was obtained by dividing the molecular weight of the protein with the number of amino acid residues in it.

To follow the folding of wild-type (WT) PAI subdomain and its mutants, CD spectra were collected at different concentrations of NaCl or pH. Mean residue ellipticity at 222 nm, $[\theta]_{222}$, was used to assess protein structural changes. CD spectra were analyzed on the DichroWeb server (Whitmore & Wallace, DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, 2004) (Whitmore & Wallace, Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases, 2008) to estimate the fraction of secondary structure elements. The dependence of the fraction of alpha-helical conformation, f_h, on pH was fitted to a modified Henderson-Hasselbalch equation (Cameselle, Ribeiro, & Sillero, 2010) given below:

$$f_h = \frac{f_a + f_b(10^{n(pKa-pH)})}{1+10^{n(pKa-pH)}} \tag{2}$$

In this equation, f_a is the fraction of the alpha-helical conformation at acidic pH prior to transition, f_b is the fraction of alpha-helical conformation at basic pH after transition, pKa is the pH value corresponding to an inflection point of the dependence, and n value (Hill coefficient) is the slope at the inflection point, which determines the number of protons involved in the transition. The Hill coefficient was set to be a free parameter during fitting.

2.3 - Intrinsic tyrosine fluorescence and Rayleigh Light Scattering

Intrinsic tyrosine fluorescence was used to monitor the change of the PAI subdomain folding state with temperature. Proteins derive their intrinsic fluorescence from the chromophores phenylalanine, tyrosine, and tryptophan. The PAI subdomain contains only one tyrosine (Tyr46) and does not have phenylalanine or tryptophan residues. The side chain of Tyr46 is oriented towards the interior of the protein (*Figure 4*, left panel), and thus could be used as a probe of the PAI subdomain unfolding.

Intrinsic tyrosine fluorescence and right-angle static light scattering (SLS) measurements were done on a PTI (Photon Technology International) QuantaMaster fluorescence spectrofluorometer on the same sample. The temperature of the jacketed cell-holder was maintained by circulating water and monitored by a Hanna Instruments 93530 K-thermocouple thermometer. The excitation wavelength was 275 nm, and the fluorescence emission was collected from 290 to 450 nm. Fluorescence data are presented as the total integral fluorescence intensity, IF, calculated by integrating the area under the fluorescence curve. IF is temperature-dependent and decreases with increasing temperature according to the Arrhenius law.

$$e^{\left(\frac{-E_{eff}}{RT}\right)} \tag{3}$$

Here, R is the gas constant and E_{eff} can be interpreted as the activation energy of the processes that lead to fluorescence quenching [26-27]. Temperature dependence was ranged from 5°C to 55°C at both pH 5.0 and 7.0. The reported data is the average of three independent measurements of the temperature dependence of IF done using a new protein sample.

2.4 - Nuclear Magnetic Resonance spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is the only technique that allows high-resolution protein structure determination in solution. It uses the intrinsic property of an atom of an odd numbered isotope called spin to analyze the amount of electromagnetic energy absorbed and released by an atom dependent on its local environment. In our work, we used NMR to determine structural changes in the PAI subdomain caused by the addition of NaCl, changing temperature, or by amino acid residue substitutions. The use of NMR chemical shifts are highly sensitive reporters of changes in the nuclear environment and thus should be a beneficial technique to study the effect of these environmental conditions on folding (Hass & Mulder, 2015).

For high-resolution NMR studies, the protein has to be labeled with ^{15}N and/or ^{13}C isotopes. We achieved this by growing bacteria in minimal medium where ^{15}N-isotopically enriched ammonium chloride was used as a sole source of nitrogen and ^{13}C isotopically enriched glycerol was used as a sole source of carbon. The purification was done according to the same procedure that was used for unlabeled protein.

All NMR experiments were performed in the range of 278 to 293 K on a Bruker Avance-III 950 MHz spectrometer equipped with CryoProbe^{TM}. When the NMR spectrum is acquired, signals must be identified and assigned to a specific proton within the protein. The backbone assignments for the PAI subdomain were derived from 2D and 3D ^{15}N-HSQC TOtal Correlated SpectroscopY (TOCSY) and ^{15}N-HSQC Nuclear Overhauser Effect SpectroscopY (NOESY) experiments using a ^{15}N-labled PAI subdomain. Inter-proton distance restrains, required for NMR structure calculation, were

derived from nuclear Overhauser enhancement (NOE) signals in ^{15}N-NOESY-HSQC experiments, collected at 80 and 150 ms mixing times. The NMR data were processed with the NMRpipe program (Delaglio, et al., 1995) and visualized using the CARA and NMRView programs (Johnson & Blevins, 1994). The NOE (Nuclear Overhauser Effect) signal is typically observed between two nuclei located within 4Å-5Å of each other (Neuhaus & Williamson, 2000), thus providing the information on which atoms within the protein are close in space. This information is used during the simulated annealing procedure to generate the ensemble of minimum energy structures with atoms positions satisfying experimental distance restraints. NMR structures were calculated using the program XPLOR-NIH (Schwieters, Kuszewski, Tjandra, & Clore, 2003) on the basis of NOE and dihedral angle restrains generated using the program TALOS (Shen, Delaglio, Cornilescu, & Bax, 2009) and chemical shift data. The 15 minimum energy structures were selected from a set of 100 calculated structures as a representative ensemble based on the absence of NOE violations greater than 0.5 Å and dihedral angle violations greater than 5°.

2.5 - Estimation of ΔΔG

The ERIS protein prediction server allows for multiple mutations within their online user interface (Yin, Ding, & Dokholyan, 2007). Starting from the protein data bank (pdb) code for the PAI subdomain, 2m8e (Carpentier, et al., 2014), we simulated mutations to the three histidine residues separately or in combinations with each other. The flexible backbone option was selected however, backbone pre-relaxation was not used (Yin, Ding, & Dokholyan, 2007).

Chapter 3: The PAI subdomain of the SB transposase must be folded for optimal functioning.

In this chapter, I review data showing that the PAI subdomain does not form a stable structure at experimental conditions close to physiologic. Some of these results have been already published, but they are presented for complete picture of the PAI subdomain folding.

3.1 - The effect of the addition of sodium chloride

Typically, the presence of NaCl at the concentration above 600 mM, solution pH above 7.0, and low temperatures (5°C) are required for the PAI subdomain to adopt a stable conformation (Carpentier, et al., 2014) (Leighton, et al., 2014).

