USING HEAT EFFECTS ON COIL HYDRODYNAMIC SIZE TO REVEAL THE NATURE AND ENERGETICS OF DENATURED STATE CONFORMATIONAL BIAS

by

Elisia Ariana Paiz, B.S.A

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a Major in Biochemistry
December 2019

Committee Members:

Steven T. Whitten, Chair

Karen A. Lewis

Sean M. Kerwin

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

DEDICATION

Para mis familias Paiz, Macias, Cantu, y Islas!

To my parents, Gabriela and Oscar Paiz,
and my very best friends Adrian, Sevy, and Angela.

For my girls,
Storm and Shade.

ACKNOWLEDGEMENTS

I would like to thank the South Texas Doctoral Bridge Program. Dr. Blake, thank you for your never-ending support and hugs! You were one of the first people in this new chapter of my life that made me feel comfortable, like I belonged, and like I could really do this. Dr. Oyajobi, you have been such a great source of encouragement, support, and drive in these last two years. I can never say thank you enough. I would also like to thank Dr. Raquel Salinas for her never-ending support of me. Thank you for everything you have done for me! I'm so lucky to have met you and look up to you. Lastly, thank you to Dr. Carolyn Chang. In the short time I've known you, you've already provided me with so much insight and advice. I know I would not be where I am today without the support and mentorship of everyone in our group. Thank you for believing in me and providing me the opportunity to find the direction I needed.

A very big thank you to my research advisor, Dr.Whitten, for his guidance and the opportunities he's given me to become a scientist. His mentorship and philosophies allowed me many chances to learn about myself and the scientific process. I'm also lucky to have collaborated with such incredible and hard-working scientists. I have to thank my lab colleagues, past and present, Sarah, George, Allie, and Lance for their never-ending source of ideas, intelligence, and perseverance. I will forever wish I could stay a member of the Whitten lab!

Behind the scenes of this all is my one-and-only support system that has helped me succeed by keeping me mentally and emotionally strong. To you: Mom, Dad, Sevy,

Angela, Adrian, Storm, and Shade, you made this all so much easier just by loving me. It's hard to stay down and frustrated when everyone around you is lifting you up.

Finally, I would like to thank my committee members, Dr. Lewis and Dr. Kerwin for their inquisitive minds and feedback through my whole project. I couldn't have asked for better scientists to see me through the whole process and help me develop the way I think about my work.

Thank you all for, at some point, seeing in me what I sometimes failed to see in myself. This change in mentality and perception has been really important to me and I know I can only continue to improve. Sincerely, thank you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

Abbreviation	Description
A_{280}	Absorbance at 280 nm
Ala^-	p53(1-93) with 12 alanine to glycine subsitutions
Alb	Chicken egg albumin
ADK	Adenylate kinase
Amp^r	Gene that codes for beta-lactamase (ampicillin-resistance)
AP	21-residue Alanine peptide
BD	Blue dextran
BD-DNP-Asp	Blue dextran-DNP-Aspartate
BME	β-mercaptoethanol
CA	Bovine carbonic anhydrase
CD	Circular dichroism
DE3	Designation given to cells carrying an IPTG-inducible chromosomal copy of T7 RNA polymerase
DEAE	Diethylaminoethyl
DLS	Dynamic light scattering
DSE	Denatured state ensemble
DNP-Asp	DNP-Aspartate
Δ131Δ	Unfolded variant of *S. aureus* nuclease that contains 18 deletions

IDP	Intrinsically disordered protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K_D	Partition coefficient
lacI	Gene that codes for the lactose repressor protein
LB	Lysogeny Broth
LB + Amp	Lysogeny Broth containing 100 mg/mL ampicillin
LP	Low Pressure
MD	Molecular dynamics
MRE	Molar residue ellipticity
MW	Molecular weight
Myo	Horse heart myoglobin
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
OD_{600}	Optical density (absorbance) at 600 nm
ORD	Optical rotatory dispersion
p53(1-93)	The intrinsically disordered N-terminal (residues 1-93) region of the human tumor suppressor protein p53
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline (10 mM sodium phosphate, 100 mM NaCl, pH 7.0)
PDB	Protein Data Bank

pdTP	Deoxythymidine 3',5- biophosphate
pLysS	Plasmid containing genes for chloramphenicol-resistance and T7 lysozyme
PGA	Poly-Glutamic Acid
PPII	Polyproline II
PXP	Short peptide form where X indicates any amino acid and P indicates 3 consecutive proline residues
R_h	Hydrodynamic radius
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SOC	Super optimal broth with catabolite repression
SOP	Standard Operating Procedures
ssDNA	Single stranded DNA
TDP	Thymidine diphosphate
UV	Ultra-violet
V_0	Void volume
V_e	Elution volume
V_t	Total volume
WT	Wildtype
XAO	Alanine heptamer
XRC	X-ray crystallography

ABSTRACT

Conformational equilibria in the protein denatured state have key roles regulating folding, stability, and function. The extent of conformational bias in the denatured state under folding conditions, however, has thus far proven elusive to quantify, particularly with regard to its sequence dependence and energetic character. To better understand the structural preferences of the denatured state, we analyzed both the sequence dependence to the mean hydrodynamic size of intrinsically disordered proteins (IDPs) and the impact of heat on the coil dimensions, demonstrating that the sequence dependence and thermodynamic energies associated with intrinsic biases for the α and polyproline II (PPII) backbone conformations can be obtained. This result was determined by using a model of the denatured state whereby the position-specific conformational preferences of IDPs were very locally determined. To test these results, two sets of experiments were performed. First, experiments were designed to determine if this method for measuring conformational bias is limited to IDP sequences, which are known to be depleted, relative to foldable protein sequences, in hydrophobic content. By designing unfolded mutant proteins from foldable sequences, it was shown that hydrophobic-rich sequences likewise experience hydrodynamic size changes that relate to heat effects on backbone structure. Second, we designed experiments to test whether the structural features we observe are indeed locally determined. By scrambling the mutant sequences multiple times, it was determined that the structural features of these proteins were, surprisingly, independent of the arrangement of amino acids, “intrinsic” to the composition, and very locally determined.

1. INTRODUCTION

A central principle of structural biology states that stable, folded structure is necessary for biological activity in proteins.[1] This view persists from the results of several key experiments and observations. In 1936, Pauling demonstrated distinct properties of native proteins, such as their ability to fold into a unique and defined states and crystallize. He further observed that pepsin, a protease found in the stomach, was unable to be crystallized and lost its activity when it was denatured, or not in its native folded structure.[2] Anfinsen observed that ribonuclease regained its structure and function upon renaturation from an inactive denatured state.[3] In addition to the many experiments that demonstrate a direct relationship between function and native structure, there are a multitude of folded protein structures available in the Protein Data Bank (PDB), such as those for the proteins myoglobin[4] and hemoglobin[5] and the enzymes lysozyme[6, 7]and carbonic anhydrase[8, 9], solved by x-ray crystallography (XRC). These early discoveries supported the prerequisite of native structure for function and there are numerous proteins for which this dogma describes them adequately.[2-10]

Challenging the notion of a strict relationship linking protein structure to biological activity were observations that some proteins or protein regions are configurationally mobile and structurally adaptive under physiological conditions.[10, 11] This suggests that fluctuations in protein structure may have a role in function. Consistent with that idea, proteins have been shown to undergo both structural and dynamic changes to regulate activity. Saavedra et al. demonstrated that local unfolding in the LID and AMPbd domains of *E. coli* adenylate kinase (ADK) regulates its activity and provides a mechanism for cold-adaption of ADK function in psychrophilic organisms.[12] Proteins

have also been shown to use conformational fluctuations to regulate turnover rates and transport across membranes. The proteasome is observed to have accelerated proteolysis when ubiquitinated substrates contained an unstructured region.[13] Proteins needing to translocate across membranes, such as *E.coli* proteins being exported to the periplasm, must be in an unfolded state[14] and it has been shown that mostly folded mutants of maltose-binding protein or ribose-binding protein kinetically slow down this process or end up trapped in the cytoplasm.[14] Taken together, these events demonstrate that it is common for the denatured state to be accessed frequently and that folding isn't always a single-occurring event in a protein's lifetime.

While the preceding examples show that conformational fluctuations associated with protein structure are common under biological conditions, there is a class of proteins that exhibit dramatic structural heterogeneity in their native states. These proteins, called intrinsically disordered proteins (IDPs), or large regions of intrinsic disorder (IDRs) are found in about 10-35% of prokaryotic proteins and 15-45% of eukaryotic proteins, depending on the method used for the estimation.[15, 16] Despite the lack of stable structure, IDPs like p53 and BRCA1, are biologically active via their disordered regions and are shown to be highly involved in cell signaling and regulation.[17] Compared to proteins that adopt stable structures, IDPs are characterized by an enrichment of amino acids that have a high flexibility index, high net charge, and low hydrophobicity, characteristics that all promote structural disorder.[18] IDPs are another example of biological activity being correlated to structural fluctuations and understanding the relationships linking sequence, structure, and function in disordered states is important for understanding IDP biology.

The examples above have demonstrated the prevalence of protein conformational fluctuations in various biological functions. However, conformational fluctuations also play a key role is the process by which a protein adopts a native, folded structure from its unfolded state. While the protein folding reaction has been studied for many years[2], key mechanistic details remain elusive due to an incomplete understanding of the process. In order to understand protein folding, it is important to detail both the final and starting states, as non-native states are the starting points for both folding and refolding in proteins and these changes are what define the reaction.[19] As previously mentioned, the final states of folded proteins are characterizable through x-ray crystallography as well as by nuclear magnetic resonance (NMR) but unfolded states are not as easy to study. Due to the dynamic nature of the unfolded state, NMR and XRC are not suitable methods for full analysis. Our group has established approaches to quantify denatured structure by measuring the hydrodynamic radius (R_h), since hydrodynamic size has been shown to be sensitive to temperature changes in many IDPs[20-22], and relating it to changes in protein backbone structure. Additionally, the hydrodynamic techniques we utilize, such as size exclusion chromatography (SEC), can separate out aggregates during analysis. This is important because our model systems (i.e., unfolded protein) often have solubility issues and are prone to aggregation, making them difficult to study without the presence of co-solvent.[23]

Because of the early difficulty of studying these conformationally flexible states, unfolded proteins have traditionally been classified as random coils[24], a term that suggests that the rotatable bonds along the protein backbone, referred to as the phi(Φ) and

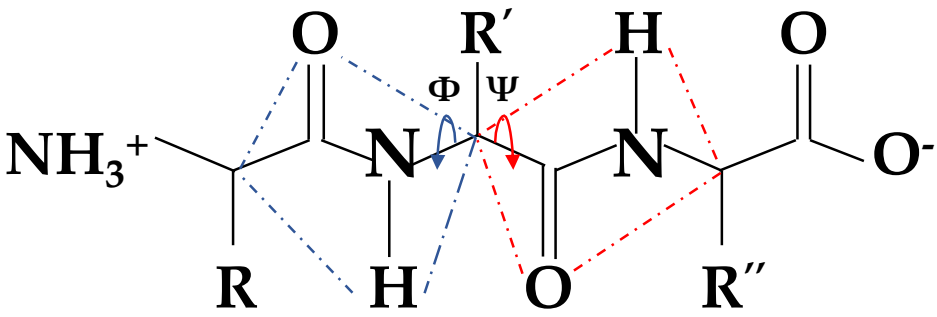

Figure 1.1. <u>Dihedral Angles.</u> Protein structure is commonly defined by the backbone dihedral angles on either side of each Cα atom, referred to as the phi(Φ) and psi(Ψ) rotation angles. The peptide bond connecting each residue has partial double-bond character and its rotation is limited to the *cis* and *trans* configurations. Pictured is an alanine tripeptide.

psi(Ψ) angles for each amino acid residue (Figure 1.1), are limited only by steric considerations.[24, 25] The idea of a random coil starting state has been supported for more than 50 years[24] by Tanford's demonstration of guanidium hydrochloride (GuHCl) denatured proteins conforming to random coil models based on Flory's Isolated Pair Hypothesis.[23] This hypothesis states that each residue in a sequence is insensitive to the chemical nature of the residues around it.[26] The generalization of a random coil unfolded state supported by these early experiments has been applied not only to proteins in the presence of denaturants but also to those in the presence of water, despite the observation that secondary structures are stable in water.[27] The past few decades have seen a change in the characterization of the denatured state by showing that, experimentally and theoretically, structure can be transiently present in intermediate states of folding.[27]

1.1. Establishing Conformational Bias

Levinthal demonstrated in 1969 that due to the large conformational search space afforded to a polypeptide chain, defined by all sterically possible configurations, it is not possible for proteins to fold correctly by a random search for a local energy minimum.[28]

If a protein were to take on its native configuration by sampling all possible combinations, this process would take much longer than the milliseconds or microseconds that is often observed. Instead, it has been proposed that protein folding begins from a mixture of unique intermediate starting states that funnel into a native state.[23] This "energy landscape" theory of protein folding suggest that rather than beginning folding at a high energy state with no structural preferences, it is more likely the case that the unfolded state is at a lower energy state, already conformationally biased for folding to the local energy minimum.[29]

These structural preferences associated with denatured protein are often described in terms of preferences for specific combinations of Φ and Ψ angles. Visually, this is demonstrated by a Ramachandran plot which contains all backbone angles (Φ,Ψ) that are theoretically possible and conformationally accessible (Figure 1.2). For example, the position (Φ,Ψ) = (0º, 0º) appears on the Ramachandran map in Figure 1.2, but this is in the disallowed region for the protein this plot represents. In contrast, (Φ,Ψ) = (-90º, +90º) is conformationally allowed in this example. When an unfolded state adopts structures with preferences for specific Φ and Ψ pairs at the expense of other conformationally accessible regions of the Ramachandran plot, we say that the protein shows a conformational bias.

Structural preference in the denatured state was first demonstrated by Tiffany and Krimm when they studied short poly-proline and poly-lysine peptides using circular dichroism (CD) and optical rotatory dispersion (ORD). Polypeptides consisting of only a few residues have insufficient length for forming compact, globular structures. Though the peptides were unfolded, Tiffany and Krimm observed strong preferences for a

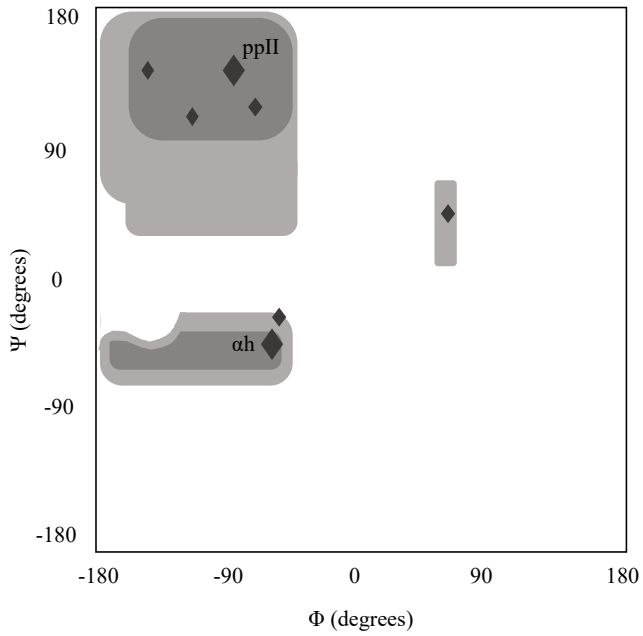

Figure 1.2. <u>Ramachandran Plot.</u> Ramachandran plot indicates combinations of phi(Φ) and psi(Ψ) backbone angles. This plot shows all possible combinations of Φ and Ψ. Shaded regions indicate conformationally accessible regions. Combinations of Φ and Ψ associated with specific secondary structures are indicated with grey diamonds.

structure called polyproline II (PPII).[30, 31] Polyproline II (PPII) in a series of residues results in an extended left-handed helix with (Φ,Ψ) ~ (-75º, +145º) at each position. Poly-glutamic acid (PGA) was also observed to transition from α-helix at low pH to PPII at neutral pH[31], indicating that transitions between one region of the Ramachandran plot to other regions can occur for some sequences due to minor solution perturbations. These discoveries support the theory that unfolded proteins are conformationally biased rather than completely sterically dictated. From these finding, Tiffany and Krimm hypothesized, to much resistance, that protein unfolded states might also contain PPII as a way to remain solvated[30], since the PPII configuration places the backbone amide and carbonyl oxygen groups in favorable positions for interactions with water. By survey of the PDB, PPII is only weakly represented in folded structures, but is found in homopolymers of

proline[30] as well as in collagen.[32] However, recent survey of the protein coil library, a denatured state model built from the segments of irregular structure in the PDB[33-35], shows that the PPII basin is one of the most populated regions of the Ramachandran map.[33, 34]

Structural preference in the denatured state is also shown by a number of spectroscopic studies on short oligomers. These peptides make adequate systems of study due to their inability to form long-range interactions that are typically found in folded structures. The first of these experiments utilized short alanine oligomers because of the simplicity of the alanine side chain.[36] NMR of XAO, an alanine heptamer flanked by basic side chains, showed that it was predominantly PPII at each of its 7 alanine positions.[37] XAO was also capable of sampling β-like structures depending on the temperature.[37] Another study examined a single alanine residue flanked by 2 glycine residues on either side (GGAGG) and was also shown to prefer PPII and shift to β structures at high temperatures.[36] These finding are replicated in alanine tripeptides[38], tetrapeptides[39, 40], and octapeptides[39] all showing a mixture of PPII and β-like structures that begin forming with a minimum of 3 residues. When XAO was examined by Asher et. al, it was determined that its ultra-violet (UV) Raman spectra was similar to a previously measured 21-residue Ala-peptide (AP).[32] AP is predominantly α-helix but was shown to be predominantly PPII as temperature increases[32], demonstrating that AP, similar to XAO, has temperature dependent conformational preferences.[41]

Conformational preferences and environmentally induced conformational changes in these short alanine peptides[36-41] prompted investigation of the preferences in the other amino acid types. The first of these studies used a short peptide form referred to as PXP,

where X indicates the residue being tested bound by two sets of three proline residues, P. For example, PQP would indicate a peptide structure of $(Pro)_3$-Gln-$(Pro)_3$. Rucker et. al used CD spectroscopy to estimate PPII structure in all residues but tyrosine and tryptophan, excluded due to the incompatibility of their aromatic side chains with measurements by CD.[42] Here, peptides with charged residues, except for histidine, had high propensities for PPII that were mostly insensitive to changes in pH.[42] The non-polar residues alanine and glycine had PPII propensities that were larger than residues with bulky side chains such as leucine, isoleucine, or valine. It was hypothesized that these differences in PPII propensities for the non-polar residues was related to backbone solvation effects. Alanine and glycine are likely to adopt a PPII conformation to maintain backbone solvation whereas a branched or bulky side chains, like those of isoleucine and phenylalanine, would partially bury the backbone.[42]

Shi et al. extended their study of the GGAGG peptide[36] to the other residues, utilizing a similar GGXGG model where X was the residue being tested examined by NMR and CD spectroscopy. They found that, qualitatively similar to Rucker et al., all residues had a different PPII propensity and, in agreement with their previous work, peptides transition from PPII to a β-like structure with higher temperatures.[43]

A third scale measured PPII propensity by using a peptide host-guest system. This system analyzed the binding of the son of sevenless (SoS) peptide to the SH3 domain of *C. elegans* Sem-5 protein. Because SoS is a short peptide, when it is unbound it is unfolded. But more importantly, when SoS is bound to the SH3 domain, it is in the PPII conformation.[44] Here, substitutions were made to solvent exposed prolines on SoS that did not affect contact with the binding pocket. Therefore, the changes in

conformation were reflective of the unbound state rather than the bound state.[45] By measuring and ranking the differences in free energy of binding ($\Delta\Delta G_{binding}$) in single-amino-acid substitution mutants relative to wildtype it is possible to gain insight to position-specific PPII bias. Once again, the rank orders were different than the other previously described rankings. Elam et al. concludes that there is a general consensus of amino acids that are high in PPII propensity (proline, lysine, glutamine, and glutamic acid) and low in PPII propensity (histidine, tryptophan, tyrosine, and phenylalanine) with the other amino acids falling in between. The rank of relative PPII propensities are quantitatively different between the three studies, possibly owing to the different host models used. This difference suggests that PPII propensity is not only intrinsic to amino acid type, but also sequence and context dependent.[43]

1.2. Conformational Bias in Unfolded Protein

The above studies show rather convincingly that unfolded peptides in water are conformationally biased. The question then is whether this is also the case for unfolded proteins in water. Several observations by our lab support that this is the case. Recently, we observed that changes in the sequence composition of the disordered N-terminal region of p53 (p53 (1-93)) resulted in distinguishable differences in hydrodynamic size at room temperature.[46] Substitution to proline at 15 sites in p53(1-93) produced the mutant with the largest mean R_h of 33.6±1.0 Å, whereas substitutions to glycine in the same spots yielded the smallest mean R_h of 30.8±1.4 Å, with mean R_h of WT p53(1-93) being 32.8±0.4 Å. We also found that the changes in the mean R_h resulting from compositional changes in the protein p53(1-93) yield amino acid preferences for PPII that recapitulate intrinsic PPII propensities measured calorimetrically in unfolded peptides.[46] This

demonstrates that sequence-dependent conformational biases are present in the disordered state.