Figure 2 shows the effect of NaCl on the folding of the PAI subdomain. The CD spectra in *Figure 2*A show that increasing NaCl concentrations results in the shift from a global minimum at 200 nm to having two local minimum at 208 and 222 nm with an additional local maximum at 198 nm. These spectra show that there is an increase in alpha-helical conformation of the PAI subdomain with increasing NaCl concentration. NMR shows similar results. *Figure 2*B shows the 2D [1H,15N]-HSQC spectrum of the PAI subdomain, where each signal originated from covalently attached proton and nitrogen. In the protein, these are typically NH amide bonds in the protein backbone. Accordingly, horizontal and vertical axes show amide proton and nitrogen chemical shifts, respectively. When the protein is in random coil conformation the local environment of amide proton and nitrogen of each amino acid in the protein is not unique (only observing its neighbors and the solvent). Therefore, the position of peaks in a 2D NMR spectrum begin to overlap with each other. In contrast, in a folded conformation the

amide proton and nitrogen exist in unique environments (uniquely surrounded by other

amino acids) resulting in a NMR spectrum with a broad dispersion of signals. The

dispersion of peaks in the 2D [^{1}H,^{15}N]-HSQC NMR spectra of the PAI subdomain at 600

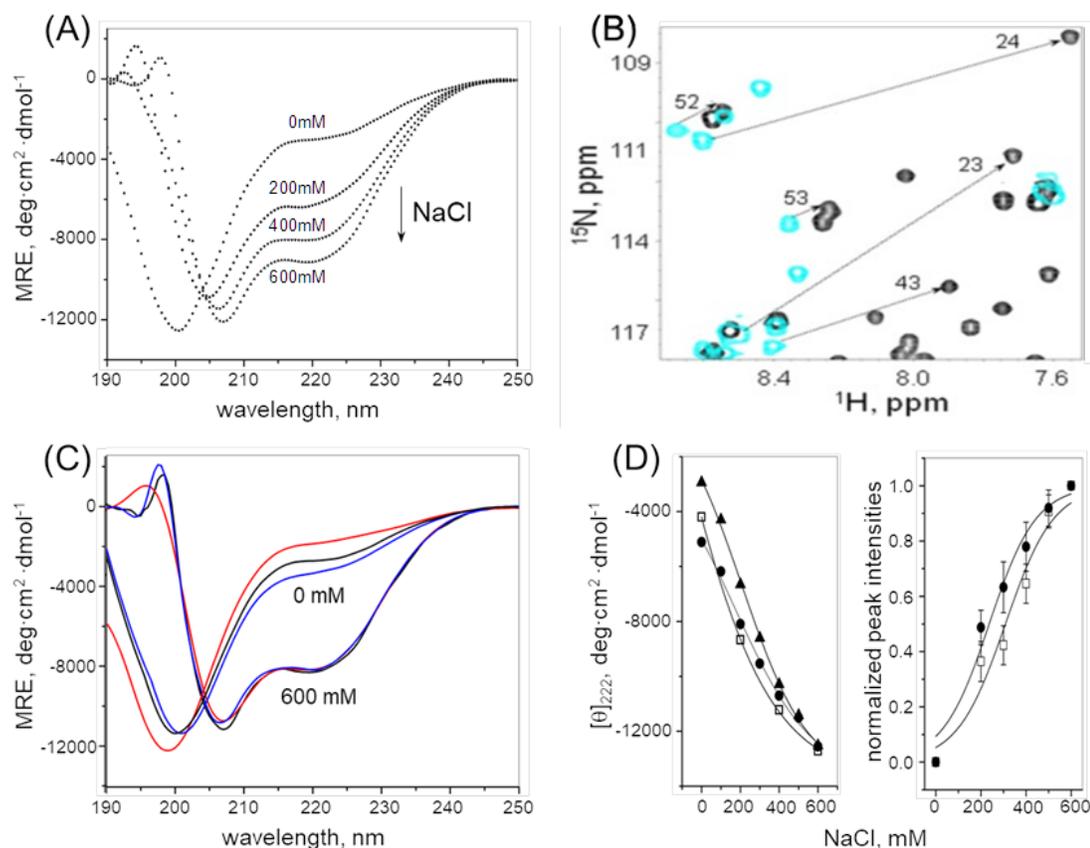

Figure 2. NaCl-induced folding of the PAI subdomain and its mutants K33A and K14RK33A. (A) Far UV CD spectra of PAI subdomain collected in 25 mM aqueous sodium phosphate buffer at pH 5.0 and NaCl concentrations of 0, 200, 400 and 600 mM. Increasing the concentration of NaCl induces an alpha-helical structure. (B) 2D [1H,15N]-HSQC spectra of the PAI subdomain. Blue and black cross-peaks represent the spectra acquired at 0 and 600 mM NaCl, respectively. The increasing chemical shift dispersion of NMR cross-peaks with the increasing concentration of NaCl indicates that the PAI subdomain folds. Representative assignments are indicated (12). (C) Far UV CD spectra of the PAI subdomain (black), K33A (red), and K14RK33A (blue), collected in 25 mM aqueous sodium phosphate buffer at pH 5.0 and NaCl concentrations of 0, and 600 mM. The content of alpha-helical conformation at 0 mM NaCl is less in K33A mutant, than in the original PAI subdomain and K14RK33A mutants, but similar at 600 mM NaCl. (D) The values of mean residue ellipticity at 222 nm of the PAI subdomain (squares), K33A (triangles), and K14RK33A (circles) (left panel) and the normalized average intensity of the NMR cross-peaks of the PAI subdomain (squares) and K14RK33A (circles) (right panel) plotted as a function of NaCl concentration. Solid lines represent a global sigmoid dose-response fit of the data yielding the concentration corresponding to the middle point of transition Cm = 230—76 mM.

mM NaCl, *Figure 2*B, shows that the PAI subdomain has a stable folded conformation (Carpentier, et al., 2014).

To find out whether the K33A and K14R mutations increase the structural stability of the PAI subdomain, as one of the possible molecular mechanisms of SB activity enhancement, we investigated the folding properties of modified PAI, where amino acid substitutions K33A or K14RK33A were introduced (K33A or K14RK33A mutants). The content of alpha-helical conformation that forms at 600 mM NaCl is similar in the original PAI subdomain and its K33A or K14RK33A mutants as judged by the resemblance of the overall shape of their CD spectra (new, unpublished data, *Figure 2*). Similar NaCl-dependent folding of the original PAI, K33A, and K14RK33A mutants, suggests that the effect of NaCl is likely due to a non-specific screening of the overall positive charge of the protein, and that both the K33A and K14R mutations do not affect the NaCl induced folding of the PAI subdomain. This indicates that a different mechanism of increasing activity is the role of these mutations. The normalized average intensity of CD data, $[\theta]_{222}$, and individual cross-peaks of 2D [^{1}H,^{15}N]-HSQC spectrum show a sigmoidal dose-response dependence on NaCl concentration (*Figure 2*). The middle of transition to the folded state, C_m, is calculated to be about 230 ± 76 mM of NaCl, averaged for the WT PAI subdomains and its mutants. This concentration is 80 mM higher than the nominal concentration of NaCl found within the nucleus of a living cell (Terry, Fernandez, Gude, Lorente, & Grant, 2011). Thus, we expect that the efficiency of the Sleeping Beauty transposon system to be diminished in mammalian cells.

3.2 - The effect of solution pH

The effect of pH on the folding of the PAI subdomain was also examined by using far-UV CD spectra (190-260 nm) in the range of pH values from 4.5 to 8.5. Mean residue ellipticity at 222 nm, $[\theta]_{222}$, was used to assess the protein structure changes with increasing pH (new, unpublished data, Figure 3). The data indicate that basic pH induces helix formation in PAI subdomain and its mutants. Analysis of alpha-helical conformation calculated using the DichroWeb server *(Whitmore & Wallace, DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data, 2004)* show and increase from about 10% to 40% when the 20 mM sodium phosphate buffered solution changes pH from 4.5 to 8.5. Solid lines represent the least squares fits of a modified Henderson-Hasselbalch equation (2) to the data, based on the assumption of a two-state transition. The pH dependence of the PAI subdomain has a folding transition around pH values 6.0-6.2, which coincides with the pKa value of histidine sidechain protonation (when protonated, it becomes positively
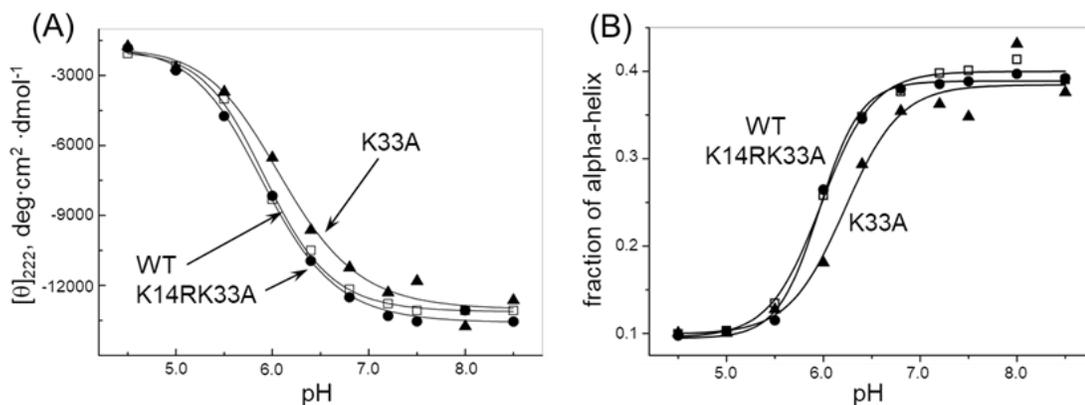
Figure 3. pH-induced folding of the PAI subdomain and its mutants K33A and K14RK33A. The dependence of $[\theta]_{222}$ (**A**) and of the alpha-helical conformation. (**B**) on solution pH is shown for the PAI subdomain (squares), K33A (triangles), and K14RK33A (circles). The alpha-helical content has been estimated using the DicroWeb server. Solid lines represent the sigmoid dose-response fit of experimental data using Equation 2.

charged). Thus, the folding of the PAI subdomain at basic pH is likely related to

deprotonation of histidine side chain(s).

3.3 - *The effect of temperature*

Temperature may affect the folding of the protein, because local motions increase at

higher temperatures, hence leading to large spatial fluctuations. Intrinsic tyrosine

fluorescence was used to monitor the change of the PAI subdomain folding state with

temperature. Proteins derive their intrinsic fluorescence from the chromophores

phenylalanine, tyrosine, and tryptophan. The PAI subdomain contains only one tyrosine

(Tyr46) and does not have any phenylalanine or tryptophan residues. The side chain of

Tyr46 is oriented towards the interior of the protein (*Figure 4*, left panel), and thus

should be sensitive to the PAI subdomain unfolding. The total integral fluorescence
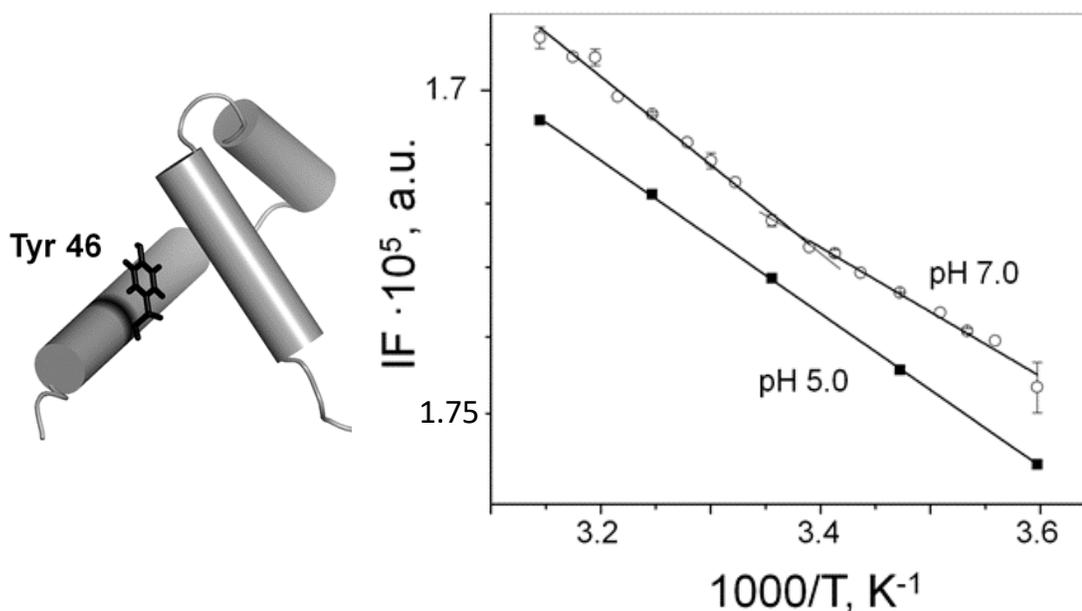
Figure 4. Intrinsic tyrosine fluorescence of WT PAI Subdomain. (Left panel) The location of Try46 on the cartoon representation of the PAI subdomain (PDB code 2m8e [7]) is shown. (Right panel) The integral fluorescence of Tyr46 at pH 5.0 (squares) and pH 7.0 (circles), calculated as 1/area under the spectrum, is plotted vs. temperature. Shown data is the average of three independent experiments. Error bars in many cases do not exceed the size of the symbol. Solid lines represent best fits of experimental data.

intensity as a function of temperature is plotted in *Figure 4* (right panel) for pH values of

5.0 and 7.0 using a semi-logarithmic scale. At pH 5.0, the temperature dependence of

Tyr46 fluorescence is linear, as expected, because the PAI subdomain remains unfolded

and the environment of Tyr46 does not change. At pH 7.0, the transition between 18 and

28°C is observed, indicating that there is a change in the Tyr46 environment due to PAI

unfolding, in agreement with our [^{1}H,^{15}N]-HSQC data shown in *Figure 5*. We note that

the effect is not very strong. This is likely due to the fact that even in a folded state Tyr46

has good water accessibility due to the small size and flexibility of the PAI subdomain.

Figure 5 shows a series of 2D [^{1}H,^{15}N]-HSQC spectra of the PAI subdomain

collected at solution pH values 5.0 and 7.0 and at temperatures varying from 5 to 45°C.