Temperature changes have also been found to induce large shifts in hydrodynamic size for disordered proteins in water[21], that can exceed the changes in size associated with heat denaturation of folded protein of the same chain length.[47] This was recently observed with "Retro nuclease" an intrinsically disordered protein that was obtained by reversing the primary sequence of the foldable protein, *Staphylococcal aureus* nuclease (SN), "WT nuclease". By utilizing this strategy, the identical composition of L-amino acids and pattern of side chains is maintained despite the reversed sequence directionality. Importantly, this method allows us the ability to study the denatured state of a protein with a foldable sequence composition to investigate how sequence patterns and composition influence the conformational dynamics of the denatured state ensemble (DSE).

What was observed was that Retro nuclease did not appear to adopt a compact denatured state at 25 °C, which was unexpected since hydrophobic residues are present at levels appropriate for a packed core in the WT.[48] In fact, Retro nuclease's hydrodynamic size, measured by hydrodynamic radius (R_h), in 10 mM sodium phosphate, 100 mM sodium chloride, pH 7 is highly expanded at 25 °C, not only larger than folded WT nuclease, but also approximately 7 Å larger than heat-denatured WT nuclease (Figure 1.3, A). Additionally, while WT nuclease adopts a stable native structure consisting of three α helices and a five-stranded, barrel-shaped β sheet[48], sequence reversal resulted in a disordered state with a temperature-dependent, reversible sigmoid-shaped transition of Retro nuclease's hydrodynamic size. This structural transition was seen by both dynamic

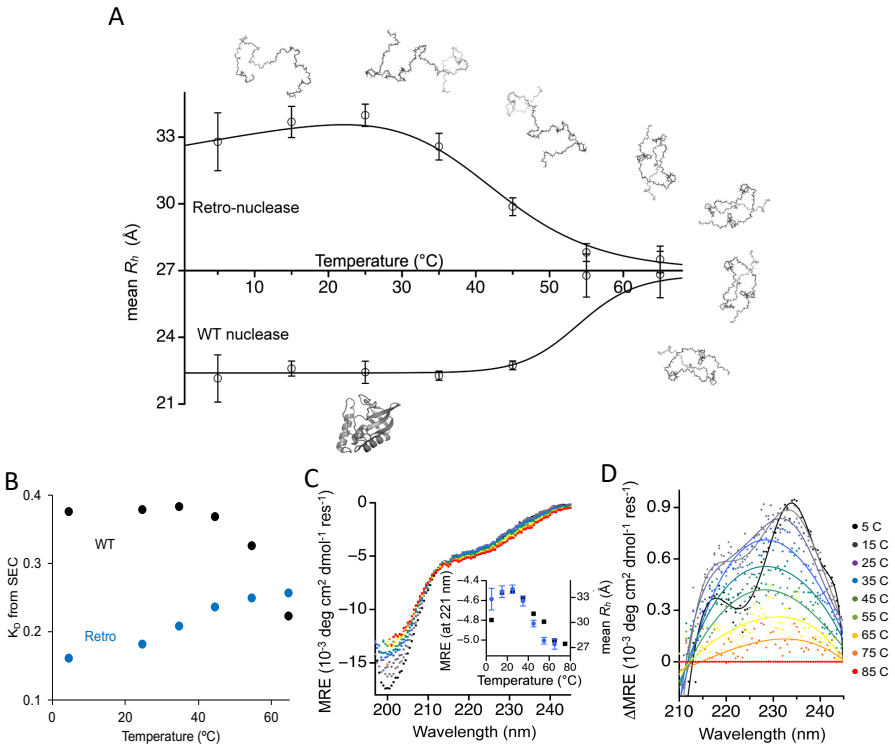

Figure 1.3. **<u>Temperature-induced structure changes in a disordered protein.</u>** A) DLS and B) SEC show hydrodynamic size is heat sensitive for both Retro nuclease and the unaltered wild type (WT) protein. C) changes in the circular dichroism spectrum are minimal D) Difference CD spectra, relative to the 85°C spectrum, show a local CD maximum at 225 nm and a local minimum at 222 nm. Adapted from English et al.[47]

light scattering (Figure 1.3,A; DLS) and size exclusion chromatography methods (Figure 1.3, B; SEC).

To examine the nature of this structural transition, several methods were employed with the greatest indicator coming from the circular dichroism (CD) spectrum. CD spectroscopy of Retro nuclease at 25°C is similar to spectra reported for IDPs[49], with molar residue ellipticity (MRE) at 221 nm close to zero (Figure 1.3, C). The Retro nuclease spectrum changed minimally with changes in temperature and was independent of the protein concentration.[47] However, close inspection of CD results by monitoring difference spectra relative to the 85°C spectrum (Figure 1.3, D) shows a local CD

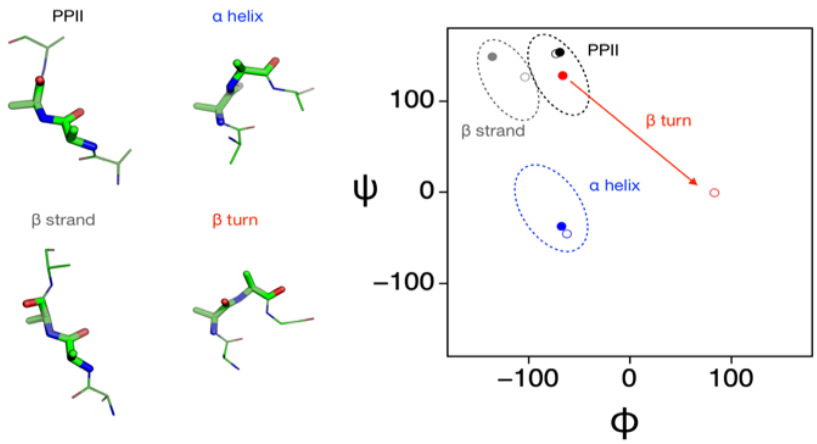

Figure 1.4. Backbone Conformations in Peptides. Computer simulated 4-residue peptides with backbone conformations that are classified as PPII, α helix, β strand, and β turn (type II), which are the common secondary structure types. The central residues in each simulated structure are highlighted and the ϕ and ψ angles are given in the Ramachandran plot. In the plot, residue 2 is shown by filled circles, and residue 3 with open circles.

maximum at 225 nm, normally seen in proteins with PPII structure[50], that increases in intensity with decreasing temperature below 85°C. This temperature-dependent bias for PPII possibly explains the increase in Retro nuclease size as it cools from high temperatures (Figure 1.3, A). Because PPII is an extended structure (Figure 1.4), an increasing level of PPII conformational sampling causes concomitant increases in the mean hydrodynamic size of simulated disordered ensembles.[22]

Temperatures < 25 °C caused a local CD minimum at around 222 nm to form (Figure 1.3, D) that also increases in intensity with cooling. Local CD minima at around 222 and 208 nm are normally found in proteins with α helix[51], suggesting that the leveling off of Retro nuclease mean R_h at temperatures below 25 ºC is possibly from a second temperature-dependent bias, this one for α backbone conformations that offsets the effects of PPII.[47] Ensemble sampling of α conformations cause transient turn-like motifs to form (Figure 1.4) that decrease the mean hydrodynamic size in simulated ensembles.[20, 46] Importantly, while chemically denatured proteins adopt macromolecular

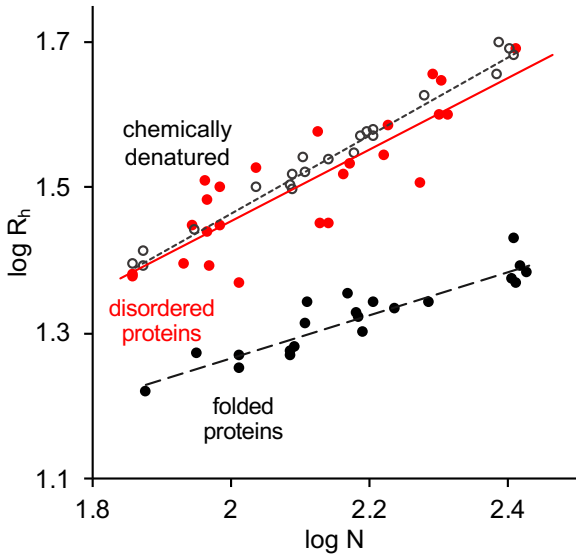

Figure 1.5. <u>Hydrodynamic size and chain length.</u> Note the scatter in the IDP trend (blue), indicating sequence effects on the hydrodynamic radius, R_h, not observed in the trend for chemically denatured protein. R_h, in Å, and length in residue number, N.

sizes that depend weakly on sequence details other than chain length, IDPs exhibit strong sequence-dependent influences on structural size (Figure 1.5). Computer simulations show that steric effects on disordered structure alone cannot account for the hydrodynamic size dependence on sequence observed in IDPs.[20]

1.3. Project Goals

Retro nuclease exhibited an anomalous temperature-induced structural behavior which was characterized by several key observations and implications.[47] First, a reversible, sigmoid-shaped structural transition was observed. In folded proteins, this is normally indicative of a cooperative, two-state transition but this was not the case for Retro nuclease. Second, a hydrophobic collapse does not appear to occur in Retro nuclease despite the presence of the necessary residues. This suggests that this system could potentially be used to investigate the nature of the sequence properties that promote the collapse of a protein chain. Finally, it was observed that changes in the CD spectrum

that monitor backbone structure scaled with temperature-dependent changes in R_h. This suggests that small changes in backbone conformational bias could potentially produce large changes in the mean hydrodynamic size. This observation is important because, if true, it implies that this relationship could be used to measure the thermodynamics of conformational bias in denatured proteins. For example, we recently designed an experimental system that, when combined with computer simulation of the denatured state ensemble, provides access to the sequence dependence and thermodynamic energies associated with protein structure in the denatured state.[46] Changes in R_h from sequence substitutions report on the differences in intrinsic conformational biases of the amino acid types involved in the substitution. This allows biases to be quantitatively measured[46] and temperature effects on R_h to yield the enthalpy and entropy contributions.[47]

Our recent results from examining p53(1-93) mutants and Retro nuclease suggest the DSE is both sequence and temperature dependent. However, the temperature dependence to mean hydrodynamic size of p53(1-93) is discernably different than the sigmoid-shaped transition observed in Retro nuclease, supporting that while both proteins are natively disordered, their difference in composition and organization produce two distinct types of denatured state ensemble structures. The first, observed in Retro nuclease, we refer to as intrinsic as it is independent of sequence organization and only dependent on the composition. The other we refer to as collective as it depends on both the composition and a specific sequence organization. We hypothesize that the role of the denatured state in folding and function can be learned from mechanistic studies on the collective properties of unfolded proteins in water. These two types of denatured state ensemble structures are discussed in further detail in Chapter 4.

Retro nuclease has been the only protein tested in this manner and thus the only protein that has been observed to exhibit this intrinsic DSE behavior. However, if we are interpreting the structural transition of Retro nuclease in terms of intrinsic conformational propensities associated with amino acid types, then this structural behavior should be general and thus present in other proteins. A primary objective of this project, presented in Chapter 3, was to structurally characterize the reverse sequence of a second protein, the B domain of *Staphylococcal aureus* virulence factor protein A, referred to as "Retro B domain." By doing this, we establish whether the observed behavior in Retro nuclease is special to the nuclease sequence or a more general observation of denatured proteins.

Further, at the present time we have no way of knowing if the behavior we observe is actually the results of conformational propensities that are indeed intrinsic. As a result, secondary to examining the general applicability of this phenomenon is the motivation to test if the structural preferences in the denatured state we observe are the result of formative interactions among neighboring elements (i.e., locally determined). To test this property, several variants of *S. aureus* nuclease that maintain the same sequence composition but have a scrambled sequence organization were structurally characterized as a lab collaborative effort and compared to Retro nuclease. These results are described in Chapter 4.

To pursue these unanswered questions, this project will have two specific aims: (1) structurally characterize Retro B domain to determine if the results from Retro nuclease are true for other sequences or unique to the nuclease sequence and (2) structurally characterize scrambled sequences of nuclease to see if sigmoidicity is maintained or disrupted as a means to determine if Retro nuclease is truly disordered or if

there is stable structure we aren't detecting.

This work is then briefly compared to structural data of Δ131Δ, an unfolded variant of *S. aureus* nuclease. Δ131Δ and WT nuclease differ by 18 deletions that result in a natively unfolded variant. However, NMR of Δ131Δ shows residual structure is present in the form of transient α-helix formation by residues corresponding to WT nuclease helix 1 and 2.[52] This protein, with its more global range interactions, provides the opportunity for comparison to the proteins we hypothesize are governed by very locally determined features. The Δ131Δ data was collected by Lance English and Allie Bebo and is included for a discussion about the different types of denatured state ensemble structures.

2. MATERIALS AND METHODS

2.1. Materials

Water used for experiments, reagents, and growth media was filtered and deionized using an Elga LabWater Purelab Classic water purification system (High Wycombe, UK). All chemicals and reagents were ACS grade or higher. Culture tubes used were VWR disposable 16×100-mm borosilicate. All dialysis was done using Spectrum Labs Spectra/Por tubing with a 12-14 kDa molecular weight cut-off (Rancho Dominguez, CA). All materials and reagents were weighed using an OHAUS Pioneer Analytical and Precision balance (Parsippany, NJ). A Thermo Fisher Orion Star A111 Benchtop pH Meter was used to measure the pH of all solutions (Waltham, MA). Degassing and vacuum filtration carried out a Welch DryFast 2032 Ultra Diaphragm Pump (Niles, IL). Equipment and growth media sterilized using a Hirayama HICLAVE HV-50 autoclave (Kasukabe-Shi, Japan). All agar plates were incubated in a VWR 120 V forced air microbiological incubator (Radnor, PA). All broth cultures were incubated in a Thermo Fisher Scientific MaxQ 5000 floor-model shaker (Waltham, MA). Bacterial cells were lysed using a Branson Sonifier S-450A (Danbury, CT). All centrifugation steps during purification were carried out using a Thermo Fisher Scientific Sorvall LYNX 6000 Superspeed centrifuge. Nickel affinity chromatography, ion exchange chromatography, and size exclusion chromatography were all performed using a Bio-Rad Biologic low pressure (LP) chromatography system (Hercules, CA). Size Exclusion Chromatography measurements were done in a Bio-Rad 30 x 1.0 cm Econo-Column Standard Jacketed Column. CD spectroscopy was performed using a Jasco J-710 spectropolarimeter equipped with a Jasco spectropolarimeter power supply, and Jasco

PFD-425S Peltier (Easton, MD). Dynamic Light Scattering was performed using a Zetasizer Nano ZS from Malvern Instruments (Worcestershire, UK). All centrifugation of samples for analysis were done in a Beckman Coulter Microfuge 16 (Indianapolis, IN).

2.2. Sequences

The B domain of virulence factor protein A from *S. aureus* was chosen as the folded protein to have its sequence reversed as a model for this project. While the wildtype B domain was not structurally examined in this project, the sequence taken from the PDB entry 1BDC is shown in Figure 2.1 for reference.

```
1     TADNKFNKEQ QNAFYEILHL PNLNEEQRNG FIQSLKDDPS QSANLLAEAK
50    KLNDAQAPKA
```

Figure 2.1. Sequence for folded, WT B domain. Colors used to demonstrate the sequence directionality reversal. Note the blue tyrosine, the first amino acid in the sequence, and the red alanine at position 60 in the WT are now residues 62 and 3, respectively, in the Retro, shown below. All neighboring amino acids are maintained.

To examine the structural character of second reverse sequenced protein, the primary sequence in Figure 2.1 was reversed. The sequence for the resulting protein, "Retro B domain," is shown in Figure 2.2. The Retro B domain sequence has 3 amino acids that were not originally present in the WT, shown in green. Glycine and Serine are residual residues as a result of thrombin cleavage to remove the histidine affinity tag. Tryptophan was added to the sequence as a means to monitor absorbance of the protein.

```
1     GSAKPAQADNLK KAEALLNASQ SPDDKLSQIF GNRQEENLNP LHLIEYFANQ
53    QEKNFKNDATW
```

Figure 2.2. Sequence for Retro B domain. Blue tyrosine and red alanine are original first and last residues, respectively, in the wildtype B domain. Green lettering represents residues that were not present in wildtype B domain but are present in the Retro B domain.

To examine the role of sequence organization in temperature dependent hydrodynamic size, three scrambled nuclease variants were created. These protein sequences (Figures 2.4-2.6) contain the same composition as Retro nuclease (Figure 2.3) but different arrangement of side chains. Sequence organization was determined using the random number generator Random.org, which uses air disturbances as its basis for its randomness.

```
1     GQGSDANDES WINLKEKKAQ AESKRLHQEH TNNPKYVYAV KALGQRVKAE
50    NVMKGDAYIY ALGRGYKDTR QGKDFEVEIK KANEVMKKTF ASAEPGYKEV
100   GKKPHKTEPT DVLLLRFTMP QGKYMLKVTD GDIAKILTAP EKHLKKTSTA
```

Figure 2.3. Sequence for Retro nuclease. Retro nuclease has the same sequence as WT nuclease, with the exception of sequence inversion and one extra amino acid, shown here in green. Colors here are used to compare to Δ131Δ constructs. Note the position of the blue alanine and the red tryptophan in Retro nuclease which are the first and last amino acids, respectively, of Δ131Δ's sequence. All neighboring amino acids are maintained.

```
1     GTHTDKKKVI NSARHGKRKE DKKPELYTNF PGTNGHEKPA YTLTHEDMKV
50    QKERFAKLPL MTLADDLSAG GQAGKAKMYE KFSKPLYVSE KEASYKVVVK
100   ETGENKAGDM TQYDDEIQNG LKKVQVNIAP RVALRWALEA TGAQYLIKKI
```

Figure 2.4. Sequence for Scrambled nuclease A. Scrambled nuclease A has the same sequence composition as Retro nuclease with a different sequence organization.

```
1     GGALDKLGDV VENTKEKLGM KAAMEDVITH HRAVKEAYGE TGMIKDSISK
50    KVAGTRYAKR LLPMYEKDGK THAWLKKDPP KGEYSFLSPV QKAQLYNKVQ
100   YVTLAFTASE QNNEENKLLG KIKIYTQKTK PERDGDVQRF NHEPAKKTAA
```

Figure 2.5. Sequence for Scrambled nuclease B. Scrambled nuclease B has the same sequence composition as Retro nuclease with a different sequence organization.

```
1     GTNLGQKNLA QKLKIGLREY FLNTGSREEV IVHDKDEKKA PEHGIGEASG
50    KPTRNHMTKS VKDDKTKLVY RKQATADNPK AGMSAMVDPA YNKKEEKYYA
100   DKQEELKAYA LWKRGTAQIS HPTLFPDMTK KKVAELVKVA TVLEGQFYIG
```

Figure 2.6. Sequence for Scrambled nuclease C. Scrambled nuclease B has the same sequence composition as Retro nuclease with a different sequence organization.

To compare local structure measured in Retro and Scrambled nuclease to global organization in an unfolded protein, *S. aureus* nuclease variant Δ131Δ, as well as Retro Δ131Δ, were purified and analyzed. Δ131Δ contains the same composition and arrangement of amino acids found in WT nuclease with the exception of 9 deletions at both the N- and C- terminus, resulting in a total of 18 deletions. The sequence taken from the PDB entry 1STN, minus residues 4-12 and 141-149, is shown in Figure 2.7. Retro Δ131Δ (Figure 2.8) contains the same composition as Δ131Δ but has reversed sequence directionality.

```
1     ATS------- --TLIKAIDG DTVKLMYKGQ PMTFRLLLVD TPETKHPKKG
50    VEKYGPEASA FTKKMVENAK KIEVEFDKGQ RTDKYGRGLA YIYADGKMVN
100   EALVRQGLAK VAYVYKPNNT HEQHLRKSEA QAKKEKLNIW ---------
```

Figure 2.7. Sequence for Δ131Δ nuclease. Δ131Δ is has the same sequence as WT nuclease, with the exception of 18 deletions. Colors here are used to demonstrate the sequence directionality reversal. Note the blue alanine, the first amino acid in the sequence, and the red tryptophan at position 131 here are now residues 131 and 1, respectively, in Retro Δ131Δ, shown below. All neighboring amino acids are maintained. Dashes used to demonstrate the position of sequence deletions.

```
1     WINLKEKKAQ AESKRLHQEH TNNPKYVYAV KALGQRVLAE NVMKGDAYIY
50    ALGRGYKDTR QGKDFEVEIK KANEVMKKTF ASAEPGYKEV GKKPHKTEPT
100   DVLLLRFTMP QGKYMLKVTD GDIAKILTST A
```

Figure 2.8. Sequence for Retro Δ131Δ nuclease. Retro Δ131Δ is has the same sequence as Δ131Δ but has a reversed sequence directionality. Colors here are used to demonstrate the sequence directionality reversal as noted above. Dashes used to demonstrate the position of sequence deletions. All neighboring amino acids are maintained.

2.3. Expression and Purification of Recombinant Retro B domain

2.3.1. Cloning and Transformation

The Retro B domain sequence shown in Figure 2.2 was synthesized and cloned into a pJ414 expression vector by ATUM (Newark, CA). The plasmid contains a high-copy origin of replication, a gene that confers ampicillin resistance (*amp*^{r}), and a gene that codes for the lactose repressor protein (*lacI)*. The gene coding for Retro B domain, an N-terminal 6-Histidine tag, and thrombin cut site was placed under the control of an isopropyl β-D-thiogalactopyranoside (IPTG) inducible T7 promoter.

The vial containing 2-5 ug plasmid DNA was centrifuged for two minutes at 2000 xg using a Beckman Coulter Microfuge 16 centrifuge. The plasmid DNA was then solubilized in 200 µL of sterile water and allowed to incubate for 10 minutes. The plasmid DNA was divided into 6 µL aliquots and stored at -80ºC.