The [^{1}H,^{15}N]-HSQC at 5°C and pH 5.0 reveals very limited chemical shift dispersion,

with the majority of the ^{1}H resonances around 8 ppm (parts per million). This indicates

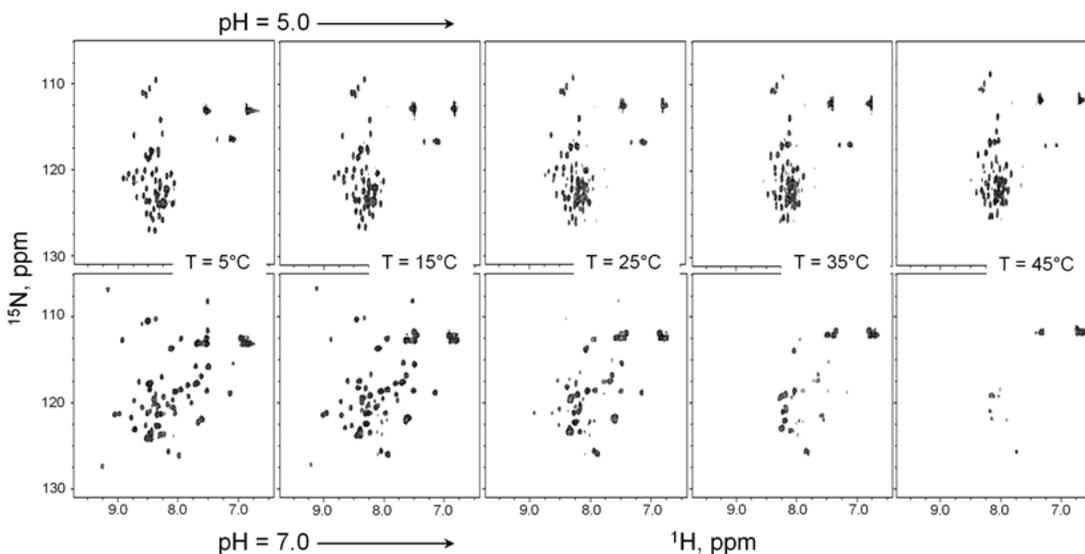that the respective amino acid residues are in random coil conformation and the PAI

Figure 5. Temperature effects on 2D [1H,15N]-HSQC NMR Spectra of the PAI subdomain. The spectra were collected in 25 mM aqueous sodium phosphate buffer at pH 5.0 (top panel) and 7.0 (bottom panel) in the range of temperatures from 5 to 45°C with a 5°C increment.

subdomain is essentially unfolded. In contrast, at pH 7.0, many resonances shift and become well dispersed, indicative of a folded structure, where different amide protons are located in diverse chemical environments, hence diverse local magnetic fields leading to the difference in amide proton and nitrogen chemical shifts. Increasing the temperature has very little effect on the chemical shift dispersion and signal intensity in the [^{1}H,^{15}N]-HSQC spectrum of the PAI subdomain at pH 5.0, implying that the PAI subdomain remains unfolded at all temperatures between 5 and 45°C. On the contrary, drastic changes are observed in the [^{1}H,^{15}N]-HSQC spectra of the PAI subdomain at pH 7.0. As the temperature increases, many signals become severely broadened and only a few signals remain observable at temperatures above 35°C. Observed signal broadening may be due to the conformational exchange between the folded and unfolded states of the protein, protein aggregation, or both. Our diffusion data excluded the PAI subdomain aggregation and confirmed that this effect was due to protein unfolding (Leighton, et al., 2014).

3.4 - The folded structure is required to bind to the DNA

The PAI subdomain is the primary DNA-recognition subdomain of SB transposase. We were interested to determine whether the folding of the PAI subdomain is required for binding the transposon DNA, because it could have direct implications for the function of SB transposase. As shown above, the PAI subdomain does not form a stable, well-folded structure, but exists in the equilibrium of slowly interconverting on the NMR time scale folded and unfolded conformations. We identified experimental conditions at which both the folded and unfolded conformations were present in solution at approximately equal

concentrations. Two sets of resonances originating from folded and unfolded

conformations are observed in the [^{1}H,^{15}N]-HSQC spectrum of the PAI subdomain at pH

5.0 in the presence of 250 mM NaCl (*Figure 6*A). This property of the PAI subdomain

provides a unique opportunity to monitor the binding of DNA to each conformation

independently. Here, we investigated the ability of folded and unfolded PAI subdomain

to bind the 18 bp direct repeats (DR)-core sequence, that represents the minimal sequence

required for transposase binding (Cui, Geurts, Liu, Kaufman, & Hackett, 2002; Cui,

Geurts, Liu, Kaufman, & Hackett, 2002). Previously, we have shown that the DNA-

binding of the PAI subdomain occurs in the intermediate regime on the NMR time scale

leading to the broadening of the PAI resonances caused by exchange between the DNA-

bound and unbound states (Carpentier, et al., 2014). We therefore analyzed the effects of

the protein binding to the transposon DNA on the basis of retention of peak intensities for

each residue.

   To determine whether the transposon DNA binds to the folded, unfolded, or both

conformations of the PAI subdomain, we monitored the changes of resonance intensities

in the [^{1}H,^{15}N]-HSQC spectra upon the addition of increasing to saturation concentrations

of DR-core DNA. A set of non-overlapping resonances observed for both PAI

conformations was used for the analysis. This set comprised the resonances originating

from Q11, D12, V18, G28, A29, R38, Y48, and H50 amino acid residues (*Figure 6*A). Of

these residues, G28, G29, and R38 were located in the DNA-binding site (*Figure 6*B).

While the folded conformation of the PAI subdomain demonstrates well-dispersed

resonances in the [^{1}H,^{15}N]-HSQC spectrum, the number of residues that can be used in

the analysis was limited due to overlap of many resonances originated from the unfolded
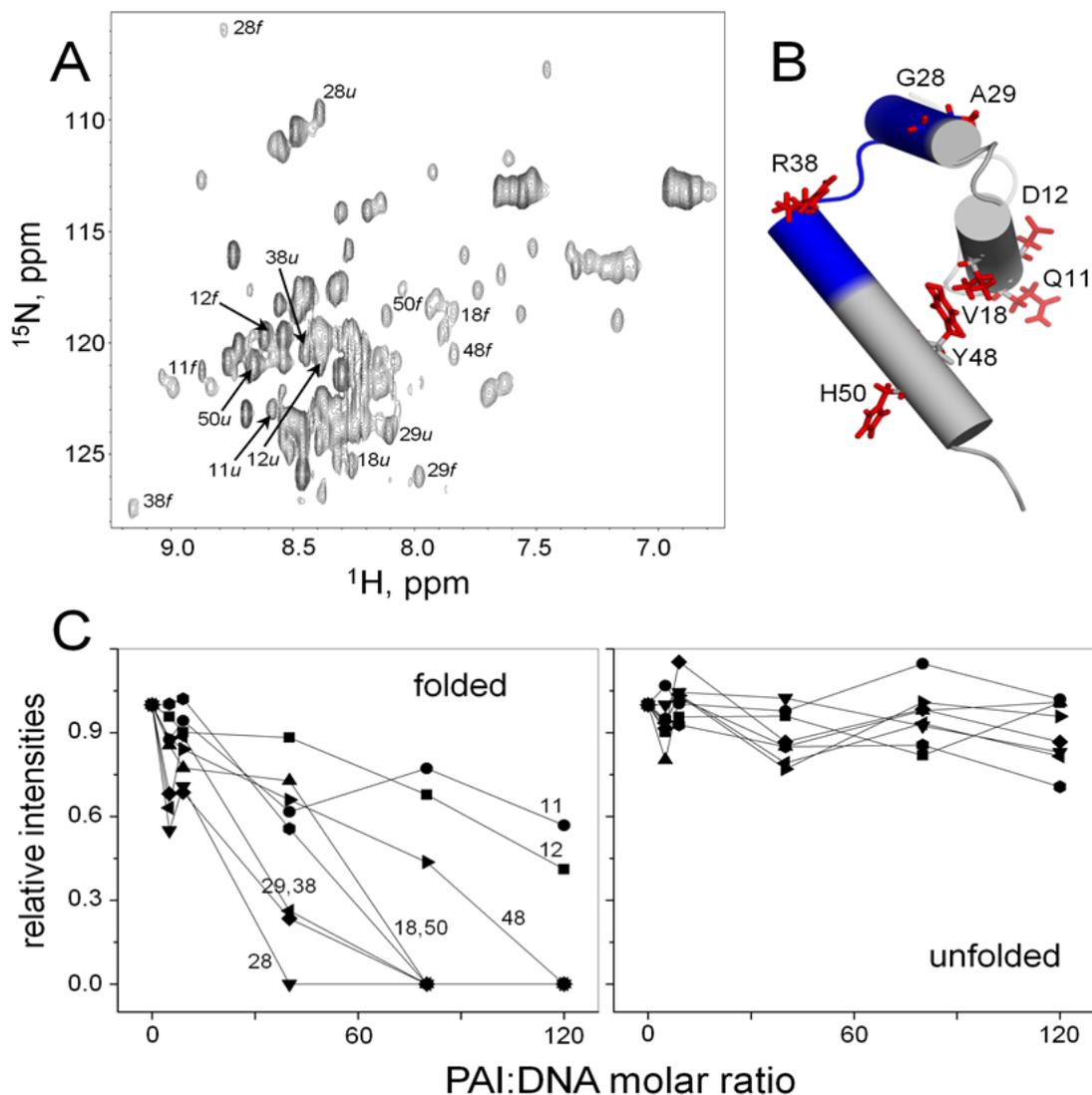
Figure 6. DNA-binding of folded and unfolded conformations of the PAI subdomain. (A) 2D [1H,15N]-HSQC spectrum of the PAI subdomain collected at the temperature of 5°C in 25 mM aqueous sodium phosphate buffer at pH 5.0 in the presence of 250 mM NaCl. The folded and unfolded conformations exist in slow exchange on the NMR time scale. Thus, two resonances are observed for each residue. Non-overlapping resonances originating from the same residue in both the folded and unfolded conformations are labeled. (B) Cartoon representation of the PAI subdomain (PDB code 2m8e [7]). The DNA-binding site is colored blue. Side chains of the residues that were used for the analysis of the DNA-binding are labeled and shown as red sticks. (C) Relative intensities of resonances corresponding to the folded and unfolded conformations are plotted as a function of PAI:DNA molar ratio. Relative intensities were calculated by dividing the resonance intensity at a given PAI:DNA molar ratio by the intensity of this resonance in the absence of DNA.

PAI subdomain. *Figure 6*C shows resonance intensities as the function of increasing to

saturation concentrations of the DNA-core, normalized by the intensity of the respective

resonance measured in its absence. Notably, only the resonances corresponding to the

20

folded PAI subdomain appear to be affected by the presence of DNA-core sequence, indicating that it is the folded PAI subdomain that binds to the transposon DNA. As expected, the resonances originating from the residues in the DNA-binding site show the largest changes. Thus, only the folded PAI subdomain is capable of binding to the transposon DNA.

In summary, the data presented in Chapter 3 show that the folding of the PAI subdomain is sensitive to the presence of salt (NaCl), solution pH, and temperature. At experimental conditions close to physiologic, the PAI subdomain does not form a stable structure. However, the PAI subdomain must be folded for binding to the transposon DNA, hence for the optimal functioning of the SB transposase. Our data suggest that the protonation of a histidine side chain is likely the cause of folding/unfolding transition in the PAI subdomain (*Figure 3*). Thus, substitution of histidine amino acid could be beneficial for folding and result in higher activity of the PAI subdomain. We therefore investigated the possibility of using the histidine substitution to promote the folding of the PAI subdomain.

Chapter 4: Histidine H19Y mutation acts as a conformational switch in the PAI subdomain leading to stable folded conformation.

The PAI subdomain contains three histidine amino acids, His19 is centrally located and directed towards the interior of the folded conformation while His48 and His49 are positioned exteriorly on helix 3 (*Figure 7*A). It can be expected that the protonation of the His19 residue is the greatest contribution to destabilization of the folded structure. Utilizing the ERIS protein stability prediction server, we estimate the change in free energy of the folded conformation (*Figure 7*B). The more negative values of Gibbs free energy correspond to more stable conformation. Calculations of protein stability results in negative changes in free energy or stabilization of structure when introducing the K14R or K33A mutations. The change is small, in agreement with our experimental data that do not show any change due to the introduction of K33A mutation; however, combining two mutations leads to a larger effect, as illustrated by group one in *Figure 7*B, although we do not observe any noticeable change experimentally. Single mutations to each of the three histidine residues results in stabilization when the substitution is of similar size and neutrally charged, displayed by group two of *Figure 7*B. Combinations of the K14RK33A mutations with favorable histidine mutations show the greatest and significant increase in structural stability, shown by group three in *Figure 7*B. These data suggest that the substitution of H19 of a similar sized and neutral charge could result in stabilization of the PAI subdomain, especially at lower pH values. Therefore, we selected the H19Y mutation for experimental verification of stabilization seen in predictions of the Eris server.

Routine expression of the mutated PAI subdomain using the procedure outlined in the

materials and methods section prove no different than when expressing the WT PAI

subdomain. Experimental verification of the purified triple mutant, K14RH19YK33A, by

NMR spectroscopy, in 25 mM sodium phosphate buffer at pH 5.0 and no NaCl show that

the dispersion of peaks relative to the WT PAI subdomain is suggestive of a folded

protein (Figure 8A, red and blue spectra, respectively). Backbone assignments were done

using the ^{15}N-HSQC NOESY and ^{15}N-HSQC TOCSY spectra. 90% of assignments are

complete and the preliminary NMR H19Y structure, (Figure 8B), shows resemblance to

the structure of the original PAI subdomain. These were performed by assigning the

individual TOCSY slices that correspond to each unique amino acid residues within the

protein (Figure 9, Left Panel). In a TOCSY spectrum, each slice taken at a particular

frequency of nitrogen, contains only those signals that originate from protons a particular

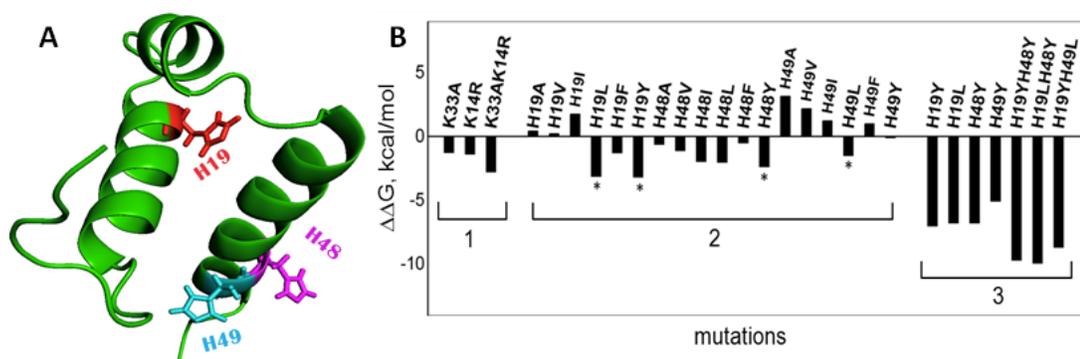amino acid residue in the protein (eg., Ser40 has 4 signals, on from amide proton H, one