An aliquot of One Shot BL21(DE3) chemically competent *E. coli* cells from Invitrogen (Carlsbad, CA) was combined with an aliquot of Retro B domain plasmid DNA and gently flicked several times to mix. The mixture was incubated on ice for 30 minutes and then heat shocked for exactly 30 seconds in a 42 ºC waterbath before being immediately returned to ice. The cells were then combined with 250 µL of super optimal broth with catabolite repression (SOC). The cells were then incubated at 37 ºC for 1 hour at 225 RPM.

After outgrowth, the cells were spread on Lysogeny Broth (LB) agar plates containing 100 mg/mL ampicillin (LB + Amp). Two plates with ampicillin were used, with 50 µL pipetted onto the first and 100 µL pipetted onto the second. A third plate without ampicillin was used as a negative control and used the remainder of the 100 µL

of the cells. A glass spreader, dipped in 100% ethanol and flamed to sterilize, was used to spread the cultures one each plate. Each plate was incubated for 15-24 hours at 37 ºC.

2.3.2. Protein Expression

After overnight incubation, a sterile metal applicator was used to pick a single colony from one of the ampicillin-containing plates. This colony was transferred to a culture tube containing 3 mL LB + Amp. As a precaution, two additional test tube broth cultures were prepared from the same plate using two different colonies. The tubes were incubated for 15-20 hours at 30 ºC with 225 RPM. Glycerol stocks were made by combining 750µL of the overnight broth culture with 250µL of 50% (v/v) glycerol in a 2-mL cryovial and stored at -80 ºC.

The OD_{600} of the test-tube overnight broth cultures were measured the following day. A test tube with an overnight OD_{600} between 2.5-3.5 was chosen to subculture. The appropriate amount of overnight broth culture was added to a 4-L flask containing 1-liter LB + Amp so that the starting OD_{600} was ~0.01. The 1-liter expression culture was incubated at 37 ºC with 225 RPM for approximately 3 hours until the culture reached OD_{600} between 0.4 and 0.6, indicative that the culture is in mid-log phase. At this point, 1 mL of 1.0 M IPTG was added to the cell culture (final concentration 1.0 mM) to induce protein expression. The culture was incubated at 37 ºC with 225 RPM shaking for 5 hours. The cells were then centrifuged at 14,000 RPM for 15 minutes at 4 ºC. The supernatant was discarded, and the cell pellets were stored at -80 ºC until purification.

2.3.3. Nickel Affinity Chromatography

Cell pellets were thawed and solubilized in 25 mL of 6M guanidine, 10 mM TrisHCl, 100 mM Na_2HPO_4, pH 8.0 (Lysis buffer) and sonicated on ice at 80% duty,

output control 5. Sonication was done using four cycles of 90 seconds on and 90 seconds off. The cell lysate was centrifuged at 14,000 RPM for 60 minutes at 4 ºC.

During this centrifugation step, a 15 x 1.5 cm Bio-Rad Econo-column was cleaned and filled with ~15 mL HIS-Select nickel affinity gel from MilliporeSigma (St. Louis, MO). The column was attached to a low pressure (LP) chromatography system and equilibrated with Lysis Buffer until the conductivity and absorbance at 280 nm (A_{280}) was consistent.

After centrifugation, the supernatant was loaded onto the nickel affinity column and then washed with 30 mL of Lysis buffer. The column was washed subsequently with 30 mL of 10 mM TrisHCl, 100 mM Na_2HPO_4, pH 8.0 (Wash buffer #2), to remove all traces of guanidine, followed by 30 mL of 10 mM TrisHCl, 100 mM Na_2HPO_4 (Wash Buffer #3), and finally 10 mM TrisHCl, 100 mM Na_2HPO_4, 10 mM imidazole, pH 8.0 (Low imidazole Wash buffer) to remove any proteins with weak Ni^{2+} affinity. The target protein was eluted with 30 mL of 10 mM TrisHCl, 100 mM Na_2HPO_4, 350 mM imidazole, pH 4.8 (Elution buffer). Collection of the eluate in a sterile 15-mL conical began as soon as the conductivity began increasing and until the A_{280} reading hit a local minimum. Traditionally, the eluate would then be dialyzed against Tris-buffered Saline (TBS), but due to the low negative charge of Retro B domain, any salt in the buffer caused affinity issues in anion purification in the next step. Due to this observation, the collected sample was instead dialyzed overnight at 4 ºC against a 10mM TrisHCl, 0mM NaCl, pH 8.0. An example chromatograph for Retro B domain is shown in Figure 2.9 at the end of this Chapter.

2.3.4. Ion Exchange Chromatography

After dialysis against 10 mM Tris HCl, pH 8.0, the nickel column eluate was placed in a sterile 15 mL conical tube. A 30 µL aliquot (50 U) of recombinant human thrombin from MilliporeSigma was thawed and added to the protein sample. The conical was wrapped with tin foil, taped to a Thermo Fisher Scientific Large 3-D rotator, and covered by a box to prevent interaction of thrombin with light while digestion occurred. Thrombin digestion of the HIS-tag proceeded for 4 hours at room temperature with gentle rotation (speed 4).

During thrombin digestion, the ethanol from the DEAE Sephacel from GE Healthcare (Chicago, IL) media was decanted and replaced with 30 mL 20 mM sodium acetate, 0 mM NaCl, pH 4.8 (Equilibration buffer). The media was placed into a slurry and poured into a 50-mL beaker. The beaker was loosely covered with parafilm and degassed for at least 30 minutes or until no bubbles were visible. After degassing, the DEAE media was gently poured into a 15 x 1.5-cm Bio-Rad Econo-column. The column was attached to a LP chromatography system and equilibrated with equilibration buffer until the conductivity and A_{280} was level.

Once digestion was completed, the protein sample was loaded on the DEAE column, and washed with 30 mL of Equilibration buffer. Because of the low net charge of Retro B domain, it was hypothesized that it would elute at a low salt concentration. The column was washed with 30 mL of 20 mM sodium acetate, 25 mM NaCl, pH 4.8 (Wash Buffer 25) and the target protein was eluted. The eluate was collected in a sterile 15-mL conical and began as soon as the conductivity began spiking until the A_{280} reading began to level out. The eluate was then dialyzed overnight at 4°C against phosphate-

buffered saline (PBS; 10 mM sodium phosphate, 100 mM NaCl, pH 7.0).

An example chromatograph for Retro B domain is shown in Figures 2.2 at the end of this chapter.

2.3.5. Gel Electrophoresis, Staining, and Imaging

Purity of each protein purification was assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). A mixture of 20 µL protein, 20 µL 95% Laemmli Sample Buffer, 5% β-mercaptoethanol (BME - Laemmli Buffer) was created in a 500 µL microcentrifuge tube. The mixture was heated for 7 minutes at 95°C using an Eppendorf Thermomixer R (Hamburg, Germany). During heating, a handcast Bio-Rad 10% Mini Criterion Tris-HCl gel was placed into a Bio-Rad Mini-Protean Cell. The unit was filled with 700 µL of 250 mM TrisHCl, 192 mM Glycine, 0.1% SDS, pH 8.3 (1X TGS Running Buffer). After heating, 15 µL of each sample was loaded into their respective wells. One well contained 4 µL Bio-Rad Precision Plus Protein Standards. Any unused wells were loaded with 4 µL of Laemmli Sample Buffer to ensure even electrophoresis. The cell was attached to a Bio-Rad PowerPac HV power supply. The gel was electrophoresed at 200 V for approximately 50 minutes at 25°C.

Following electrophoresis, the gel was put into ddH_2O for approximately 5 minutes. The gel was then soaked in 10% acetic acid, 40% methanol, 0.05% Coomasssie Brilliant Blue R-250 (Coomassie Stain solution) overnight. The next morning, the gel was soaked in 10% acetic acid, 40% methanol (Destain solution) for one hour. After one hour, the Destain solution was replaced and this was repeated until the gel was sufficiently destained. The gel was then left in water for 30 minutes before imaging. The gel was imaged with a Bio-Rad Molecular Imager ChemiDoc XRS+ imaging system.

Example gel for Retro B domain is shown in Figure 2.3 at the end of this chapter.

2.3.6. Protein Concentration and Storage

Due to the concentration requirements of later techniques, Retro B domain needed to be concentrated following purity assessment. After dialysis, Retro B domain was evenly aliquoted into 2-mL microcentrifuge tubes. The tubes were placed in a ThermoFisher Scientific Savant DNA 120 SpeedVac and run without heat until the volume of each microcentrifuge tube was ~0.5 mL. The samples were consolidated and concentrated until an approximate concentration of 0.8-1.0 mg/mL. Approximate concentrations were assessed with a ThermoFischer NanoDrop™ 8000 Spectrophotometer. Once the desired concentration was reached, the samples were combined and dialyzed at 4 °C overnight against PBS. The following day, the protein sample was transferred to a sterile 15 mL tube, aliquoted into 500 µL microcentrifuge tubes, and stored at -80 ºC for future use.

2.4. Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy was used to gain insight to the secondary structure and backbone properties of Retro B domain. A Hellma Analytics High Precision quartz cell with 1-mm path length was filled with 300 µL of Retro B domain at a concentration of 0.17 mg/mL. The cell was placed in the spectropolarimeter, which was then purged with N_2 gas for 15 minutes at 22 PSI. The spectropolarimeter was turned on, followed by the water pump and Peltier.

JASCO Spectra Manager software was used to record spectropolarimeter measurements with the following settings: Standard sensitivity, 0.5 nm pitch, continuous 20 nm/min scanning, 2.0 second response, and 1.0 nm band width. Absorption was

measured from 198 nm to 245 nm with 8 accumulations. Temperature was set to 5°C. A blank of the same solution concentration made by previous lab members was used. The temperature within the spectropolarimeter was allowed to equilibrate for 10 minutes before measurements began. Before the first temperature scan (5°C), the N_s pressure was decreased to 17 PSI. After the first scan, the N_2 pressure was decreased to 12 PSI where it remained for the remainder of the experiment. CD spectra was measured from 5–85°C in 10° increments with careful HT voltage monitoring. Upon a voltage reading greater than 700 V, the absorption range was decreased for the next temperature scan. After all temperature scans were complete, the Peltier and water pump were turned off followed by the spectropolarimeter. The equipment was purged for an additional 15 minutes with N_2 gas before the quartz cell was removed.

All CD measurements were converted to mean residue ellipticity (MRE) using the formula below where CD is circular dichroism (degrees), MW is molecular weight (g/dmol), l is path length (cm), n is residues (res), and c is concentration (g/mL).

$$\theta_{MRE} = \frac{CD \cdot MW}{l \cdot n \cdot c}\ deg\ cm^2\ dmol^{-1}\ res^{-1}$$

CD difference spectra was created with the 85 ºC spectra as reference. For example, the 75 ºC spectra is shown in orange in Figure 3.1 and is the result of subtracting the 85 ºC MRE value at a given wavelength from the 75 ºC MRE value at the same wavelength, demonstrated by $MRE_{75\text{-}85} = MRE_{75^oC\ at\ wavelength\ x} - MRE_{85^oC\ at\ wavelength\ X}$. This is because it is hypothesized that there is little structural character at high temperatures and more residual structure at lower temperature. By doing this, we are

able to see the small changes that are happening as temperature decreases from 85 ºC and at what wavelengths so that we might understand what secondary structure components may be being sampled by the polypeptide.

2.5. Size Exclusion Chromatography

Size exclusion chromatography (SEC) is a method that can be used to determine the size of proteins based on diffusion of molecules. In SEC, larger proteins are able to pass through the fluid volume of the column at a quicker rate than smaller proteins that are slowed down by their diffusion into the porous media. SEC was used to measure both the partition coefficients (K_D) and the R_h of Retro B domain. Chicken egg albumin (Alb), bovine carbonic anhydrase (CA), and horse heart myoglobin (Myo) were used as folded standards from 5 º-45 ºC. At temperatures where folded proteins could not be used (i.e, T>45 ºC) the IDPs p53(1-93), p53(1-93) with substitutions of its 12 alanine residues to glycine (p53 Ala^-), and Retro nuclease were used as they all have experimental mean R_hs measured previously by DLS. The R_h used for each protein standard was determined experimentally by DLS previously by other lab members.[46, 47, 53]

Table 2.1. Folded Protein Standards

Folded Protein Standard	PDB	Residues	Molecular Weight (Da)	$R_{h,crys}$ (Å)
Chicken Egg Albumin	1OVA	386	172,436.66	35.763
Bovine Carbonic Anhydrase	1V9E	259	58,171.94	27.339
Horse heart myoglobin	2O58	153	17,791.22	21.83

Table 2.2. IDP standards

IDP Standard	Residues	Molecular Weight (Da)	$R_{h,exp}$ (Å) at 55ºC	$R_{h,exp}$ (Å) at 65ºC
p53(1-93)	94	10,123	29.478	28.239
p53 Ala-	94	9,954.92	28.152	26.864
Retro nuclease	149	16,868	27.649	27.474

2.5.1. Preparation of Running Buffer

Four liters of PBS were filtered through GE Healthcare Whatman 1 qualitative filter paper using a filter flask and a Welch DryFast Ultra vacuum pump. The buffer was allowed to sit for 10 minutes with the vacuum running to degas. The degassed PBS was transferred to a 5-L beaker and placed on a surface near the pump on the Biologic LP chromatography system.

2.5.2. Preparation of Media

Approximately five grams of GE Healthcare Sephadex G-75 media was hydrated with filtered and degassed PBS to a volume of 150 mL and allowed to swell overnight at room temperature. The next day, a volume of PBS was gently pipetted out to leave a 75:25 media:solvent ratio. The media was placed in suspension, transferred to a filter flask, and placed on a hot plate to heat to approximately 75ºC. Once heated, the media was degassed for at least an hour or until no more gas was released by the mixture.

After thorough degassing, the media was placed in suspension and poured in one continuous motion into the jacketed column. For high temperature SEC experiments, the column was equilibrated to 65ºC before adding the media into the column. For low temperature SEC experiments, the degassed media was allowed to cool and equilibrate in a 4 ºC VWR chromatography refrigerator for 30 minutes. The column was equilibrated

to 5 ºC and the media was poured in one continuous motion into the jacketed column. The bottom of the column was attached to an LP chromatography system while the top of the column was attached to a 90-cm segment of Saint-Gobain Tygon tubing with an inner diameter of 1.59 mm and an outer diameter of 3.18 mm (Courbevoie, France). The other end of Tygon tubing was anchored to the bottom of the 5-L beaker containing filtered and degassed PBS to create a syphon. The column was packed by setting the LP Biologic pump to 0.6 mL/min overnight.

2.5.3. Preparation of Protein Standards, Blue Dextran, and DNP-Aspartate

Albumin, carbonic anhydrase, and myoglobin were chosen as folded protein standards due to their globular, folded structure, wide R_h range, and their previous structural characterization. These proteins have been solved by XRC and as a result the R_h for each folded protein standard can be estimated from the crystallographic structure by calculating the maximum distance between the two furthest alpha carbons in the chain. Similarly, these proteins have been measured by DLS previously[46] and the two methods produce R_h values that are in good agreement.[47]

Aliquots of Chicken egg albumin (Thermo-Fisher Scientific Acros 40045) and bovine carbonic anhydrase (Sigma-Aldrich C5024), previously prepared by other lab members, were used. Albumin was prepared by hydrating Albumin dry powder with PBS to a concentration of ~5 mg/mL. The hydrated protein was filtered with a 0.45 μm polyvinylidene fluoride (PVDF) membrane followed by a second filtration through a 0.2 μm PVDF membrane before it was stored at -80°C. Carbonic anhydrase was prepared by hydrating lyophilized protein with 5 mL PBS. Once hydrated, the sample was filtered with a 0.2 μm PVDF membrane. The filtered protein was stored as 200 μL aliquots at -

80°C.

Horse heart Myoglobin (Sigma-Aldrich M1882) was hydrated with PBS to a concentration of 20 mg/mL. The hydrated protein was filtered with a 0.2 μm PVDF membrane and stored for short-term use at -80°C.

Recombinant proteins p53(1-93), p53 Ala-, and Retro nuclease were purified in the same manner detailed in Section 2.3 by various lab members, including Lance English, Sarah Voss, and Allison Bebo, respectively, for other projects, were used when folded protein standards could not be used (T>45 ºC).

Blue dextran is a dyed sugar of approximately 2,000 kDa and is used to establish the void volume, or the volume of a molecule that does not experience retention by the porous beads of the media, of the column. Blue Dextran D5751 from MilliporeSigma was prepared by solubilizing 0.1 g in 10 mL of PBS Running Buffer to achieve a Blue Dextran (BD) Stock with a concentration of 10 mg/mL.

Along the same lines of logic, the maximum retention volume of the column must be established in order to calculate the K_D of the protein of interest. Calculation of K_D is used to describe how the protein being examined is distributed between the mobile phase (the fluid volume) and the stationary phase (the porous media). DNP-Aspartate, a single amino acid containing dinitrophenyl and approximately 300 Da in size, was used to establish the total volume (V_t) of the column. DNP-aspartate (DNP-Asp) from Sigma-Aldrich was solubilized with 10 mL of PBS Running Buffer to achieve a stock with a concentration of 1.5 mg/mL. 0.6 mL of BD, 0.4 mL of DNP-Asp, and 1 mL of PBS Running Buffer were combined to create a BD-DNP-Asp stock solution to combine with samples or run individually.

2.5.4. Size Exclusion Measurements

Proteins were removed from the -80ºC and allowed to thaw. Once thawed, proteins were transferred into 1.5 mL microfuge tubes and centrifuged at 10,000 RPM for at least 2 hours.

During this time, the column was flushed with PBS at a flow rate of 0.3 mL/min to equilibrate the column and allow the UV lamp to warm up for measurements.

To measure K_D, each protein and protein standard was analyzed in one of the following manners. One way that proteins were measured was by running a sample of protein mixed with BD-DNP-Asp. Immediately before running the sample, 4 µL BD-DNP-Asp (3 mg/mL Blue Dextran D5751, 0.75 mg/mL DNP-Aspartate) was mixed with, in most cases, 50 µL of protein. If a sample was particularly dilute, the volume of protein used was increased. The mixture was then pipetted directly on to the top of the matrix bed and covered with 3 mL PBS. Once reconnected, the pump was started, and the start time noted by using the mark option on the Biologic LP system.

Another way that proteins were measured was by running BD-DNP-Asp and the protein of interest separately. To do this, a 4 µL sample of BD-DNP-Asp was measured, as detailed above, followed immediately by a sample of protein. This was done if resolution of the peaks was poor using this first method or for albumin, which is very large and often overlaps in elution with BD. Each protein and protein standard was measured a minimum of three times at each temperature using one or both of these methods. Standard deviation of the mean of K_D, reported in Table 3.1, are from repeat measurements.

2.5.5. Measuring Temperature Effects

The entire procedure was repeated at temperatures 10 ºC, 35 ºC, 45 ºC, 55 ºC, and 65 ºC with water from the water bath running through the jacketed part of the column. To accomplish this, the following adjustments were made.

For temperatures greater than 25 ºC, hoses were connected from a VWR water bath to the column, and from the column to a food grade Chugger Pump CPPS-in-1 115V beer pump (Farmingdale, NJ). When turned on, heated water from the water bath was pumped through the outer layer of the column, effectively changing the immediate environment of the SEC media to the respective heat of the water bath.

For the 10 ºC measurement, the pump, a Bel-Art H-B DURAC Calibrated Electronic Thermometer (Wayne, NJ), and a 4L glass beaker were placed inside of a VWR Chromatography refrigerator set to 4 ºC. The Biologic LP system was moved on to a mobile cart to be relocated directly adjacent to the fridge so that the set-up from before (water source to column to pump) would be maintained. To retain cold temperatures, tubular foam pipe insulation from Lowe's was wrapped around the column. When turned on, cold water from the glass beaker in the fridge was pumped through the outer layer of the column, effectively changing the immediate environment of the SEC media to the temperature read by the Durac thermometer in the beaker in the fridge. To further check the accuracy of the column temperature, a second Durac thermometer was placed between the column and the foam insulation.

2.5.6. K_D and R_h Calculation

K_D for each protein measured was calculated according to the equation below.

$$K_D = \frac{V_e - V_0}{V_t - V_0}$$

Here, V_0 is the void volume, the volume in which BD elutes from the volume. V_e is the elution volume, the volume in which the protein of interest elutes from the column. V_t is total volume, the volume in which DNP-Asp elutes from the column.

As seen previously, this project found that K_Ds are only interpretable within the context of the same column. However, within the same SEC experiment (no resuspension of column media), K_Ds are very repeatable. Hence, a full set of standards were run for each temperature as resuspensions often had small but significant enough differences to affect calculations.

The crystallographic R_h of each protein standard (Table 2.1) was plotted against their experimental K_D value. The formula for the regression line, shown below, was used to extrapolate the R_h for Retro B domain at each temperature:

$$R_h = m_{regr} * K_D + b_{regr}$$

2.6. Dynamic Light Scattering

Dynamic light scattering is a standardized technique for particle size analysis in the nanometer range based on the Brownian motion of dispersed particles in which particles, constantly colliding with solvent molecules, are continuously moving in solution. The rate of Brownian motion can be quantified by the translational diffusion coefficient, *D*, which can then be used to determine the particle size. This relationship between the speed of the particles and the particle size is given by the Stokes-Einstein equation:

$$D = \frac{kT}{6\pi\eta R_h}$$

Here, *k* is Boltzmann's constant, T is temperature, η is viscosity, and R_h is the hydrodynamic radius. DLS uses an auto-correlation function of fluctuations in the

scattering of light overtime, to determine *D* and quickly measure the R_h of a sample.