Figure 7. The effect of mutations on the Gibbs free energy of unfolding of the PAI subdomain. (A) The three-dimensional cartoon representation of the PAI subdomain with histidine residues displayed as sticks. (B) Group 1 shows the effect of K14R and K33A mutations alone or in combination. Group 2 shows the effect of H19, H48, and H49 single mutants. The most energetically favorable mutations are labeled by stars. They were used to calculate for G calculations in group 3. Accordingly, group 3 shows the effect of H19, H48, and H49 single or double mutants in the presence of two hyperactive mutations K14R and K33A. The values of G were calculated using the Eris protein stability prediction server (37). **(C)** The location of K14 residue on helix 1 and two negatively charged residues D10 and D17 at positions (i, i-4) and (i, i+3) from K14, respectively, are shown.

from HA proton, and two from HB proton). The unique chemical structure of each amino acids allows for differentiation between different amino acids (e.g., compare Ser40 and Val41 in Figure 9). Unlike the TOCSY spectrum, the NOESY spectra contains signals from all protons that are in close proximity to each other, but not necessarily belonging to the same amino acid residue. Hence, the slice from the NOESY spectrum, taken at the same nitrogen frequency as TOCSY slice, contains additional signals from neighbor amino acid residues. All backbone NH bonds are clearly visible and must be aligned with the corresponding neighbors in both forward, red arrow, as well as backwards, green arrow, signals shown in the right panel of Figure 9. Any additional interactions between protons of the sidechain will appear as new peaks in the NOESY spectra, easy identified when compared to the TOCSY spectrum. These NOEs are the basis for sequential
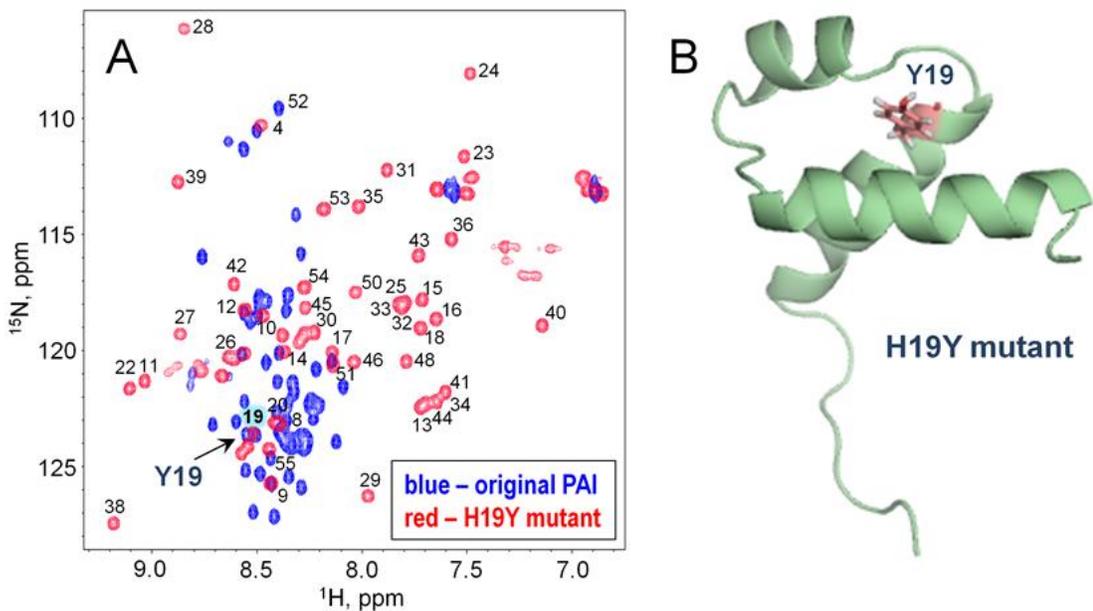

Figure 8. NMR data on the H19Y mutant. (A) The [15N-1H]-HSQC spectra of the PAI subdomain (blue) and H19Y mutant (red) recorded at identical experimental conditions, at 5oC in 25 mM sodium phosphate buffer at pH 5.0, no NaCl. Peak dispersion in H19Y spectrum clearly shows that the protein is folded. About 90% of NMR assignments is completed (shown). (B) Preliminary solution NMR H19Y structure in the cartoon representation.

assignments of amino acid residues to NMR signals. Such NMR spectra analysis

provided signal assignments labeled on the HSQC spectrum in *Figure 8*. Remaining

signals, which do not belong to neighbor amino acid residues, originate from resides

distant in sequence, but close in space, so called long-range NOEs, provide distance

information used for structural calculations and examples are shown as dashed circles

within Figure 9. The intensity of any NOE is inversely proportional to the sixth power of

the distance between interacting protons, $\sim 1/r^6$ (Poulsen, 2002). Therefore, in structure

calculation, we start with the most intense peaks in preliminary calculations to minimize

errors in the accidental selection of noise over data. Using these NOE data, we calculated

a preliminary structure of the PAI H19Y mutant show in *Figure 8*B. This structure needs