2.6.1 Preparation of Sample, Cuvette, and Instrument

An aliquot of Retro B domain was removed from the -80 ºC and allowed to thaw. Once thawed, protein was transferred into a 1.5 mL microfuge tube and centrifuged at 10,000 RPM for at least 2 hours. Following centrifugation, samples were filtered using 0.45 um PVDF syringe-driven filters from EMD Millipore Corporation (Billerica, MA).

A quartz cuvette with path length of 1 cm was used for DLS measurements. To prepare it for use, the cuvette was cleaned with concentrated nitric acid. After the nitric acid was disposed of and the cuvette was allowed to dry thoroughly, 200 µL of PBS was pipetted up and down several times. This was repeated two times. Finally, the concentration of the protein was taken by measuring the absorbance at 280 nm. Based on previous research in the group, a concentration around 0.5 mg/mL was desired for measurements. When necessary, the sample was adjusted to an appropriate concentration.

While the cuvette was being cleaned and the protein sample was being prepped, the Zetasizer Nano was turned on and allowed to warm up for a minimum of 30 minutes to allow for thorough thermal equilibration of the instrument.

2.6.2 Dynamic Light Scattering Measurements

DLS readings used noninvasive backscatter optics and were measured using a Zetasizer Nano ZS with Peltier temperature control from Malvern Instruments (Worcestershire, UK). The use of Standard Operating Procedures (SOPs) were used to alleviate the need for constant attention. Addionally, an SOP file contains all the measurement settings including equilibration times, number of measurements, and

material and dispersant specific values required for performing repetitive and consistent measurements. Solvent viscosity was calculated by the solvent builder program provided by Malvern, which uses Sednterp[54] to estimate η from the solution contents. This dispersant viscosity was also applied as sample viscosity. Measurements were taken using 173º backscatter. These settings were kept consistent in all SOPs for all measurements at all temperatures.

The sample temperature was cycled in 10 °C steps and back from 5 to 65 °C to establish the reversibility of heat effects on mean Rh. Samples were equilibrated at each temperature for 5 minutes before measurement. Fresh samples were prepared daily from frozen stock. Mean R_h was calculated using the Stokes-Einstein relationship, detailed above. Mean R_h was measured over 3 times for each sample at each temperature. The standard deviation (σ) of the mean R_h from N number of measurements calculated by

$$\sigma = \sqrt{\Sigma \frac{(mean\ R_{h,i} - average\ mean\ R_h)^2}{N}}$$

2.7. Additional Figures

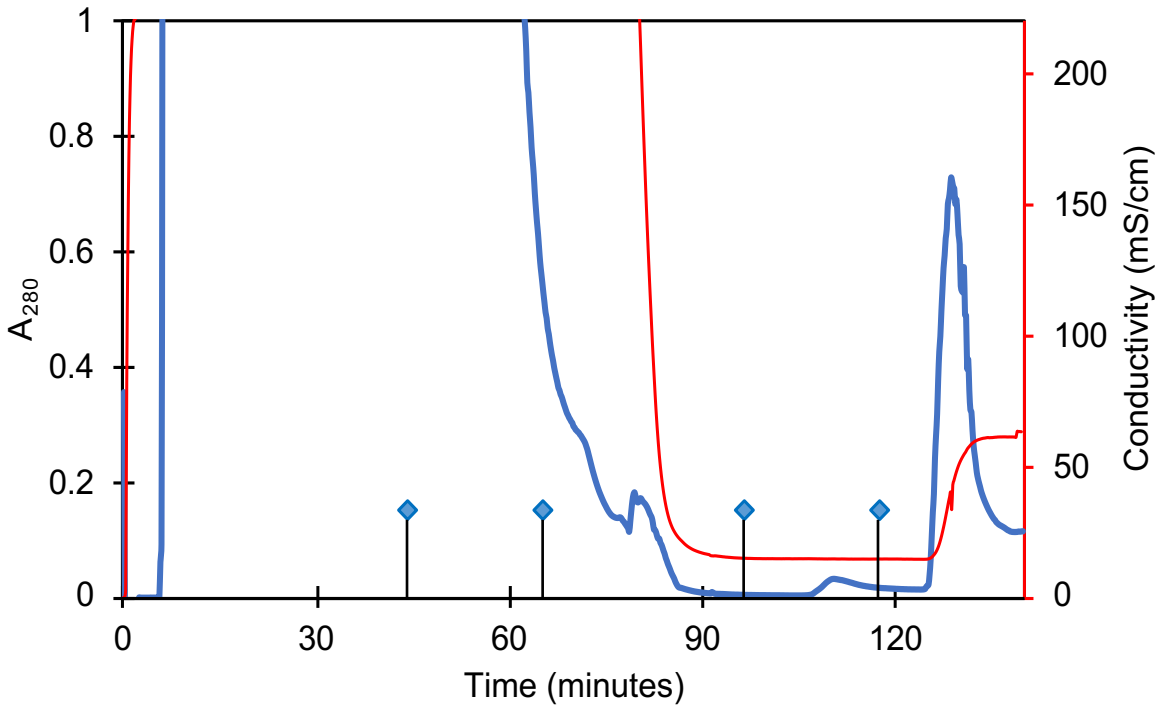

Figure 2.9. <u>Chromatogram for nickel affinity purification of Retro B domain.</u>
The blue line represents absorbance at 280 nm. The red line represents conductivity. The blue diamonds are event markers. In chronological order: sample load (5 min), addition of Lysis Buffer (Event Marker 1), addition of Wash Buffer 2 (Event Marker 2), addition of Low Imidazole Wash Buffer (Event Marker 3), and addition of Elution Buffer (Event Marker 4). The peak at 130 min is the target protein.

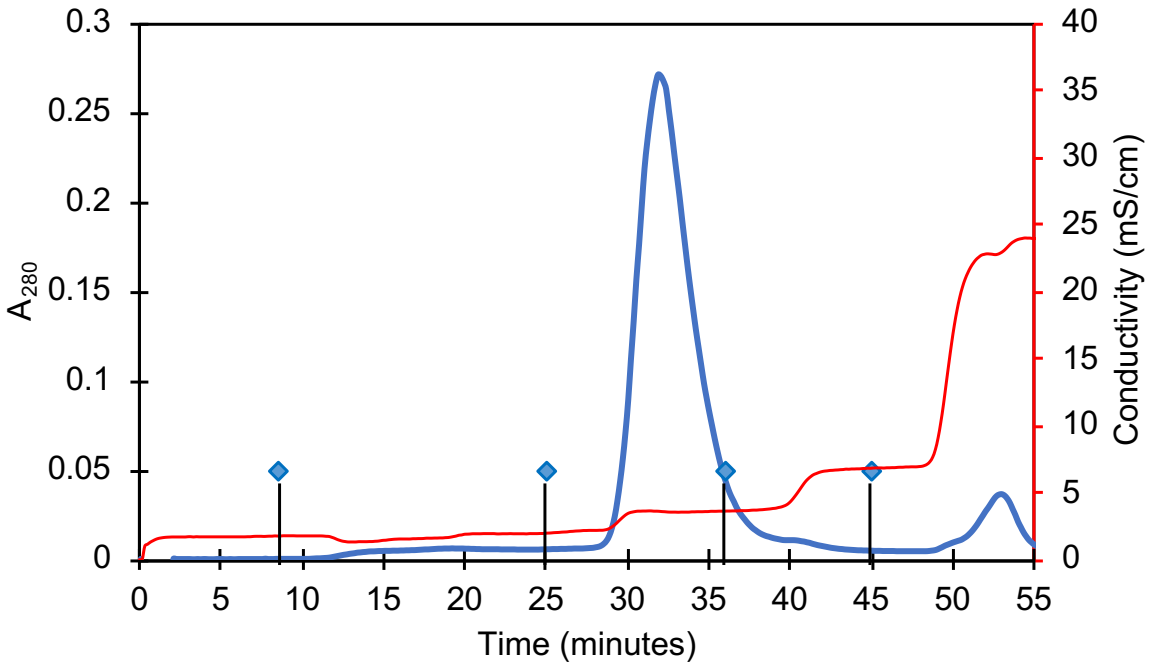

Figure 2.10. <u>Chromatogram for anion exchange purification of Retro B domain.</u>
The blue line represents absorbance at 280 nm. The red line represents conductivity. The blue diamonds are event markers. In chronological order: Post- thrombin digest sample load, addition of Equilibration Buffer (Event Marker 1), addition of Wash Buffer 25 (Event Marker 2), addition of Wash Buffer 50 (Event Marker 3), and addition of Wash Buffer 250 (Event Marker 4). The peak at ~30 min is the target, digested protein.

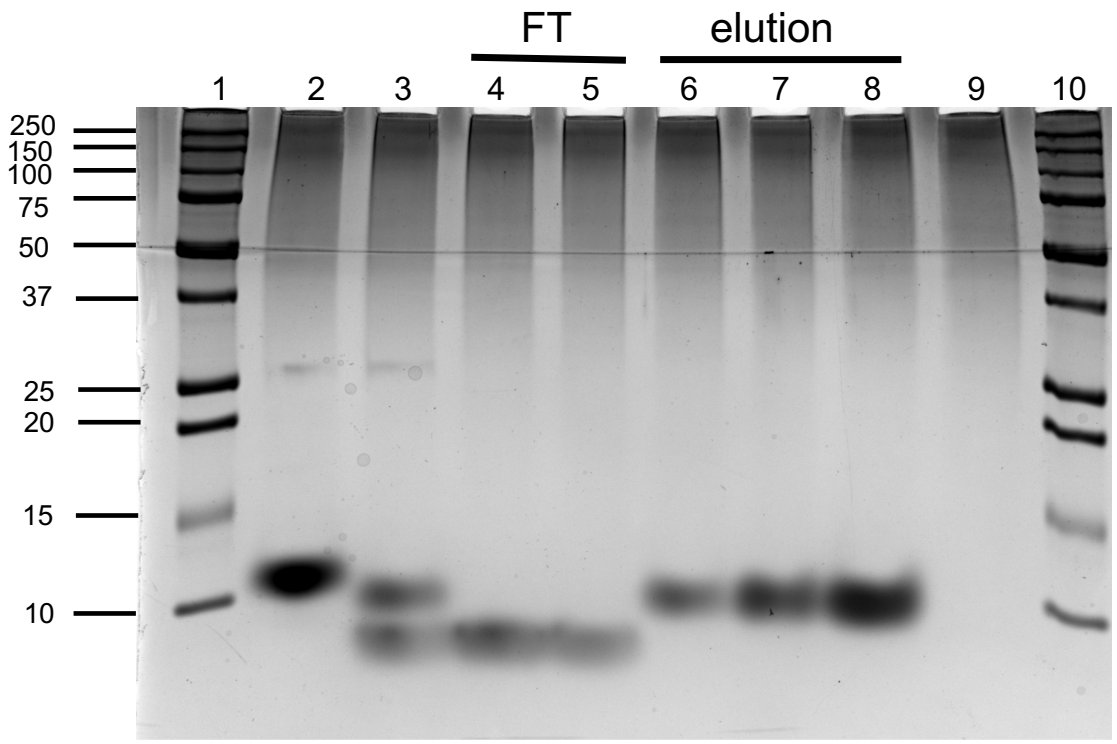

Figure 2.11. **<u>Coomassie stained polyacrylamide gel for Retro B domain purification.</u>** Retro B domain after Ni-NTA and anion exchange purification. The first and last lane is Precision Plus protein standards by Bio-Rad. The second lane is Retro B domain after Ni-NTA purification (before digest). The third lane is Retro B domain after digestion but before anion exchange purification. Lanes 4 and 5 are fractions from the flow-through (FT) of the anion exchange column when the sample was added to the column. Lanes 6-8 are fractions from the elution of Retro B domain using Wash Buffer 25. Lane 9 is BME-Laemli Running Buffer (empty lane).

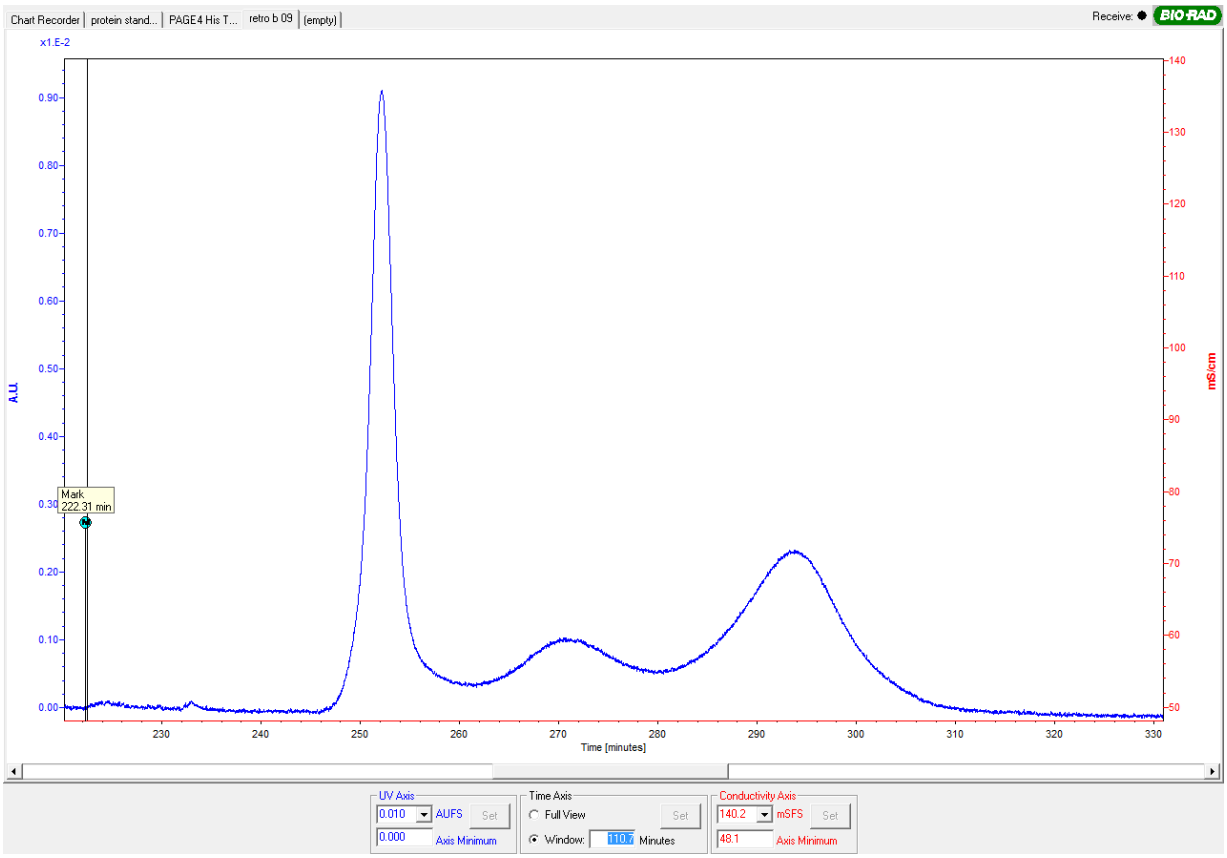

Figure 2.12. <u>Example SEC chromatogram.</u> The blue line represents absorbance at 280 nm. The blue circles marked "M" is an event marker that indicates the addition of a sample. K_D was calculated from elution time of BD (first peak), protein of interest (second peak), DNP-Asp (third peak).

Size Distribution Report by Intensity

v2.1

Malvern

Sample Details

Sample Name:	1		
SOP Name:	25 Run.sop		
General Notes:			
File Name:	Retro B domain 08.18.19	**Dispersant Name:**	10 mM Sodium Phosphate, 10...
Record Number:	15	**Dispersant RI:**	1.331
Material RI:	1.45	**Viscosity (cP):**	0.9087
Material Absorbtion:	0.001	**Measurement Date and Time:**	Sunday, August 18, 2019 2...

System

Temperature (°C):	25.0	**Duration Used (s):**	180
Count Rate (kcps):	26.9	**Measurement Position (mm):**	4.65
Cell Description:	Glass cuvette with square apert...	**Attenuator:**	11

Results

			Size (r.nm):	**% Intensity**	**Width (r.nm):**
Z-Average (r.nm):	139.3	**Peak 1:**	2.068	100.0	0.3823
PdI:	0.431	**Peak 2:**	0.000	0.0	0.000
Intercept:	0.479	**Peak 3:**	0.000	0.0	0.000

Result quality Refer to quality report

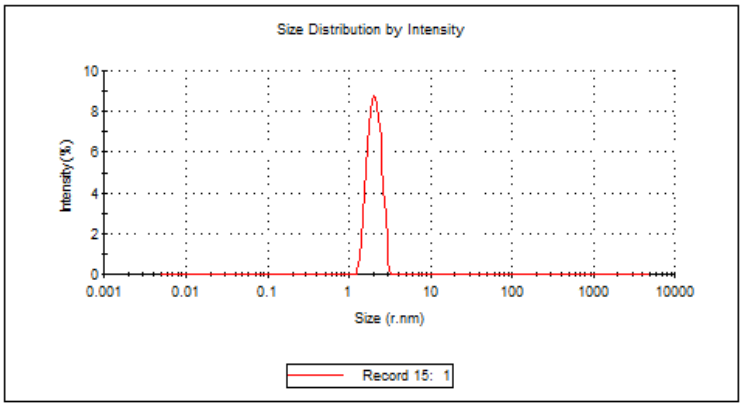

Figure 2.13. <u>Example DLS report.</u> Report given by Malven Zetasizer. Each report contains important information including duration of scan, dispersant and dispersant RI, viscosity, temperature, size in nanometers, and any polydispersity in the sample.

3. RETRO B DOMAIN RESULTS AND DISCUSSION

3.1. Introduction

It has been known for over 80 years that protein macromolecules are conformationally flexible[2], accessing an ensemble of conformational states in their folded and unfolded forms. While a folded, native conformation is typically fixed, unfolded proteins in solution enter a large and diverse ensemble of conformations, eliminating their study by traditional methods such as x-ray crystallography and NMR. The protein denatured state was slow to build academic interest and the lack of compelling evidence, the result of the challenges in characterization, led to the acceptance of the unfolded state as random coil. This notion has been challenged over the last 60 years with evidence that local and intermediate range states may persist in the denatured state[30, 31, 55] as well as the observation that the denatured state has important roles in folding, stability[20], and function.[10] And while the importance of DNA is unquestioned, how that genetic information is then translated into molecular activity via proteins, how proteins adopt their native configurations, and how that native configuration is dependent on sequence are arguably of equal consideration. This is especially true for biological process that are facilitated by proteins that rely on the properties and energies associated with the denatured state.

In Section 1.2, I detailed two recent findings by our lab. The first analyzed both the sequence dependence to the mean hydrodynamic size of IDPs and the impact of heat on the coil dimensions. These results were determined by using a model of the denatured state where conformational preferences were determined at the residue level and, importantly, show that the sequence dependence and thermodynamic energies

associated with intrinsic biases for PPII and α backbone conformations can be obtained.[46] Next, we designed experiments to determine if this method for measuring conformational bias is unique to IDPs or also applicable to sequences that contain a greater proportion of hydrophobic residues. By designing unfolded mutant proteins from foldable sequences, such as Retro nuclease, we found that changes in temperature caused large shifts in the hydrodynamic size that track with changes in the backbone structure and are close to expected values based on the sequence dependence and known magnitude of intrinsic polyproline II propensities, both measured in small peptides.[42, 43, 45]

These data were not only surprising but also support the idea that the unfolded state is not determined solely by steric considerations and may have very locally determined structural preferences that are also influential. However, Retro nuclease has been the only folded sequence composition tested and thus the only protein to exhibit this intrinsic behavior. To further test this hypothesis, this project extends this experimental system to the reverse-sequenced B domain of *Staphylococcal aureus* virulence factor protein A, "Retro B domain," to determine if temperature-induced changes are similarly observed in a second protein and thus not a unique result of *S. aureus* nuclease.

3.2. Structural Characterization of Retro B domain

Retro B domain was obtained for structural studies by recombinant expression in bacterial cells followed by purification from cell lysate using affinity chromatography. To analyze temperature-induced changes in its structural features, Retro B domain was evaluated using CD spectroscopy, SEC, and DLS. The results show that Retro B domain is an intrinsically disordered monomeric protein under physiological like conditions. Changes in temperature were observed to cause distinct changes in the coil structures of

Retro B domain, which is presented, discussed, and analyzed below.

3.2.1. Circular Dichroism Spectrum of Retro B domain

CD spectroscopy was used to examine the secondary structure of Retro B domain. At 25ºC, the Retro B domain spectrum is similar in character to the spectra reported for Retro nuclease (Figure 1.3, C) and for IDPs[49], with molar residue ellipticity (MRE) at 221 nm near zero (Figure 3.1). The spectrum exhibits a local maximum observed at ~225 nm that is usually seen in proteins with PPII secondary structure. This supports that Retro B domain does not take on native-like structure and is indeed disordered under physiological conditions.

Based on a survey of CD spectra from approximately 100 proteins[47], it has been argued that residual structure in IDPs can be categorized based on MRE values at 200 and 222 nm. IDPs are "mostly unfolded" with MRE (10^{-3} degree cm^2 $dmol^{-1}$ $residue^{-1}$) values from 12.0 to -19.4 at 200 nm and -5.0 to -2.8 at 222 nm while MRE (10^{-3} degree cm^2 $dmol^{-1}$ $residue^{-1}$) values between -21.7 to -16.1 and -2.4 to -1.0, respectively, are consistent with the "existence of some residual secondary structure."[49] Using that dichotomy, the Retro B domain CD spectrum at 25 ºC is showing signs of some residual secondary structure (Figure 3.1, A).