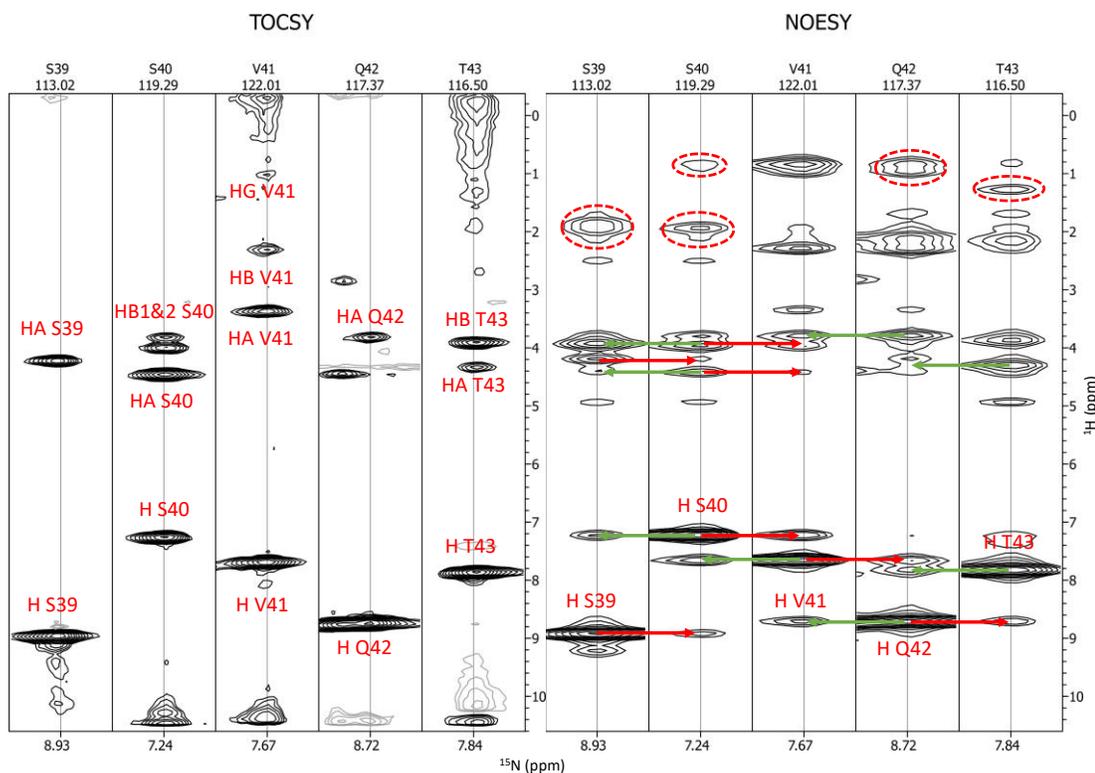*Figure 9. Example 15N-HSQC slices from TOCSY and NOESY spectra.*(Left Panel) A fragment of backbone-aligned 3D [1H,15N]-HSQC TOCSY strips labeled with side chain assignments. (Right Panel) Same fragment of backbone aligned 3D [1H,15N]-HSQC NOESY with neighboring spin systems indicated by left (green) or right (red) facing arrows. Example NOE signals are indicated by dashed circles (red).

further refinement, because we did not fully assign the NOESY spectrum. The full NMR

structural analysis will be completed using the ^{15}N, ^{13}C-isotopically labeled

K14RH19YK33A mutant to verify ambiguous assignments and generate a significant

amount of long-range side chain to side chain NOE signals needed to resolve fine

structural differences between the WT PAI subdomain and its H19Y mutant.

Observations of unfolding due to increasing temperature from 5°C to 45°C in 10-degree

increments are greatly diminished upon introduction of the H19Y mutation (*Figure 10*).

Peak broadening can be observed evenly across the spectrum caused by increasing

temperatures at both pH 5 and 7; overall the structure remains folded in contrast to the

WT PAI subdomain (*Figure 5*). Our NMR data, and preliminary structure, clearly show

that substituting His19 results in a folded PAI subdomain structure. This experimental

result corresponds to the predicted increase in stability by the ERIS server, and, thus, it is

likely that the addition of mutations to the other histidine residues would result in similar

stabilization. Data not shown provide evidence that DNA binding and subsequent

transposition has been observed to increase upon introduction of this H19Y mutation in
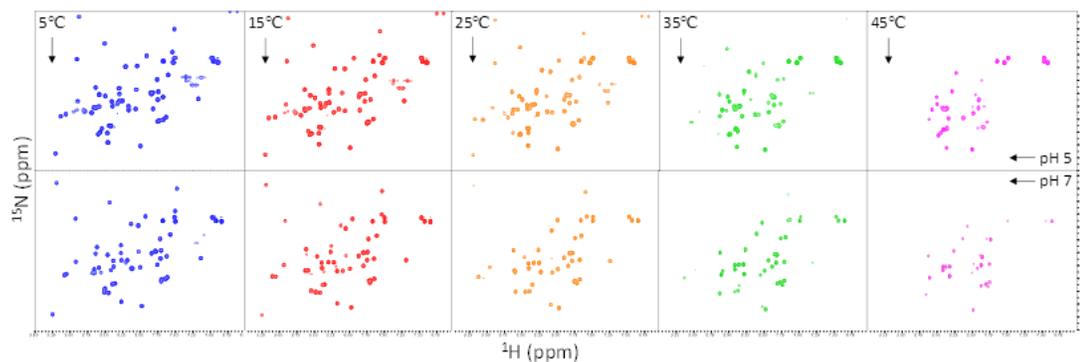

Figure 10. Increasing temperature on H19Y folding by NMR spectroscopy.2D[1H,15N]-HSQC NMR spectra of the K14RH19YK33A mutated PAI subdomain at pH 5 (top row) or pH 7 (bottom row) as a response to increasing temperature.

the conditions tested. We thus expect that subsequent utilization of the H19Y mutation in combination with the previous SB100X hyperactive mutation should result in more efficient human gene therapy trials.

Conclusion

The addition of the H19Y mutation, in addition to the K14RK33A hyperactive mutations show increased structural stability under a wide range of conditions as predicted computationally by the ERIS protein stability prediction server and experimentally verified by NMR spectroscopy. Preliminary data further show that the introduction of the H19Y mutation into the PAI subdomain results in greater affinity for the transposon sequence and increases the efficiency of transposition. We thus propose that increasing the structural stability of the PAI subdomain could lead to increased transposition activity of the SB system. Our data also provide insight into the molecular mechanism of SB transposon – we show that the preliminary DNA recognition subdomain of the SB transposase binds to the transposon DNA via the conformational selection mechanism (Vogt & Cera, 2013). As a next step, I propose to experimentally investigate other histidine mutations in the PAI subdomain and determine the transposition activity of these mutants.

References

Cameselle, J., Ribeiro, M., & Sillero, A. (2010). Derivation and Use of a Formula to Calculate the Net Charge of Acid-Base Compounds. Its Application to Amino Acids, Proteins and Nucleotides. *Biochemical Education*, 131-136.

Carpentier, C. E., Schreifels, J. M., Aronovich, E. L., Carlson, D. F., Hackett, P., & Nesmelova, I. V. (2014, February). NMR structural analysis of Sleeping Beauty transposase binding to DNA. *Protein Science, 23*, 23-33. doi:10.1002/pro.2386

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. *J. Mol. Biol., 318*(5), 1221–1235. Retrieved 6 17, 2017

Danaei, G., Finucane, M. M., Lu, Y., Singh, G. M., Cowan, M. J., Paciorek, C., . . . Ezzati, M. (2011). National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2·7 million participants. *Lancet*, 31-40.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, 277-293.

Gencer, A. G., Pelin, Z., Kucukali, C. I., Topcuoglu, O. B., & Yilmaz, N. (2011). Creutzfeld-Jakob disease. *Psychogeriatrics*, 119-124.

Geurts AM, Y. Y. (n.d.). Gene transfer into genomes of human cells by the sleeping beauty transposon system. *Mol. Ther., 8*(1), 108–117. Retrieved 6 17, 2017

Hackett, P. B., Largaespada, D. A., Switzer, K. C., & Cooper, L. J. (2013). Evaluating risks of insertional mutagenesis by DNA transposons in gene therapy. *Transl Res*, 265-283.

Hackett, P., & Zerull, D. (2011, May 9). *Generic diagram of Sleeping Beauty transposase-mediated cut-and-paste transposition.* Retrieved from https://commons.wikimedia.org/wiki/File:SBTS1.png

Hass, M. A., & Mulder, F. A. (2015). Contemporary NMR Studies of Protein Electrostatics. *Annual Review of Biophysics, 44*, 53-75.