To look at the temperature dependence of this disordered state, CD spectroscopy was repeated at a range of temperatures. The Retro B domain spectra has subtle, but minimal changes across the temperature range. For instance, similar to the 25 ºC spectrum, a local maximum is observed in all spectra at ~225 nm that is usually seen in proteins with PPII secondary structure. However, this local maximum at ~225 nm can be seen to decrease in intensity with increasing temperature, consistent with the favorable

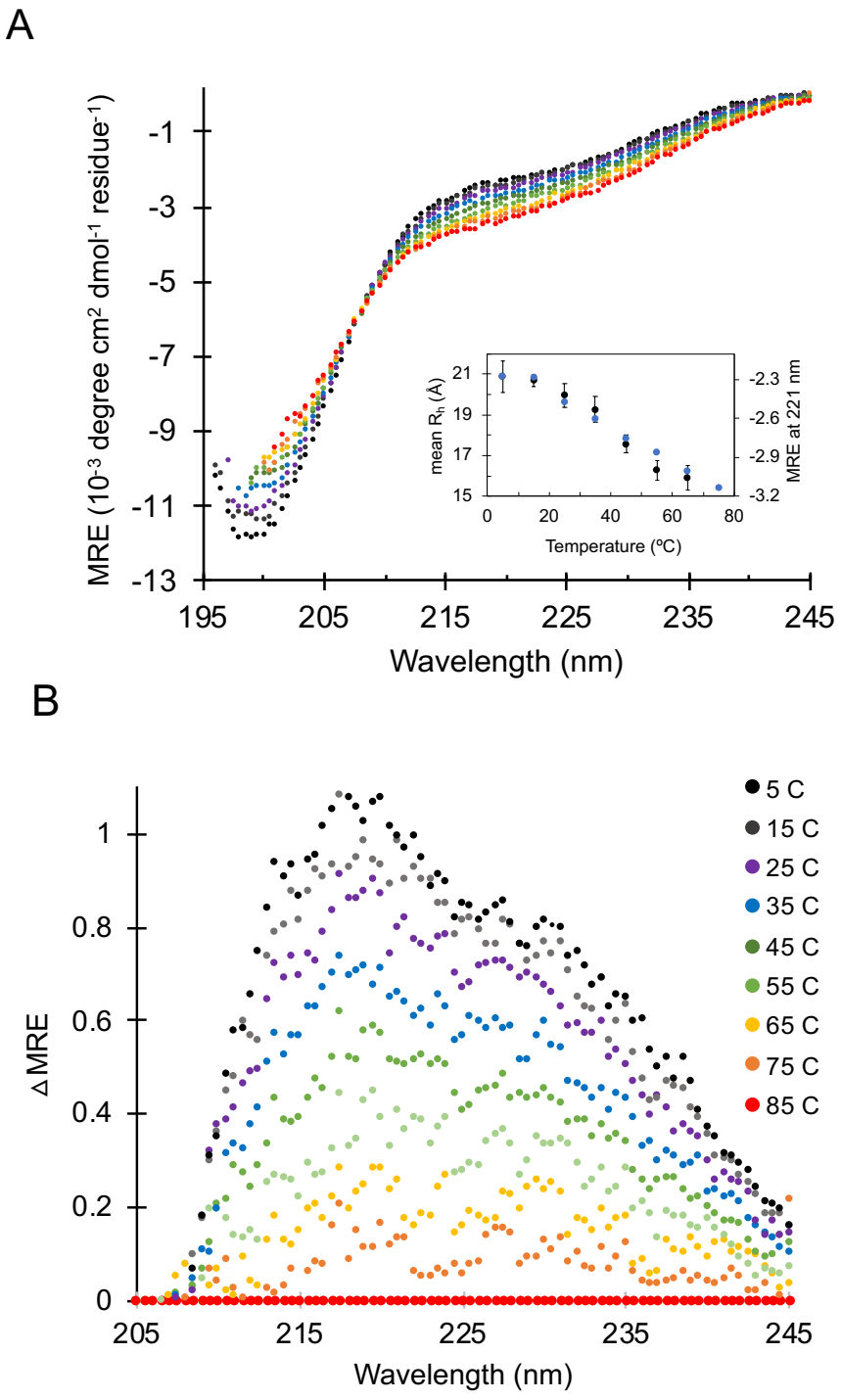

Figure 3.1. **<u>Temperature dependence of Retro B domain CD spectra.</u>**
A) CD spectra of Retro B domain from 5 to 85 ºC in 10º increments, Inset) DLS measured R_h (blue) tracks with changes in MRE at 221 nm (black) which is associated with PPII conformational sampling, B) Difference spectra in relation to 85 ºC spectrum.

enthalpy of the PPII bias in the denatured state.[32] Temperature induced changes in structural character are not easily characterized by traditional deconvolution of the CD spectra as satisfactory reference sets for disordered proteins are limited, owing to the spectral similarity of denatured structures.[56]

In order to elucidate the details within these subtle changes, CD spectroscopy results were also monitored by the difference spectra relative to the 85 ºC spectrum (Figure 3.1B). The 85 ºC spectrum was chosen as reference so that subtle changes relative to a high temperature state, which is presumably more of a random coil distribution because of the thermal agitation to the chain, could be better observed. In the difference spectra, there is a local maximum at ~222 nm that increases in intensity with decreasing temperature. A maximum at this wavelength is generally seen in proteins with PPII structures.[47] The temperature dependence of this maximum is indicative that PPII conformational sampling is occurring and increases in magnitude as the protein is cooled. This suggests that Retro B domain has increased PPII conformational sampling as temperature decreases. PPII conformational sampling is prominent in IDPs and unfolded proteins and is known to decrease in magnitude as it is heated.[21, 22, 57]

At low temperatures, the difference spectra changes and appears to have a local minimum at ~225 nm, first appearing at 25ºC. This is similar to the minima observed at ~222 nm seen in the Retro nuclease CD difference spectra, which is thought to be indicative of α-helical conformational sampling as local minima at or near 222 and 208 nm are observed in proteins with α-helices.[47] Based on the spectral indications of temperature-sensitive PPII and α-helix content (Figure 3.1, B), we investigated by DLS whether Retro B domain was experiencing significant hydrodynamic size changes such

as those observed previously in Retro nuclease.

3.2.2. Characterization of Retro B domain structure by Dynamic Light Scattering

Dynamic light scattering (DLS) was performed from 5 to 65 ºC in 10º increments to measure the temperature dependence of Retro B domain mean R_h. DLS shows a similar trend in Retro B domain as Retro nuclease, again exhibiting a sigmoid dependence with temperature (Figure 3.2). Temperatures greater than 25 ºC caused compaction, reducing the mean R_h from 19.9±0.6 Å at 25 ºC to 15.9±0.6 Å at 65 ºC. At temperatures less than 25 ºC, there is a slight increase in size that levels off at 20.8±0.8 Å at 5 ºC. At all temperatures, error bars indicate the standard deviation from the mean from repeat measurements of Retro B domain. It was observed in Retro nuclease that changes in MRE at 221 nm, the wavelength associated with PPII content, tracked with changes in DLS-measured mean R_h.[47] This was also found to be the case with Retro B domain (Figure 3.1, inset), again exhibiting a sigmoid dependence with temperature and correlation between temperature-induced changes in MRE at 221 nm and mean R_h.

As such, DLS has been shown to sensitive to the protein concentration.[47] As explained in Section 2.6, DLS determines R_h by examining the fluctuation of scattered light by particles as they diffuse overtime. Size plays a role in the diffusion of particles (i.e, larger particles diffuse more slowly than smaller particles) and polydisperse ensembles of proteins or interactions between proteins in solution affect the calculated R_h. To account for this sensitivity to concentration, DLS measurements are repeated at several concentrations so that the concentration dependence of mean R_h can be extrapolated back to zero protein concentration[47], giving a more accurate mean R_h at each temperature. However, Retro B domain precipitates out of solution at concentrations

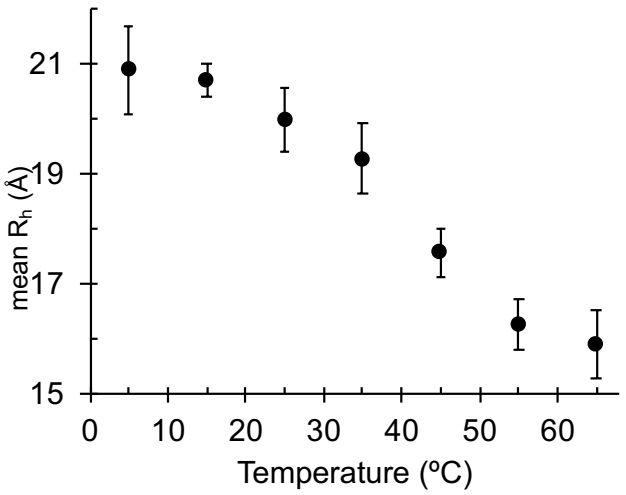

Figure 3.2. <u>Temperature dependence of Retro B domain mean R_h.</u> Retro B domain hydrodynamic size is temperature sensitive. Circles show mean R_h measured by DLS. The error is the standard deviation from repeat measurements.

around and above 0.8 mg/mL, limiting the range of concentrations at which Retro B domain could be measured by DLS. Previous experiments suggest that extrapolating the concentration dependence of mean R_h to zero protein concentration gave a near matching measurement at 0.5 mg/mL.[47] This suggests that ~0.5 mg/mL of protein, or lower, offers good evaluation of hydrodynamic size with minimal concentration effects. The mean R_h reported here is from DLS experiments with Retro B domain at a concentration of 0.3 mg/mL.

Overall, the changes in mean R_h from temperature effects on Retro B domain structure are qualitatively significant but, at each temperature, the error bars are quite large. This increased uncertainty stems from several factors. Protein hydrodynamic size analysis by DLS is a sensitive technique which functions optimally when samples are homogenous and of a single population.[58] When used to examine the hydrodynamic size changes of a disordered state such as here, the data by DLS is noisy. Biologically, Retro B domain is a smaller protein than those our group has previously characterized by DLS.

This can be observed by comparing the overall change in mean R_h of Retro B domain (just under 5 Å) to those observed in Retro nuclease (>7 Å, greater in size than the WT unfolding). Comparison of light scattering data to data obtained from other hydrodynamic methods allow for better, more meaningful interpretation. For these reasons and the inability to study Retro B domain at higher concentrations, SEC was used to compare Retro B domain's DLS measurements as well as look at other structural characteristics.

3.2.3. Hydrodynamic Size Analysis of Retro B domain by Size Exclusion Chromatography

Size exclusion chromatography (SEC) was utilized in several ways to analyze Retro B domain's hydrodynamic character. First, it was used to examine the solubility and stability overtime of Retro B domain. This would be descriptive qualitatively as well as help determine if Retro B domain would be suitable for measurement by DLS in the future as experimentation can often be several hours long to lasting overnight. The sample was initially centrifuged for 2 hours and then loaded on to the column at time points after centrifugation. This experiment demonstrated that Retro B domain did not seem to significantly aggregate over the necessary time range (Figure 3.3).

SEC was also utilized as a second method to analyze Retro B domain's hydrodynamic size. Thermodynamic retention factors (K_D) were calculated from elution volumes, with smaller K_D signifying larger mean hydrodynamic size (Figure 3.4, B). SEC chromatograms at temperatures from 8 to 65 ºC show that the hydrodynamic size of Retro B domain is temperature sensitive (Figure 3.4, A). Retro B domain's K_D increases from 0.45 ± 0.02 at room temperature (~23 ºC) to $\sim 0.53 \pm 0.01$ at 45 ºC where it levels off with increasing temperature (Figure 3.4, A). Because K_D and DLS-measured mean

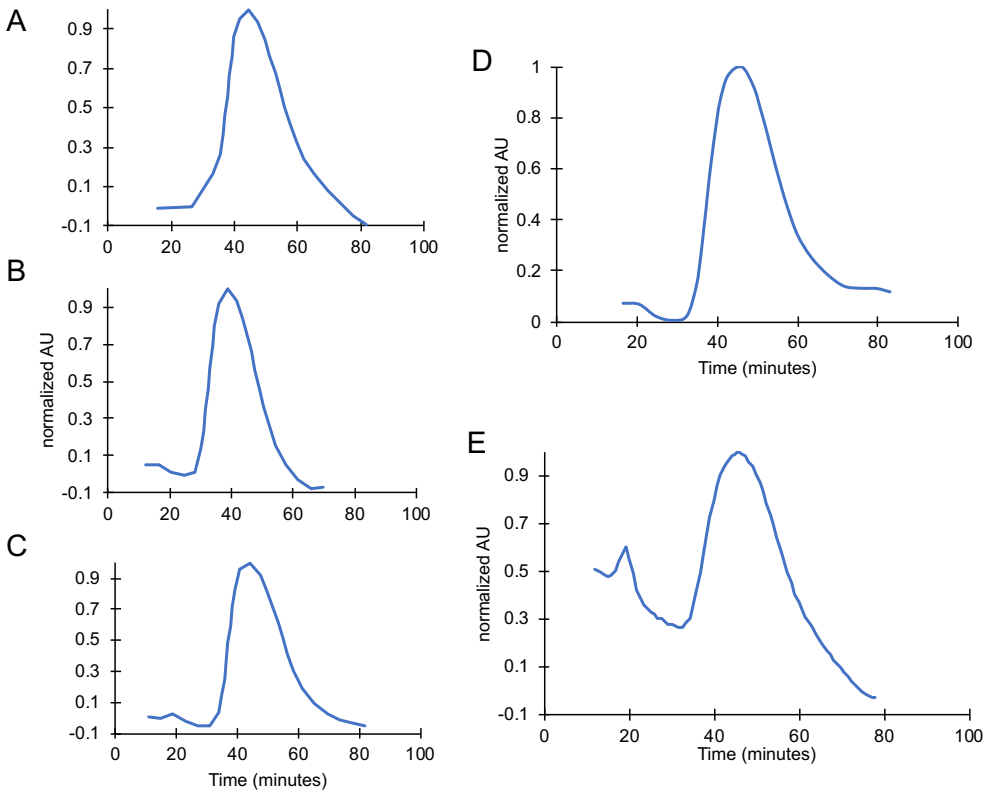

Figure 3.3. <u>Retro B domain solubility and stability overtime.</u> After initial 2 hour centrifugation, Retro B domain was analyzed by SEC at 23 ºC at A) 0 hours, B) 2 hours, C) 4 hours, D) 6 hours, and E) 24 hours after centrifugation

R_h showed some agreement and are correlated linearly when compared at common temperatures (Figure 3.4, B), this implies that, consistent with DLS measurements, Retro B domain decreases in hydrodynamic size as temperature increases from 25 ºC to 65 ºC, supporting the idea of a heat-induced compaction. From 45 to 65 ºC and 25 to 8 ºC, the K_D changes minimally. This supports that Retro B domain mean hydrodynamic size depends on temperature and exhibits a sigmoidal shape in the K_D dependence on temperature, which is again in good agreement with DLS measured changes in hydrodynamic size shown in the previous section.

SEC-estimated mean R_h was determined using the linear correlation of mean R_h and as established by globular proteins at each temperature. The SEC estimated values

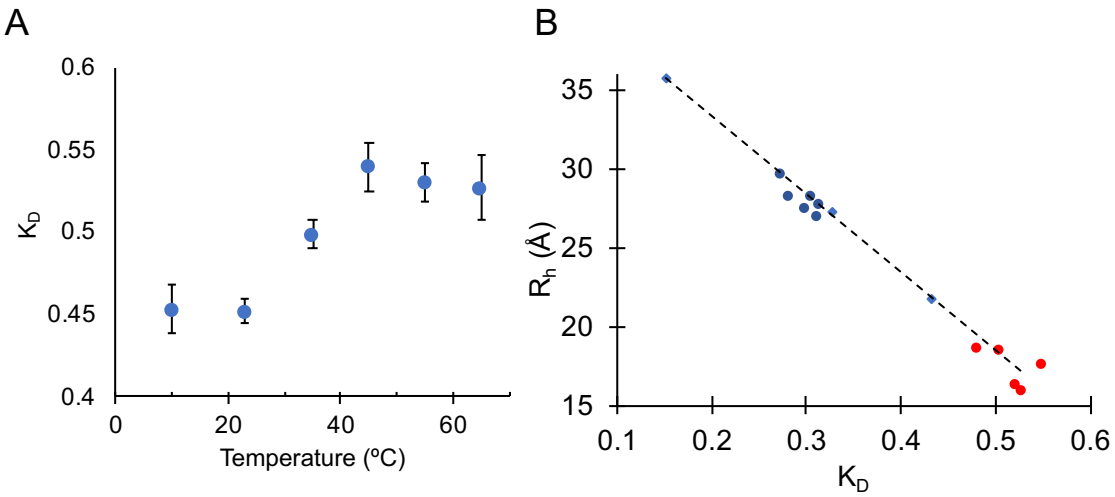

Figure 3.4. <u>Retro B domain mean K_D measured by SEC</u>. A) Temperature dependence of Retro B domain K_D, B) Linear comparison of DLS mean R_h and K_D for duplicates of folded proteins (dark blue dots), IDP standards (light blue diamonds) and Retro B domain (red) at temperatures from 5 to 65ºC.

for the mean R_h show some good agreement with the DLS results from 5 to 45ºC. From 23 to 8 ºC, SEC-estimated R_h values don't exhibit significant change, from 20.69 Å (Figure 3.6, B) to 20.8 Å (Figure 3.5, B), respectively, which is consistent with DLS measurements (Table 3.1). Similarly, SEC predicted R_h at 35 (Figure 3.6, B) and 45 ºC (Figure 3.7, B) are similar in trend and overall change as DLS-measured mean R_h, with differences in the magnitude of change at 35 ºC (Table 3.1).

At temperatures above 45 ºC, myoglobin, carbonic anhydrase, and albumin were unable to be used due to each sustaining structural effects from high heat, such as oligomerization or aggregation. As a result, disordered proteins for which the R_h has been approximated previously were used to estimate Retro B domain R_h at 55 and 65ºC.[47, 53] However, because these standards were far larger than Retro B domain and looking at the linear regression that was used to predict the size, we see that the 3 proteins used are of similar K_D and predicted R_h. Thus, small differences in the measured K_Ds

can, and do, change the linear regression significantly which in turn affects the predicted R_h of Retro B domain. This can be seen in Table 3.7 at 65 ºC where the average K_D is similar to the one measured at 55 ºC (Table 3.6) but the predicted R_h is smaller and with greater deviation. This may account for the difference in magnitude of size changes between SEC and DLS at 55 and 65 ºC (Table 3.1).

when examining the relationship of molecular weight based on primary sequence with K_D, Retro B domain shows separate linear correlations than the folded protein standards and trends with IDPs p53(1-93) and Retro nuclease, seen in the A panels of Figures 3.5-3.10. Overall, when comparing SEC measurements to DLS measurements, it can be stated that the hydrodynamic size of Retro B domain follows the trend of temperature sensitivity that was observed in Retro nuclease, larger at low temperatures and smaller at high temperatures.

Table 3.1. Comparing R_h measured by SEC and DLS. Summary of hydrodynamic measurements at each temperature. R_h measurements most similar from 5-45 ºC when folded protein standards used and deviate much more at high temperatures when IDPs used to predict structure. *SEC temperature was 8ºC, DLS was 5 ºC

Temperature (ºC)	Average K_D	K_D standard deviation	SEC predicted R_h (Å)	DLS measured R_h (Å)
5*	0.453	0.015	20.98 ± 0.8	20.87 ± 0.8
25	0.452	0.008	20.69 ± 0.4	19.96 ± 0.6
35	0.499	0.009	18.63 ± 0.1	19.27 ± 0.6
45	0.540	0.014	17.16 ± 0.4	17.57 ± 0.4
55	0.530	0.012	17.86 ± 0.5	16.27 ± 0.5
65	0.527	0.020	17.44 ± 0.9	15.90 ± 0.6

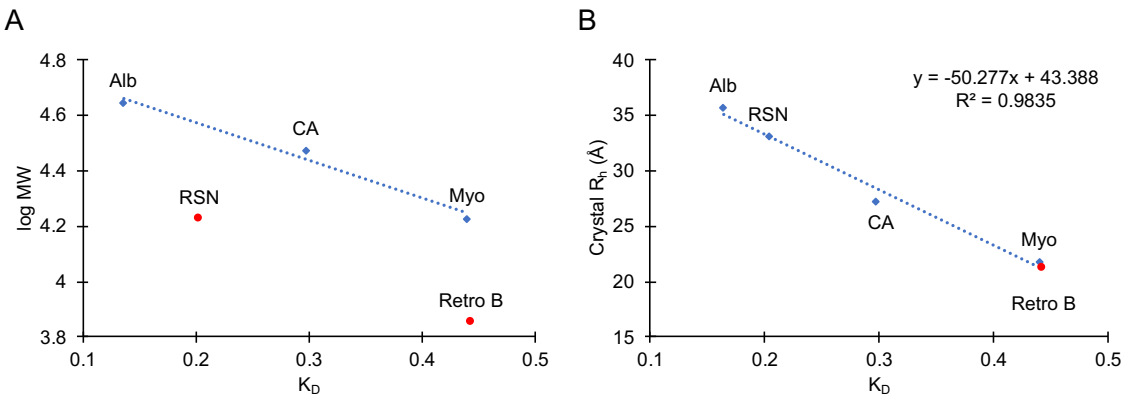

Figure 3.5. **SEC Results at 8 ºC.** Blue diamonds are folded protein standards. Retro B and Retro nuclease are in red. Log M.W. based on primary sequence vs. K_D shows correlation trends for both folded and disordered proteins. Retro B domain R_h was extrapolated from folded protein crystal structures and the folded protein regression line.

Table 3.2 SEC Results at 8ºC

Protein	Average K_D	K_D standard deviation	Molecular Weight (Da)	R_h (Å)
Chicken Egg Albumin	0.136	0.040	44,287	35.76
Bovine Carbonic Anhydrase	0.297	0.015	29,844	27.34
Horse heart myoglobin	0.440	0.006	16,950	21.83
Retro Staphylococcal nuclease	0.203	0.002	16,868	33.16
Retro B domain	0.453	0.015	7,100	20.98

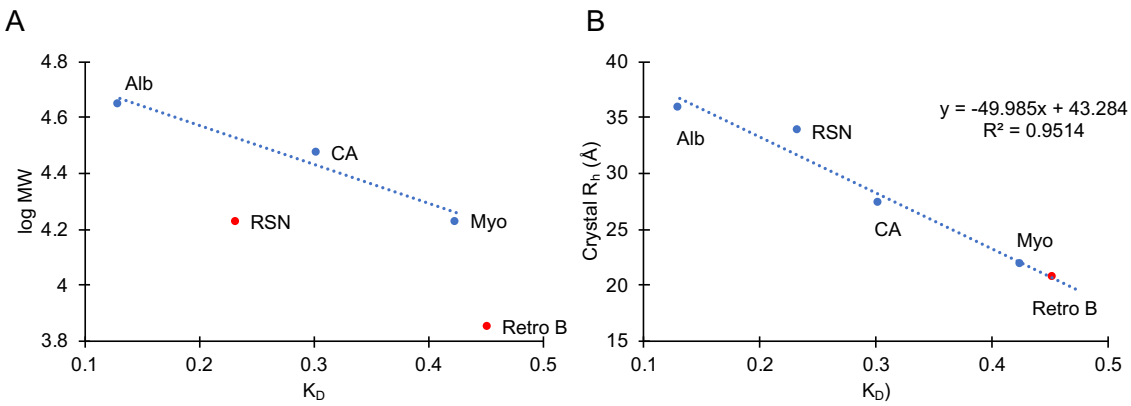

Figure 3.6. **SEC Results at Room Temperature (23 ºC).** Blue diamonds are folded protein standards. Retro B and Retro nuclease are in red. Log M.W. based on primary sequence vs. K_D shows correlation trends for both folded and disordered proteins. Retro B domain R_h was extrapolated from folded protein crystal structures and the folded protein regression line.

Table 3.3 SEC Results at Room Temperature (23 ºC)

Protein	Average K_D	K_D standard deviation	Molecular Weight (Da)	R_h (Å)
Chicken Egg Albumin	0.130	0.005	44,287	35.76
Bovine Carbonic Anhydrase	0.303	0.009	29,844	27.34
Horse heart myoglobin	0.425	0.003	16,950	21.83
Retro Staphylococcal nuclease	0.233	0.009	16,868	33.7
Retro B domain	0.452	0.008	7,100	20.69

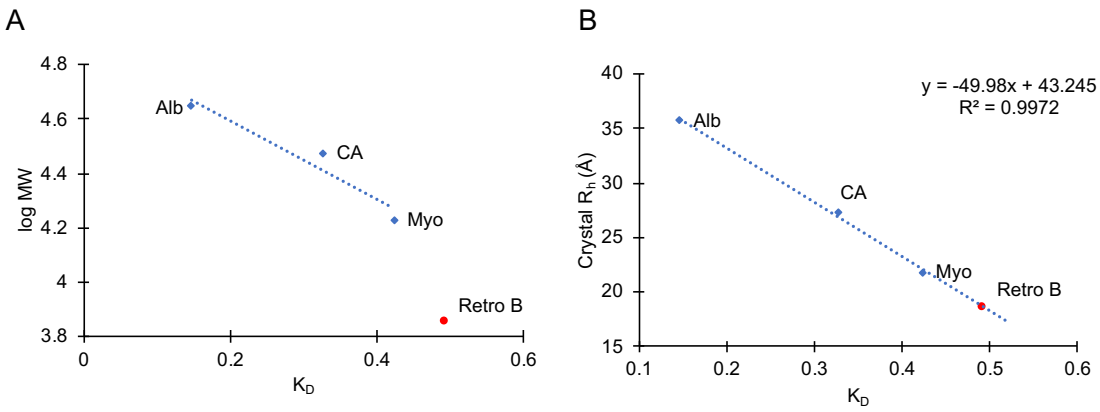

Figure 3.7. SEC Results at 35 ºC. Blue diamonds are folded protein standards. Retro B and Retro nuclease are in red. Log M.W. based on primary sequence vs. K_D shows correlation trends for both folded and disordered proteins. Retro B domain R_h was extrapolated from folded protein crystal structures and the folded protein regression line.

Table 3.4 SEC Results at 35 ºC

Protein	Average K_D	K_D standard deviation	Molecular Weight (Da)	R_h (Å)
Chicken Egg Albumin	0.147	0.004	44,287	35.763
Bovine Carbonic Anhydrase	0.327	0.017	29,844	27.339
Horse heart myoglobin	0.423	0.009	16,950	21.83
Retro B domain	0.493	0.017	7,100	18.63

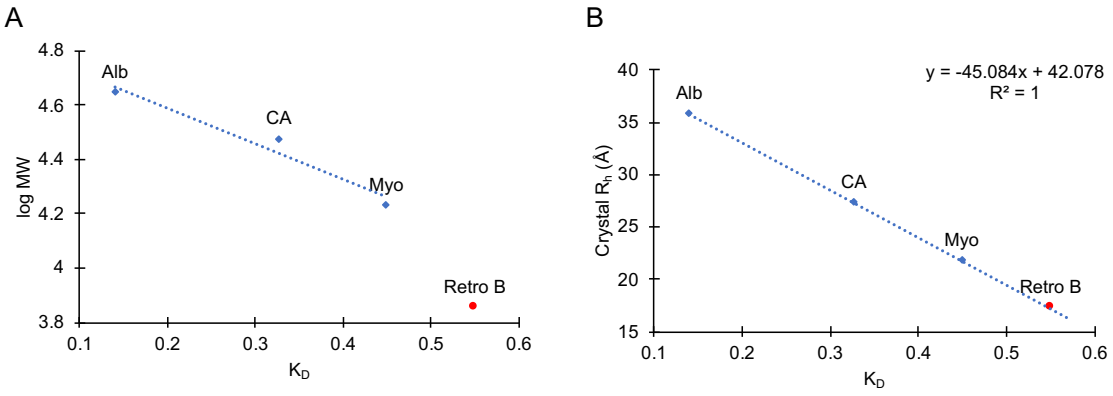

Figure 3.8. **SEC Results at 45 ºC.** Blue diamonds are folded protein standards. Retro B and Retro nuclease are in red. Log M.W. based on primary sequence vs. K_D shows correlation trends for both folded and disordered proteins. Retro B domain R_h was extrapolated from folded protein crystal structures and the folded protein regression line.

Table 3.5 SEC Results at 45 ºC

Protein	Average K_D	K_D standard deviation	Molecular Weight (Da)	R_h (Å)
Chicken Egg Albumin	0.151	0.017	44,287	35.763
Bovine Carbonic Anhydrase	0.326	0.006	29,844	27.339
Horse heart myoglobin	0.448	0.009	16,950	21.83
Retro B domain	0.539	0.014	7,100	17.28

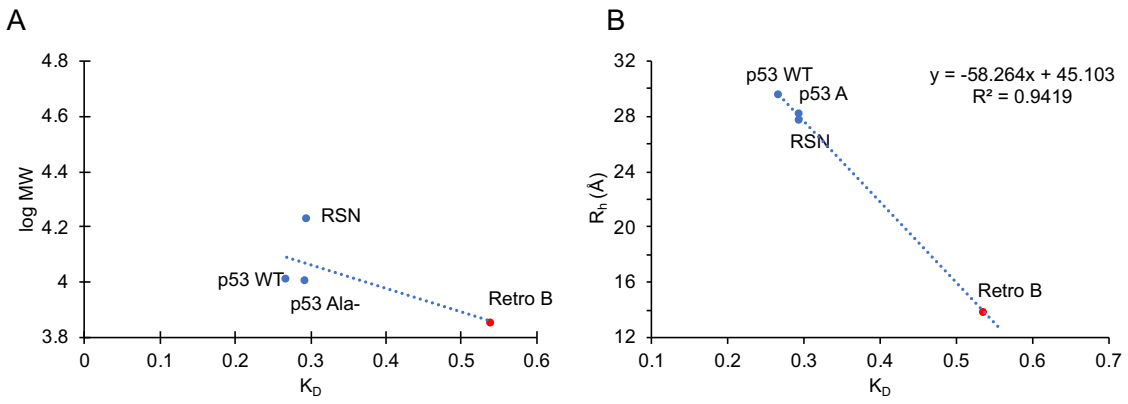

Figure 3.9. **SEC Results at 55 ºC.** Blue diamonds are folded protein standards. Retro B and Retro nuclease are in red. Log M.W. based on primary sequence vs. K_D shows correlation trends for both folded and disordered proteins. Retro B domain R_h was extrapolated from folded protein crystal structures and the folded protein regression line.

Table 3.6 SEC Results at 55 ºC

Protein	Average K_D	K_D standard deviation	Molecular Weight (Da)	R_h (Å)
Retro Staphylococcal nuclease	0.296	0.025	16,868	27.649
p53(1-93) WT	0.268	0.003	10,123	29.478
p53(1-93) Ala-	0.295	0.005	9,955	28.152
Retro B domain	0.530	0.012	7,100	15.98

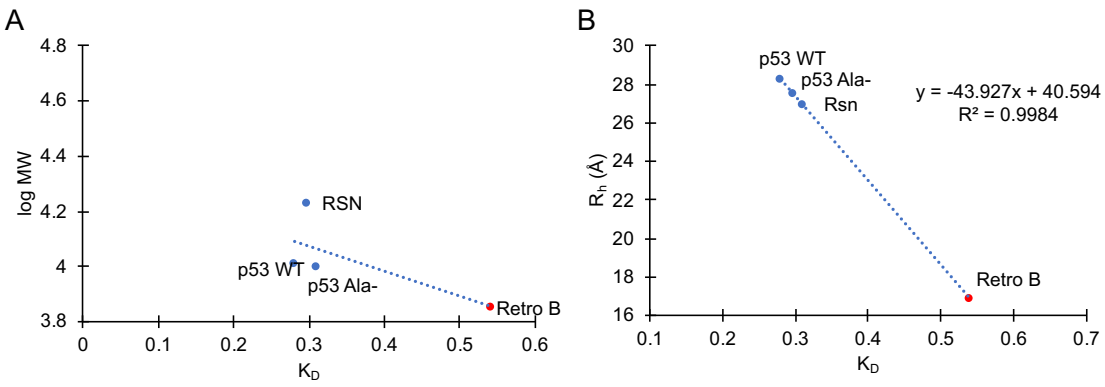

Figure 3.10. **SEC Results at 65 ºC.** Blue diamonds are folded protein standards. Retro B and Retro nuclease are in red. Log M.W. based on primary sequence vs. K_D shows correlation trends for both folded and disordered proteins. Retro B domain R_h was extrapolated from folded protein crystal structures and the folded protein regression line.

Table 3.7 SEC Results at 65 ºC

Protein	Average K_D	K_D standard deviation	Molecular Weight (Da)	R_h (Å)
Retro Staphylococcal nuclease	0.299	0.003	16,868	27.474
p53(1-93) WT	0.298	0.023	10,123	28.239
p53(1-93) Ala-	0.312	0.019	9,955	26.864
Retro B domain	0.527	0.020	7,100	17.44

3.3. Discussion

The protein coil library is a widely used denatured state model built from the segments of irregular structure found in the Protein Data Bank and has been shown to reproduce the intrinsic conformational preferences of the amino acids for helix, sheet, and PPII[33], as well as the effects on the conformational preferences from neighboring residues.[46, 59] These coil libraries are also able to accurately describe structural propensities and NMR parameters for denatured proteins within a reasonable degree. Interestingly, while survey of the PDB shows that PPII is only weakly represented in folded structures, when constructed from the protein coil library, the two most populated regions of the Ramachandran map are the α and PPII basins.[33, 34, 46] While this supports our hypothesis that PPII is the dominant backbone conformation in the protein denatured state[31], the extent to which α and PPII populations in the denatured state change in response to changes in temperature is not yet fully understood. Even less understood is how subtle conformational biases in the denatured state are utilized to facilitate folding and function or the role of temperature in describing coil structure. However, we have seen that heat does indeed modulate coil populations. This is evidenced by the large temperature-dependent changes in hydrodynamic size exhibited by IDPs[20-22, 47, 53] as well as by other temperature dependent characteristics of proteins including their physical abilities to fold[60], phase separate[61], and recognize binding partners.[62]

Here, it was shown that Retro B domain exhibited structural character and trends that were similar to those measured in Retro nuclease. Like Retro nuclease, Retro B domain has the sequence composition of foldable protein. This means that the sequence

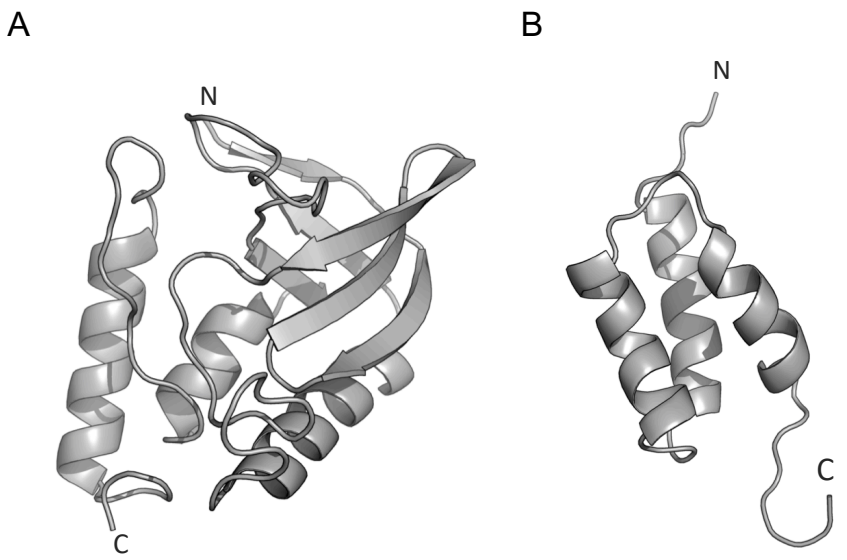

Figure 3.11. <u>Native structures.</u> The native structures of A) nuclease and B) the B domain of protein A, both from *Staphylococcal aureus*. Nuclease is composed of α helices and β sheets and the Retro demonstrated PPII and α conformational propensities. The B domain is made of only α helices and will be tested similar to Retro to determine what, if any, conformational propensities it has. Cartoon structures were rendered with the program PyMol, using the protein data bank files 1STN and 1BDC for nuclease and the B domain, respectively.

is compositionally similar to proteins which adopt a folded structure, containing around 40% hydrophobic residues. This is in contrast to IDPs which tend to have low sequence complexity, low hydrophobic group content, high amounts of disorder promoting residues, and high net charge.[18] Even though the hydrophobic residues that form a stable core in the WT are also present, Retro B domain was shown to be disordered by CD spectroscopy. Both Retro B domain and Retro nuclease were shown to trend with IDPs when examining (log MW vs K_D) (Figure 3.5, A). Both proteins exhibited significant changes in SEC and DLS measured hydrodynamic size that were dependent on temperature. Similarly, these temperature-dependent changes in hydrodynamic size track with temperature induced maxima and minima formation in the CD difference spectra that are associated with transient sampling of PPII and α, respectively.

The main differences between Retro B domain and Retro nuclease are the magnitude of the changes observed. Retro B domain is a small protein composed of a three-helix bundle (Figure 3.11, B) and only 63 amino acids, whereas Retro nuclease is composed of both α helices and β sheets (Figure 3.11, A), is 150 amino acids in size, and had greater magnitude of change. Retro B domain, chosen for its previous structural examination that demonstrated its ability to be expressed and analyzed[63], was not as soluble, limiting concentration studies, and didn't express as well as Retro nuclease did, typically yielding 2-3 mg of protein per one-liter prep.

3.4. Conclusions

To better understand the structural preferences of the denatured state, the structural character and temperature dependence to mean hydrodynamic size of Retro B domain was analyzed. Our data seem to show that the structural character of Retro B domain, presented in this chapter, are consistent with those found in the study of Retro nuclease. Both denatured state ensembles had sigmoid-shaped temperature dependent hydrodynamic size changes that tracked with changes in backbone structure, reported by CD spectroscopy.

The data also support that proteins rich in hydrophobic content exhibit heat sensitive hydrodynamic size that relate to heat effects on backbone structure. This shows that this property is not unique to *S. aureus* nuclease and thus a general result that could be widely applicable to other proteins unfolded or disordered states. Moreover, it is shown that Retro B domain may be PPII- dominant temperatures > 25 ºC with a bias for the α basin of the Ramachandran map at cold temperatures. This indication in the CD difference spectra is very similar to the one observed in Retro nuclease and may support

the long-argued claim[31] that PPII is indeed the dominant conformation in the denatured state.

More broadly, Retro B domain results demonstrate that we continue to observe marginal structural stability of unfolded proteins under biological conditions. Consequently, the biological processes that are facilitated by protein macromolecules are indirectly, if not directly, reliant on the properties and energetic character of the denatured state.

4. SCRAMBLED NUCLEASE RESULTS AND DISCUSSION

4.1. Introduction

Experiments demonstrating that denatured protein in 6M guanidinium conform to the expected size of a random coil, as Shortle notes[19], fail to prove that all chains in all conditions, including water, are random coils. The unfolded state has been shown both theoretically[3, 20-22] and experimentally[21, 47, 53] to be conformationally-biased. This was first tested in short peptides[32-39], an adequate but incomplete model of the denatured state in that, like an unfolded protein, they don't form globular structures and yet they are incapable of making the long-range interactions that are characteristic of full-length proteins. Similarly, IDPs have been shown to have sequence dependent influences on structural size[46] but their compositional differences make their application to folded proteins difficult. Structural characterization of Retro nuclease and Retro B domain support the presence of temperature sensitive conformational biases in the denatured state of proteins that are full length and compositionally similar to folded proteins.

What we have observed with Retro nuclease is that, at high temperatures, there is essentially a random coil condition in which there is little bias in the sampling of sterically accessible states. As temperature cools from 65 to 25 ºC, a preference for PPII, an extended structure, is observed that increases in intensity, and subsequently hydrodynamic size, with lower temperature. At temperature below 25 ºC, a second sampling preference develops, this time in the α region that results in the hydrodynamic size leveling off. This hypothesis is supported by molecular dynamic (MD) simulation if we are assuming intrinsic conformational propensities based on amino acids in the sequence and not from the formation of stable helices. If this behavior that we have

observed repeatedly in IDPs is truly owing to intrinsic conformational propensities, a few hypotheses can be drawn.

First, this phenomenon should be general and thus found in other proteins. This was shown with Retro B domain in Chapter 3. Additionally, intrinsic propensities that are very locally defined (i.e., at the single residue level) imply that what we observe is largely the result of composition and not organization of the sequence. If true, we should be able to reorganize the sequence and see similar trends with temperature. To test this hypothesis, the sequence of *Staphylococcal* nuclease was randomly scrambled and analyzed by similar hydrodynamic methods. Additionally, by scrambling the sequence, we can test whether Retro nuclease is truly disordered or if it has some stable structure that we did not observe. *Staphylococcal* nuclease was used in these experiments due to promising characteristics of Retro nuclease including the magnitude of its size change, its solubility, and high expression yields, relative to Retro B domain.

The final hypothesis is a different check of the same premise. It follows that if the trend we are seeing is due to locally defined structures, then residual, global structure in the unfolded state would yield a distinct temperature dependence of the mean R_h. To test this, Δ131Δ nuclease (Δ131Δ), a natively unfolded variant of *Staphylococcal* nuclease that has nine residues deleted from both the C- and N-terminal sides, was structurally characterized with similar hydrodynamic methods. Despite being unfolded, NMR of Δ131Δ has shown native turn-like structures as well as transient, but persistent helical structures at residues corresponding to α-helices 1 and 2.[52] Δ131Δ has also been shown to be active upon the addition of active site ligands calcium ion and 3',5'- thymidine diphosphate (TDP)[64], substrate analog inhibitor deoxythymidine 3',5- biophosphate

(pdTP), or single stranded DNA (ssDNA).[19] Further, if the distinct behavior observed in Δ131Δ is the result of global patterns that are maintained in the forward sequence directionality and not the results of the deletion, disruption of this global pattern should result in a temperature dependence to hydrodynamic size that is distinct from Δ131Δ. To test this, we reversed the sequence of Δ131Δ, yielding Retro Δ131Δ, and analyzed its structural character using the same hydrodynamic techniques. These experiments will also test the capability of the techniques we use (i.e, measuring R_h) in identifying and distinguishing between intrinsic and collective properties of a sequence.

The examination of Scrambled nucleases A, B, and C, Δ131Δ, and Retro Δ131Δ aim to further define the role of sequence composition and organization in determining intrinsic conformational propensities in the denatured state.