Holtzer, M. E., & Holtzer, A. (1992). Alpha-helix to random coil transitions: determination of peptide concentration from the CD at the isodichroic point. *Biopolymers*, 4306-4312.

Hudecek, M., Izsvak, Z., Johnen, S., Renner, M., Thumann, G., & Ivics, Z. (2017). Going non-viral: the Sleeping Beauty transposon system breaks on through to the clinical side. *Mol Biol*, 355-380.

Ivics Z, H. P. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell, 91*(4), 501–510.

Ivics Z, I. Z. (1996). Identification of functional domains and evolution of Tc1-like transposable elements. *Proc. Natl. Acad. Sci. U.S.A., 93*(10), 5008–5013.

Izsvak, Z., Hackett, P. B., Cooper, L. J., & Ivics, Z. (2010). Translating Sleeping Beauty transposition into cellular therapies: Victories and challenges. *BioEssays*, 756-767.

Izsvak, 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. *Biol Chem*, 34581-34588.

Izsvak, 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. *Biological Chemistry*, 34581-34588.

Johnson, B. A., & Blevins, R. A. (1994). NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, 603-614.

Judd, S. J. (2013). 10. Blood clotting deficiency disorders. In *Genetic disorders sourcebook.* Detroit: Omnigraphics.

Judd, S. J. (2013). 11.3 Sickle cell disease. In *Genetic disorders sourcebook.* Detroit: Omnigraphics.

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*. doi:10.1371/journal.pone.0112114

Mátés L, C. M.-S.-K. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat. Genet., 41*(6), 753–761. doi:10.1038/ng.343

Mort, M., Ivanov, D., Cooper, D. N., & Chuzhanova, N. A. (2008). A meta-analysis of nonsense mutations causing human genetic disease. *Human Mutations, 29*(8), 1037-1047.

Mrabet, N. T., Van den Broeck, A., Van den brande, I., Stanssens, P., Laroche, Y., Lambeir, A. M., . . . van Tilbeurgh, H. (1992). Arginine residues as stabilizing elements in proteins. *Biochemistry*, 2239-2253.

Nesmelova IV, H. P. (n.d.). DDE transposases: Structural similarity and diversity. *Advanced Drug Delivery Reviews, 62*(12), 1187–95. Retrieved 6 20, 2017

Neuhaus, D., & Williamson, M. P. (2000). *The Nuclear Overhauser Effect in Structural and Conformational Analysis* (2nd ed.). New York: WILEY-VCH.

Poulsen, F. M. (2002). *A brief introduction to NMR spectroscopy of proteins.*

RIchardson, J. M., Colloms, S. D., Finnegan, D. J., & Walkinshaw, M. D. (2009). Molecular architecture of the Mos1 paired-end complex: the structural basis of DNA transposition in a eukaryote. *Cell*, 1096-1108.

Sandberg, W. S., Schlunk, P. M., Zabin, H. B., & Terwilliger, T. C. (1995). Relationship between in Vivo Activity and in Vitro Measures of Function and Stability of a Protein. *Biochemistry*, 11970-11978.

Schwieters, C. D., Kuszewski, J. J., Tjandra, N., & Clore, G. M. (2003). The Xplor-NIH NMR molecular structure determination package. *Magn Reson, 160*, 65-73.

Shen, Y., Delaglio, F., Cornilescu, G., & Bax, A. (2009). TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Biomol NMR*, 213-291.

Singh, H., Huls, H., & Cooper, L. J. (2014). A new approach to gene therapy using Sleeping Beauty to genetically modify clinical-grade T cells to target CD19. *Immunological Reviews*, pp. 181-190.

Smith, C., Gore, A., Yan, W., Abalde-Atristain, L., Li, Z., He, C., . . . Ye, Z. (2014). Whole-genome sequencing analysis reveals high specifi city of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell*, 12-13.

Sokalingam, S., Madan, B., Raghunathan, G., & Lee, S. G. (2013). In silico study on the effect of surface lysines and arginines on the electrostatic interactions and protein stability. *Biotechnol. Bioproc.*, 18-26.

Sokalingam, S., Raghunathan, G., Soundrarajan, N., & Lee, S. G. (2012). A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS One*.

Suzuki, K., Yu, C., Qu, J., Li, M., Yao, X., Yuan, T., . . . Izpisua, J. C. (2014). Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specifi c induced pluripotent stem cell clones. *Cell Stem Cell*, 31-36.

Terry, C. A., Fernandez, M.-J., Gude, L., Lorente, A., & Grant, K. B. (2011). Physiologically relevant concentrations of NaCl and KCl increase DNA photocleavage by an N-substituted 9-aminomethylanthracene dye. *Biochemistry, 50*, 10375-10389.

Tomoshige, K., Matsumoto, K., Tsuchiya, T., Oikawa, M., Miyazaki, T., Yamasaki, N., . . . Nagayasu, T. (2015). Germline mutations causing familial lung cancer. *Human Genetics, 60*(10), 597-603.

van Pouderoyen, G., Ketting, R. F., Perrakis, A., Plasterk, R. H., & Sixma, T. K. (1997). Crystal structure of the specific DNA-binding domain of Tc3 tranposase of C.elegans in complex with transposon DNA. *EMBO*, 6044-6054.

Vogt, A. D., & Cera, E. D. (2013). Conformational Selection Is a Dominant Mechanism of Ligand Binding. *Biochemistry*, 5723-5729.

Voigt, F., Wiedemann, L., Zuliani, C., Querques, I., Sebe, A., Mates, L., . . . Barabas, O. (2016). Sleeping Beauty transposase structure allows rational design of hyperactive variants for genetic engineering. *Nat. Commun.*

Wang, Y., Pryputniewicz-Dobrinska, D., Nagy, E. E., Kaufman, C. D., Singh, M., Yant, S., . . . Izsvak, Z. (2017). Regulated complex assembly safeguards the fidelity of Sleeping Beauty transposition. *Nucleic Acids Research*, 311-326.

Wang, Y., Sarkar, M., Smith, A. E., Krois, A. S., & Pielak, G. J. (2012). Macromolecular Crowding and Protein Stability. *J. Am. Chem. Soc.*, 16614-16618.

Watkins, S., van Pouderoyen, G., & Sixma, T. K. (2004). Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. *Nucleic Acids Res.*, 4306-4312.

Whitmore, L., & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*, 668-673.

Whitmore, L., & Wallace, B. A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, 392-400.

Yamamoto, T. (2015). *Targeted genome editing using site-specific nucleases : ZFNs, TALENs, and the CRISPR/Cas9 system.* Tokyo: Springer.

Yant, S. R., Park, J., Huang, Y., Mikkelsen, J. G., & Kay, M. A. (2004). Mutational analysis of the N-terminal DNA-binding domain of sleeping beauty transposase: critical residues for DNA binding and hyperactivity in mammalian cells. *Mol. Cell Biol.*, 9239-9247.

Yin, S., Ding, F., & Dokholyan, N. V. (2007). Eris: an automated estimator of protein stability. *Nat Methods, 4*, 466-467.

Yuan, Y.-W., & Wessler, S. R. (2011). The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc Natl Acad Sci U S A*, 7884-1889.