4.2. Structural Characterization of Scrambled nuclease variants

The data in the following sections was collected by various lab members who are credited at the beginning of each section. Scrambled nuclease A, B, and C were obtained for structural studies by recombinant expression in bacterial cells followed by purification from cell lysate using affinity chromatography. To analyze temperature-induced changes in its structural features, these scrambled proteins were evaluated using CD spectroscopy, SEC, and DLS, when possible. The results show that Scrambled nucleases A, B, and C are marginally soluble disordered monomeric proteins under physiological like conditions. Changes in temperature measured by SEC were observed to cause the same changes in the coil structures for each variant despite varying sequence organization. DLS of the most soluble mutant shows that Scrambled B has the same temperature dependent change in hydrodynamic size as Retro nuclease. The importance

and meaning of these findings are presented, discussed, and analyzed below.

4.2.1. Hydrodynamic Size Analysis of Scrambled nuclease A, B, and C by Size Exclusion Chromatography

This section, for which I organized and was heavily involved in, was a collaborative effort completed with the help of Allison M. Bebo and George L. Parra.

SEC chromatograms from 5 to 65 ºC in 20º increments were measured for each variant and used to evaluate differences in mean hydrodynamic size. Thermodynamic retention factors (K_D) were calculated from elution volumes, with smaller K_D signifying larger mean hydrodynamic size. SEC chromatograms for each variant show that the hydrodynamic sizes of Scrambled A, B, and C are temperature sensitive. From 25 to 5 ºC there is minimal change in the K_D of all variants. From 25 to 65 ºC, the K_D of each variant increases (Table 4.1). Because K_D and DLS-measured mean R_h are correlated linearly (Figure 3.4, B), this implies that each scrambled variant is decreasing in hydrodynamic size at temperatures above 25 ºC, once again supporting the idea of a heat-induced compaction. This confirms that the mean hydrodynamic sizes of Scrambled A, B, and C depend on temperature and exhibit a broad sigmoidal shape in the K_D dependence on temperature, which is in good agreement with DLS measured changes in hydrodynamic size shown in the following section.

The results are of significant importance when compared to each other and with the sequence organization in mind. It is apparent that each scrambled variant shows almost exactly the same temperature-dependent structural behavior as Retro nuclease (Figure 4.1), demonstrating that scrambling the sequence did not have a substantial effect

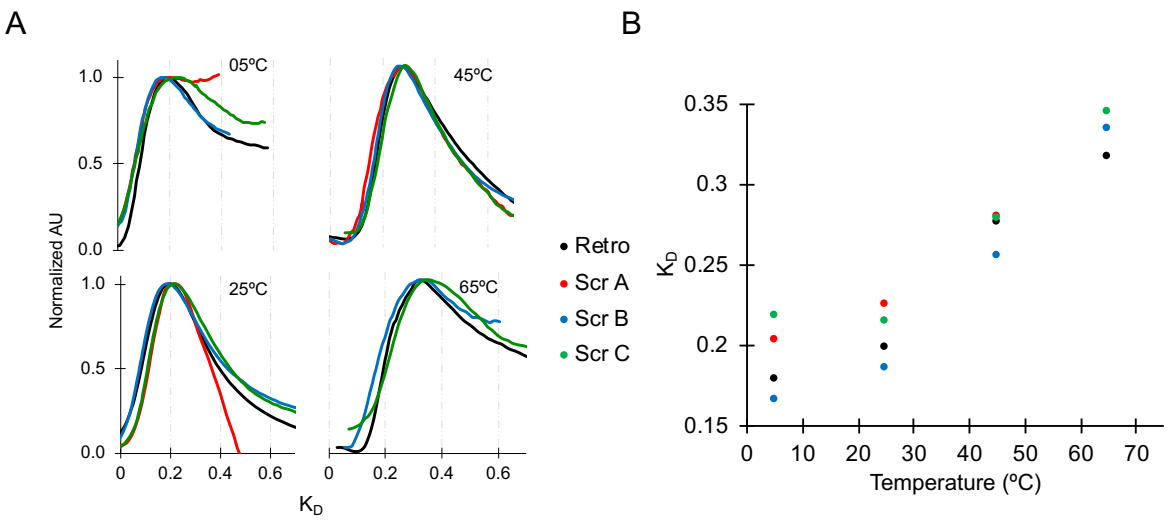

Figure 4.1. <u>Scrambled nuclease hydrodynamic size trend measured by SEC.</u> A) Elution chromatograms of Scrambled A (red), Scrambled B (blue), Scrambled C (green) and Retro nuclease (black) from 5 to 65 ºC. Absorbance was measured at 280 nm and normalized to a maximal value of 1, B) Temperature dependence of Scrambled nucleases A, B, and C and Retro nuclease K_D.

on hydrodynamic size or its temperature dependence. This supports that the conformational biases we are seeing are likely the result of intrinsic propensities that are characteristic of amino acids that the sequence is composed of.

We observed that scrambling the sequence often yielded insoluble protein. The three scrambled variants examined here, Scrambled A, Scrambled B, and Scrambled C, had the best solubility when compared to other variants not reported here. However, the marginal stability of the scrambled variants is still visible in the SEC chromatograms of Scrambled A, B, and C shown in Figure 4.1, A. Here, elution peaks were often distorted and not well defined. These characteristics made this data fairly difficult to obtain as our SEC column would need daily cleaning, to remove degradation or aggregation products, and overnight packing. As mentioned in Section 2.5.6 and in other lab members notes and theses, K_Ds are found to be only interpretable within the context of the same column (without media resuspension). So, this constant state of maintaining the integrity of the

Table 4.1. SEC Results for Retro and Scrambled nucleases

Protein	5ºC Average K_D	25ºC Average K_D	45ºC Average K_D	65ºC Average K_D
Retro nuclease	0.179	0.199	0.276	0.318
Scrambled nuclease A	0.204	0.226	0.280	---
Scrambled nuclease B	0.166	0.186	0.256	0.335
Scrambled nuclease C	0.218	0.215	0.279	0.345

SEC column, long experimentation times to reproducibly test all variants with the same column, and the marginal stability of the proteins themselves presented many challenges. Further, extreme temperatures exacerbated these issues. Overall, these observations suggest that there is more to be uncovered about the relationship between protein solubility, hydration, and the pattern of side chains in the sequence.

4.2.2. Circular Dichroism Spectra of Scrambled A, B, and C

The data in this section was collected by Lance English and Allison Bebo. It is included for accurate analysis of the Scrambled nuclease data.

CD spectroscopy used to examine secondary structure shows that, at 25 ºC, the spectrum for Scrambled A (Figure 4.2), Scrambled B (Figure 4.3), and Scrambled C (Figure 4.4) is similar in character to spectra reported for Retro nuclease[47] and IDPs[49], with molar residue ellipticity (MRE) at 221 nm near zero. The local maximum observed at 221 nm is characteristic of PPII secondary structure which is common in IDPs.

Based on survey of CD spectra from approximately 100 proteins[47], described previously in Section 3.2.1 in which IDPs can be categorized as “mostly unfolded” or are

consistent with the "existence of some residual secondary structure" based on MRE values (10^{-3} degree cm^2 $dmol^{-1}$ $residue^{-1}$) at 200 and 222 nm, Scrambled A, B, and C CD spectra show signs of some residual secondary structure (Figures 4.2-4.4). To look at the temperature dependence of this disordered state, CD spectroscopy was repeated at a range of temperatures. Figures 4.2-4.4 shows that the Scrambled nuclease spectra have subtle, but minimal changes across the temperature range. Once again, the changes in character from temperature are not easily characterized by traditional deconvolution of the CD spectra into specific structural elements as satisfactory reference sets for disordered proteins are limited, owing to the spectral similarity of denatured structures.[56] For the purposes of this project, the CD difference spectra were not examined and described.

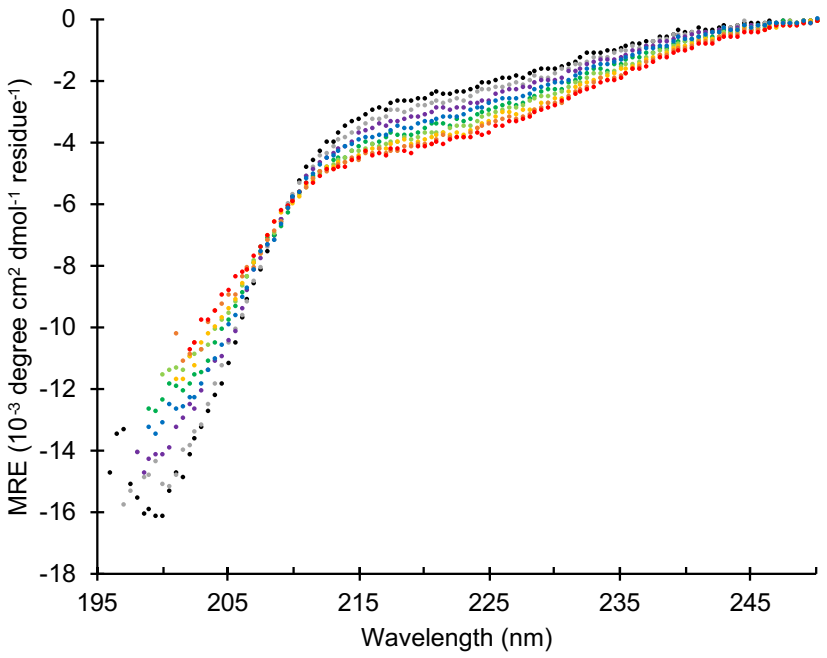

Figure 4.2 **CD spectra of Scrambled A.** The spectra for Scrambled A is shown from 5 to 85ºC in 10º increments; 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (dark green), 55°C (light green), 65°C (yellow), 75°C (orange), 85°C (red). CD measurement done by Lance English.

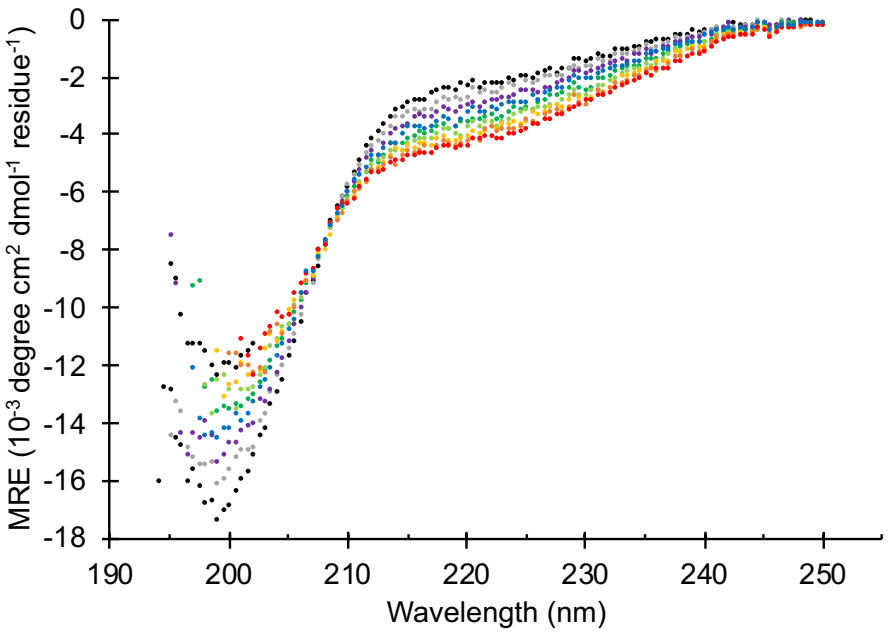

Figure 4.3. <u>CD spectra of Scrambled B.</u> The spectra for Scrambled B is shown from 5 to 85ºC in 10º increments; 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (dark green), 55°C (light green), 65°C (yellow), 75°C (orange), 85°C (red). CD done by Lance English.

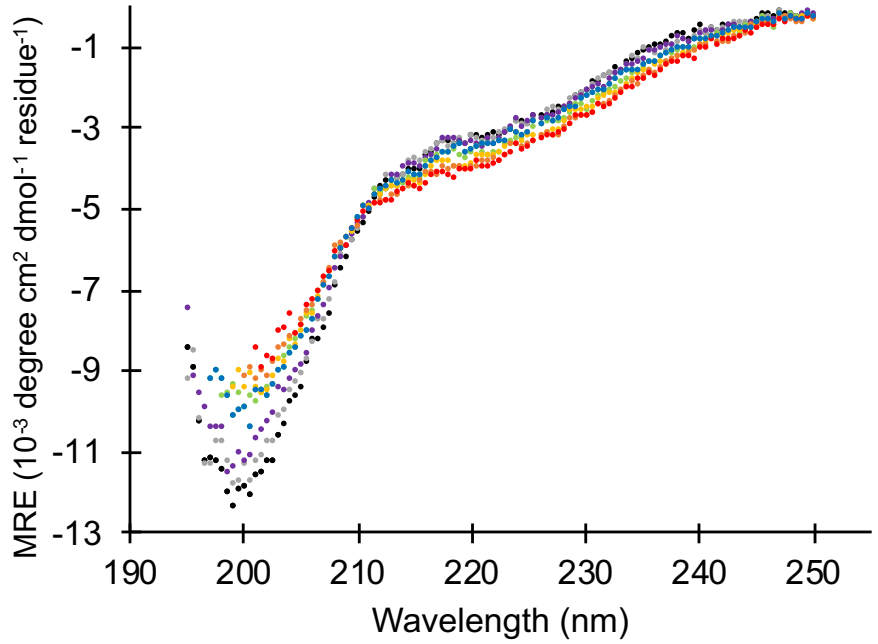

Figure 4.4. <u>CD spectra of Scrambled C.</u> The spectra for Scrambled C is shown from 5 to 85ºC in 10º increments; 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (dark green), 55°C (light green), 65°C (yellow), 75°C (orange), 85°C (red). CD done by Allie Bebo.

4.2.3. Characterization of Scrambled nuclease B Structure by Dynamic Light Scattering

The data in this section was collected by Lance English and George Parra. It is included for accurate analysis of SEC described in Section 4.2.1.

Dynamic light scattering, performed from 5 to 65 ºC, was used to measure the mean R_h of the most soluble variant, Scrambled B. Figure 4.5 gives the temperature dependence of Scrambled B mean R_h showing gradual shifts with temperature. Temperature-induced changes in mean R_h were fully reversible and show a near identical trend in Scrambled B as Retro nuclease[47], again exhibiting a sigmoid dependence with temperature. Temperatures greater than 25ºC caused compaction, reducing the mean R_h from 34.0 Å at 25ºC to 27.6 Å at 65ºC. At temperatures less than 25ºC, there is a decrease in size that levels off at 32.3 at 5ºC. At all temperatures, error bars indicate the standard deviation from the mean from repeat measurements. These values are in good

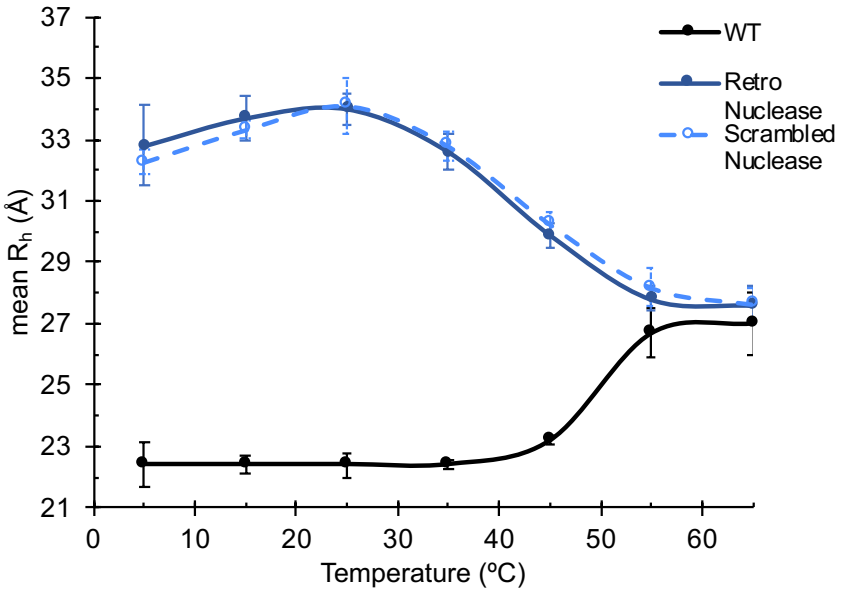

Figure 4.5. <u>Temperature dependence of Scrambled B mean R_h</u>. Scrambled B (light blue, open circles) R_h is temperature sensitive and follows the same trend as Retro nuclease (blue, closed circles). WT nuclease (black) is shown for reference. All lines have no physical meaning and are provided to highlight the temperature trends. The error is the standard deviation from repeat measurements.

agreement with the SEC data presented in Section 4.2.1. with the exception of SEC reporting a slight increase in mean R_h (decrease in K_D), rather than a slight decrease from 25 to 5ºC (Figure 4.1, B). Interestingly, this was also observed in Retro nuclease here, and when examined previously.[47] Overall, the changes in mean R_h from temperature effects on Scrambled B structure are highly similar to those of Retro nuclease.

4.3. Structural Characterization of Δ131Δ nuclease

To compare temperature-induced changes in local structure to those in global structure, Δ131Δ and Retro Δ131Δ nuclease variants were evaluated using CD spectroscopy and DLS. The data in this section was collected by Lance R. English and Allison M. Bebo and is included with the purpose of better interpreting not only our scrambled nuclease data but also the impact of amino acid propensities and sequence organization. As such, the findings from structural examination of Δ131Δ and Retro Δ131Δ are discussed very briefly below.

CD spectroscopy of Δ131Δ (performed by Allison Bebo) shows that, at 25ºC, the Δ131Δ spectrum is similar in character to spectra reported for Retro nuclease[47] and IDPs[49], with molar residue ellipticity (MRE) at 225 nm near zero (Figure 4.6) with minimal changes to its spectrum with temperature. For the purposes of this project, the CD difference spectra was not examined and described.

DLS, performed from 5 to 65 ºC, was used to measure the mean R_h of Δ131Δ. Figure 4.7 gives the temperature dependence of Δ131Δ mean R_h showing gradual shifts with temperature. Temperature-induced changes in mean R_h were fully reversible and show a distinct trend in Δ131Δ, exhibiting a dependence with temperature that is not the same sigmoid shape we observed thus far in Retro nuclease, Scrambled nucleases, and

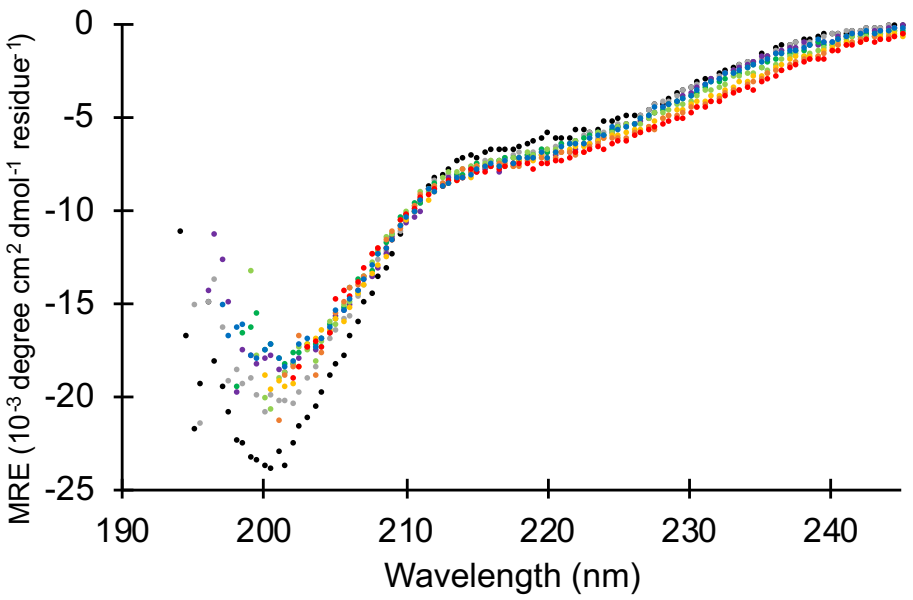

Figure 4.6. <u>CD spectra of Δ131Δ.</u> The spectra for Δ131Δ is shown from 5 to 85ºC in 10º increments; 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (dark green), 55°C (light green), 65°C (yellow), 75°C (orange), 85°C (red). CD measurements done by Allie Bebo.

Retro B domain (Figure 4.7). Temperatures greater than 25 ºC caused compaction, reducing the mean R_h from 32.2±0.5 Å at 25 ºC to 28.1±1.3 Å at 65 ºC, still supporting the presence of heat induced compaction. At temperatures less than 25 ºC, the hydrodynamic size increases to 33.4 Å at 15 ºC and to 35.0 Å at 5 ºC.

To confirm that the changes we observed were the result of residual global characteristics and not from the deletions in Δ131Δ, the reverse sequence of Δ131Δ (Retro Δ131Δ) was analyzed by DLS. DLS measurements of Retro Δ131Δ taken from 5 to 65 ºC trend (Figure 4.7) exhibit a dependence with temperature that is similar to Retro nuclease and Scrambled nucleases, only shifted down due to its smaller size, the result of the 18-residue deletion. This supports that when residual structure is not present, intrinsic propensities are dominant. Temperatures greater than 25 ºC caused compaction, reducing the mean R_h from 33.3±0.3 Å at 25 ºC to 26.7±0.8 Å at 65 ºC. At temperatures

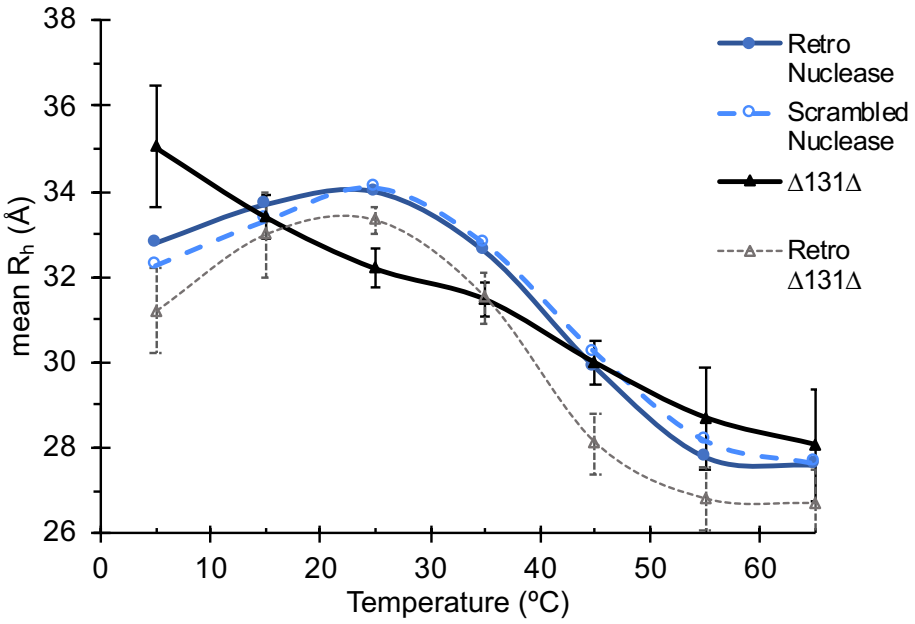

Figure 4.7. <u>Temperature dependence of Δ131Δ and Retro Δ131Δ mean R_h.</u> Δ131Δ has been shown by NMR to have residual structure[52] and shows a mean R_h trend with temperature (black triangles) that is different than Scrambled (blue open circles) and Retro Nuclease (blue closed circles). Reversal of the Δ131Δ sequence results in a temperature dependent trend (grey open triangles) that is like Retro and Scrambled nuclease. DLS measurements done by Lance English.

less than 25 ºC, the hydrodynamic size decreases to 33.0 Å at 15 ºC and to 31.2 Å at 5 ºC.

4.4. Discussion

In the temperature range that CD spectroscopy suggests PPII is the dominant conformation (i.e., T >25 °C), the measured change in mean R_h for Retro nuclease compares favorably to the sequence-predicted change when computed using either of the values for the enthalpy of the PPII to non-PPII transition ΔH_{PPII} measured in peptides, ~10 or ~13 kcal mol^{-1}.[46] This shows that Retro-nuclease's large ~7 Å reduction in mean R_h owing to the 25-65 °C temperature change is in good quantitative agreement with experimental PPII propensities for the thermodynamic energies of the PPII to non-PPII transition, ΔH_{PPII}, and ΔS_{PPII}.[46] Here, it was shown that Scrambled nuclease variants had structural character and trends that were nearly identical to those measured in Retro

nuclease. This includes CD spectroscopy results showing the Scrambled variants to be disordered and SEC and DLS measurements that reported on hydrodynamic size. Like Retro nuclease, Scrambled nucleases A, B, and C had the sequence composition of a folded protein. This means that the sequence is compositionally similar to proteins which adopt a folded structure, as opposed to having low sequence complexity, low hydrophobic group content, and high net charge like an IDP.[49] Following these observations, the data support that the PPII bias present in Retro nuclease is likely also present in Scrambled nuclease, thus supporting that the PPII bias is locally determined and mostly insensitive to organizational details of the primary sequence.

Moreover, scrambling the sequence did not seem to affect the temperature dependence of the mean R_h. CD spectroscopy showed all these scrambled variants to be disordered and DLS showed the same temperature dependent trend in Scrambled nuclease B as in Retro nuclease. This supports that Retro nuclease is indeed truly disordered, as scrambling of the sequence would have disrupted any stable structure it may have had that we failed to detect. The main differences between Scrambled nuclease and Retro nuclease were the affects to solubility and stability. If the sequence was scrambled, rather than reversed, we typically obtained insoluble protein. Some scrambled nucleases were marginally soluble and those were the ones we analyzed with SEC, but only one was able to be measured with DLS. All three scrambled nuclease proteins examined exhibited similar changes, relative to Retro nuclease, in SEC measured hydrodynamic size that were dependent on temperature This supports that the conformational propensities we are measuring are substantial and intrinsic to amino acids in the sequence.

Δ131Δ is an important test of our hypothesis about locally determined intrinsic propensities. As discussed, Δ131Δ is a variant of *S. aureus* nuclease with 18 total deletions. This means that it is compositionally similar to the variants of *S. aureus* nuclease we have examined, allowing a more direct comparison between all nuclease variants and the intrinsic properties of the sequence. Further, Δ131Δ has been shown to have residual structure in the form of transient sampling of helical and turn structures despite being an unfolded variant.[52] We wondered if examination of Δ131Δ using the same methods would be able to distinguish between Δ131Δ's global organization and the local structure we believe we are identifying in the Retro and Scrambled proteins. This is indeed what was observed with DLS (Figure 4.7). Retro Δ131Δ, having lost its global sequence organization, has a temperature dependent behavior that is much more similar to what we observed in Retro and Scrambled nuclease, only shifted down a few angstroms due to the residue deletions and subsequent smaller size. This supports, once again, that intrinsic conformational propensities are substantial and only able to be overcome with a network of interactions. This collective property of the sequence is able to form a specific network of interactions, much like the folding reaction would. How structure transitions from intrinsic behavior to collective behavior is the essence of the folding reaction and we hypothesize that the organizational principals of the denatured state, and thus how a protein folds, can be revealed from studying the collective properties of a sequence.

4.5. Conclusions

The data seem to indicate that small changes in bias can yield large changes in mean hydrodynamic size and that there are two distinctly different types of DSE

structure: intrinsic and collective. Both depend on the protein sequence, but intrinsic depends only on the composition, demonstrated by our ability to scramble the sequence and retain near identical structural character. Collective denatured states depend on both the composition and the specific arrangement of amino acids, demonstrated by Δ131Δ's unique temperature dependence to mean R_h that seemingly reverts back to its intrinsic propensities upon disruption of its global organization via reversal of its primary sequence. In other words, locally defined intrinsic features are hard to overcome unless you have a fairly well-developed networks of interactions. Moreover, it is unsurprising that these proteins are most similar at high temperatures, where intrinsic propensities are weakest, and less similar at low temperatures, where they seem to strengthen, allowing their effects to be more readily detected.

Protein solubility seems to also depend strongly on sequence organization since the scrambled nuclease proteins varied widely in their solubility. This is important because good solubility is needed so that native folding has minimal competition. While hydrophobic burial is favorable, it can be non-specific, resulting in misfolded states, kinetic traps, or aggregation products.

The data we obtained raises important questions about the ways that the denatured state ensemble is non-random, how it's pre-organized and biased structurally, and how the transition from intrinsic to collective behavior is important to folding. While what we think we are seeing is a general result (seen in Retro nuclease, its scrambled variants, Retro B domain, and Retro Δ131Δ) that could be widely applicable to other protein's unfolded states, this is not to say that we think all denatured states will behave in the same way for all proteins. However, it is the case that the ensemble for an unfolded

protein that is able to fold behaves differently.

This ability to form networks and allow cooperative folding, which is what we typically see in protein folding, does require a highly specific sequence organization. Thus, the organizational principals of the denatured state and its role in folding and function can be revealed from mechanistic study of the collective properties of the sequence. These details, the correlation of intrinsic conformational propensities to amino acid type and how structure transitions from intrinsic to collective behavior, are what we aim to establish.

APPENDIX SECTION

List of Equipment and Reagent Suppliers

Supplier	Location	Website
A&D Company	Tokyo, Japan	www.aandd.jp
Airgas	Radnor, PA	www.airgas.com
Beckman Coulter	Brea, CA	www.beckmancoulter.com
Bio-Rad Laboratories	Hercules, CA	www.bio-rad.com
Branson Ultrasonics	Danbury, CT	www.sonifier.com
Atum	Menlo Park, CA	www.atum-bio.com
Eppendorf	Hamburg, Germany	www.eppendorf.com
GE Healthcare	Chicago, IL	www.gehealthcare.com
Geoglobal Partners	Palm Beach, FL	www.ggp-us.com
Hellma Analytics	Müllheim, Germany	www.helma-analytics.com
Hirayama Manufacturing	Kasukabe-Shi, Japan	www.hirayama-hmc.co.jp
JASCO Analytical Instruments	Easton, MD	www.jascoinc.com
MilliporeSigma	Burlington, MA	www.emdmillipore.com
Medline Industries	Mundelein, IL	www.medline.com
Pall Life Sciences	New York, NY	www.pall.com
Saint-Gobain	Courbevoie, France	www.saint-gobain.com
Spectrum Laboratories	Rancho Dominguez, CA	www.spectrumlabs.com
Thermo Fisher Scientific	Waltham, MA	www.thermofisher.com
VWR International	Radnor, PA	www.vwr.com
Welch Vacuum	Niles, IL	www.welchvacuum.com

REFERENCES

1. Wright, P. E.; Dyson, H. J., Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* **1999,** *293* (2), 321-31.
2. Mirsky, A. E.; Pauling, L., On the Structure of Native, Denatured, and Coagulated Proteins. *Proc Natl Acad Sci U S A* **1936,** *22* (7), 439-47.
3. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science* **1973,** *181* (4096), 223-30.
4. Watson, H. C., The Stereochemistry of the Protein Myoglobin. Prog. Stereochem: 1969; Vol. 4, p 299.
5. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C., A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **1958,** *181* (4610), 662-6.
6. Blake, C. C.; Koenig, D. F.; Mair, G. A.; North, A. C.; Phillips, D. C.; Sarma, V. R., Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* **1965,** *206* (4986), 757-61.
7. Rose, D. R.; Phipps, J.; Michniewicz, J.; Birnbaum, G. I.; Ahmed, F. R.; Muir, A.; Anderson, W. F.; Narang, S., Crystal structure of T4-lysozyme generated from synthetic coding DNA expressed in Escherichia coli. *Protein Eng* **1988,** *2* (4), 277-82.
8. Eriksson, A. E.; Jones, T. A.; Liljas, A., Refined structure of human carbonic anhydrase II at 2.0 A resolution. *Proteins* **1988,** *4* (4), 274-82.
9. Kisker, C.; Schindelin, H.; Alber, B. E.; Ferry, J. G.; Rees, D. C., A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon Methanosarcina thermophila. *EMBO J* **1996,** *15* (10), 2323-30.
10. Saavedra, H. G.; Wrabl, J. O.; Anderson, J. A.; Li, J.; Hilser, V. J., Dynamic allostery can drive cold adaptation in enzymes. *Nature* **2018,** *558* (7709), 324-328.
11. Dhar, A.; Girdhar, K.; Singh, D.; Gelman, H.; Ebbinghaus, S.; Gruebele, M., Protein stability and folding kinetics in the nucleus and endoplasmic reticulum of eucaryotic cells. *Biophys J* **2011,** *101* (2), 421-30.

12. Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W., Structure-function-folding relationship in a WW domain. *Proc Natl Acad Sci U S A* **2006,** *103* (28), 10648-53.
13. Prakash, S.; Tian, L.; Ratliff, K. S.; Lehotzky, R. E.; Matouschek, A., An unstructured initiation site is required for efficient proteasome-mediated degradation. *Nat Struct Mol Biol* **2004,** *11* (9), 830-7.
14. Huber, D., A selection for mutants that interfere with folding of Escherichia coli thioredoxin-1 in vivo. **2005,** *102* (52), 18872–18877.
15. Tompa, P., Intrinsically unstructured proteins. *Trends Biochem Sci* **2002,** *27* (10), 527-33.
16. Tompa, P., Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* **2012,** *37* (12), 509-16.
17. Liu, J.; Perumal, N. B.; Oldfield, C. J.; Su, E. W.; Uversky, V. N.; Dunker, A. K., Intrinsic disorder in transcription factors. *Biochemistry* **2006,** *45* (22), 6873-88.
18. Tompa, P.; Lemke, E. A., Structure and Function of Intrinsically Disordered Proteins. 2011.
19. Shortle, D., The expanded denatured state: an ensemble of conformations trapped in a locally encoded topological space. *Adv Protein Chem* **2002,** *62*, 1-23.
20. English, L. R.; Tilton, E. C.; Ricard, B. J.; Whitten, S. T., Intrinsic α helix propensities compact hydrodynamic radii in intrinsically disordered proteins. *Proteins* **2017,** *85* (2), 296-311.
21. Langridge, T. D.; Tarver, M. J.; Whitten, S. T., Temperature effects on the hydrodynamic radius of the intrinsically disordered N-terminal region of the p53 protein. *Proteins* **2014,** *82* (4), 668-78.
22. Tomasso, M. E.; Tarver, M. J.; Devarajan, D.; Whitten, S. T., Hydrodynamic Radii of Intrinsically Disordered Proteins Determined from Experimental Polyproline II Propensities. *PLoS Comput Biol* **2016,** *12* (1), e1004686.
23. Tanford, C., Protein denaturation. *Adv Protein Chem* **1968,** *23*, 121-282.
24. Toal, S.; Schweitzer-Stenner, R., Local order in the unfolded state: conformational biases and nearest neighbor interactions. *Biomolecules* **2014,** *4* (3), 725-73.

25. Ramachandarn, G. N.; Ramakrishnan, C.; Sasisekharah, V., Stereochemistry of polypeptide chain configurations. *J Mol Biol* **1963,** *7*, 95-9.
26. Flory, P. J., Statistical Mechanics of Chain Molecules. Wiley & Sons: New York, NY, USA, 1969; pp 30-31.
27. Baldwin, R. L., A new perspective on unfolded proteins. *Adv Protein Chem* **2002,** *62*, 361-7.
28. C, L., How to fold graciously. DeBrunner JTP, Munck E, editors. Mossbauer spectroscopy in biological systems. Allerton House: Monticello, Illinois: University of Illinois Press, **1969**; pp 22-24.
29. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G., Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **1995,** *21* (3), 167-95.
30. Tiffany, M. L.; Krimm, S., Circular dichroism of poly-L-proline in an unordered conformation. *Biopolymers* **1968,** *6* (12), 1767-70.
31. Tiffany, M. L.; Krimm, S., New chain conformations of poly(glutamic acid) and polylysine. *Biopolymers* **1968,** *6* (9), 1379-82.
32. Chen, K.; Liu, Z.; Kallenbach, N. R., The polyproline II conformation in short alanine peptides is noncooperative. *Proc Natl Acad Sci U S A* **2004,** *101* (43), 15352-7.
33. Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F., Helix, sheet, and polyproline II frequencies and strong nearest neighbor effects in a restricted coil library. *Biochemistry* **2005,** *44* (28), 9691-702.
34. Perskie, L. L.; Street, T. O.; Rose, G. D., Structures, basins, and energies: a deconstruction of the Protein Coil Library. *Protein Sci* **2008,** *17* (7), 1151-61.
35. Fitzkee, N. C.; Fleming, P. J.; Rose, G. D., The Protein Coil Library: a structural database of nonhelix, nonstrand fragments derived from the PDB. *Proteins* **2005,** *58* (4), 852-4.
36. Ding, L.; Chen, K.; Santini, P. A.; Shi, Z.; Kallenbach, N. R., The pentapeptide GGAGG has PII conformation. *J Am Chem Soc* **2003,** *125* (27), 8092-3.

37. Shi, Z.; Olson, C. A.; Rose, G. D.; Baldwin, R. L.; Kallenbach, N. R., Polyproline II structure in a sequence of seven alanine residues. *Proc Natl Acad Sci U S A* **2002,** *99* (14), 9190-5.
38. Eker, F.; Cao, X.; Nafie, L.; Schweitzer-Stenner, R., Tripeptides adopt stable structures in water. A combined polarized visible Raman, FTIR, and VCD spectroscopy study. *J Am Chem Soc* **2002,** *124* (48), 14330-41.
39. McColl, I. H.; Blanch, E. W.; Hecht, L.; Kallenbach, N. R.; Barron, L. D., Vibrational Raman optical activity characterization of poly(l-proline) II helix in alanine oligopeptides. *J Am Chem Soc* **2004,** *126* (16), 5076-7.
40. Schweitzer-Stenner, R.; Eker, F.; Griebenow, K.; Cao, X.; Nafie, L. A., The conformation of tetraalanine in water determined by polarized Raman, FT-IR, and VCD spectroscopy. *J Am Chem Soc* **2004,** *126* (9), 2768-76.
41. Asher, S. A.; Mikhonin, A. V.; Bykov, S., UV Raman demonstrates that alpha-helical polyalanine peptides melt to polyproline II conformations. *J Am Chem Soc* **2004,** *126* (27), 8433-40.
42. Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P., Host-guest scale of left-handed polyproline II helix formation. *Proteins* **2003,** *53* (1), 68-75.
43. Shi, Z.; Chen, K.; Liu, Z.; Ng, A.; Bracken, W. C.; Kallenbach, N. R., Polyproline II propensities from GGXGG peptides reveal an anticorrelation with beta-sheet scales. *Proc Natl Acad Sci U S A* **2005,** *102* (50), 17964-8.
44. Elam, W. A.; Schrank, T. P.; Campagnolo, A. J.; Hilser, V. J., Evolutionary conservation of the polyproline II conformation surrounding intrinsically disordered phosphorylation sites. *Protein Sci* **2013,** *22* (4), 405-17.
45. Ferreon, J. C.; Hilser, V. J., The effect of the polyproline II (PPII) conformation on the denatured state entropy. *Protein Sci* **2003,** *12* (3), 447-57.
46. English, L. R.; Voss, S. M.; Tilton, E. C.; Paiz, E. A.; So, S.; Parra, G. L.; Whitten, S. T., Impact of Heat on Coil Hydrodynamic Size Yields the Energetics of Denatured State Conformational Bias. *J Phys Chem B* **2019**.

47. English, L. R.; Tischer, A.; Demeler, A. K.; Demeler, B.; Whitten, S. T., Sequence Reversal Prevents Chain Collapse and Yields Heat-Sensitive Intrinsic Disorder. *Biophys J* **2018,** *115* (2), 328-340.
48. Hynes, T. R.; Fox, R. O., The crystal structure of staphylococcal nuclease refined at 1.7 A resolution. *Proteins* **1991,** *10* (2), 92-105.
49. Uversky, V. N., Natively unfolded proteins: a point where biology waits for physics. *Protein Sci* **2002,** *11* (4), 739-56.
50. Rucker, A. L.; Creamer, T. P., Polyproline II helical structure in protein unfolded states: lysine peptides revisited. *Protein Sci* **2002,** *11* (4), 980-5.
51. Saxena, V. P.; Wetlaufer, D. B., A new basis for interpreting the circular dichroic spectra of proteins. *Proc Natl Acad Sci U S A* **1971,** *68* (5), 969-72.
52. Alexandrescu, A. T.; Abeygunawardana, C.; Shortle, D., Structure and dynamics of a denatured 131-residue fragment of staphylococcal nuclease: a heteronuclear NMR study. *Biochemistry* **1994,** *33* (5), 1063-72.
53. Perez, R. B.; Tischer, A.; Auton, M.; Whitten, S. T., Alanine and proline content modulate global sensitivity to discrete perturbations in disordered proteins. *Proteins* **2014,** *82* (12), 3373-84.
54. Laue, T. M., Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L., Eds. Computer-aided interpretation of analytical sedimentation data for proteins **1992**, 90-125.
55. Dill, K. A.; Shortle, D., Denatured states of proteins. *Annu Rev Biochem* **1991,** *60*, 795-825.
56. Sreerama, N.; Venyaminov, S. Y.; Woody, R. W., Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal Biochem* **2000,** *287* (2), 243-51.
57. Kjaergaard, M.; Nørholm, A. B.; Hendus-Altenburger, R.; Pedersen, S. F.; Poulsen, F. M.; Kragelund, B. B., Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II? *Protein Sci* **2010,** *19* (8), 1555-64.
58. Lorber, B.; Fischer, F.; Bailly, M.; Roy, H.; Kern, D., Protein analysis by dynamic light scattering: methods and techniques for students. *Biochem Mol Biol Educ* **2012,** *40* (6), 372-82.

59. Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M., Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J Mol Biol* **1996,** *255* (3), 494-506.

60. Guo, M.; Xu, Y.; Gruebele, M., Temperature dependence of protein folding kinetics in living cells. *Proc Natl Acad Sci U S A* **2012,** *109* (44), 17863-7.

61. Dignon, G. L.; Zheng, W.; Kim, Y. C.; Mittal, J., Temperature-Controlled Liquid-Liquid Phase Separation of Disordered Proteins. *ACS Cent Sci* **2019,** *5* (5), 821-830.

62. Hamburger, J. B.; Ferreon, J. C.; Whitten, S. T.; Hilser, V. J., Thermodynamic mechanism and consequences of the polyproline II (PII) structural bias in the denatured states of proteins. *Biochemistry* **2004,** *43* (30), 9790-9.

63. Lacroix, E.; Viguera, A. R.; Serrano, L., Reading protein sequences backwards. *Fold Des* **1998,** *3* (2), 79-85.

64. Wang, Y.; Shortle, D., The equilibrium folding pathway of staphylococcal nuclease: identification of the most stable chain-chain interactions by NMR and CD spectroscopy. *Biochemistry* **1995,** *34* (49), 15895-905.