ELECTROSTATIC EFFECTS ON THE STRUCTURE OF INTRINSICALLY DISORDERED PROTEINS

by

Lance R. English, B.S.

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
August 2016

Committee Members:

Steven T. Whitten, Chair

Karen A. Lewis

Sean M. Kerwin

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

Duplication Permission

DEDICATION

To my parents, Kenneth A. English and E. Jill English, and my brother, Clay "Clayton" English. None of this would have been possible without your support.

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Dr. Steven T. Whitten, for giving me the opportunity to work in a flexible lab environment that encourages creativity, discovery, and freedom. I would like to thank my committee members, Dr. Karen A. Lewis and Dr. Sean M. Kerwin, and my undergraduate research advisor, Dr. Robert J. C. McLean, for always being available when I needed assistance with troubleshooting or needed to borrow equipment or reagents. I would also like to thank my lab colleagues Erin C. Tilton and Leona A. Martin. We started and finished this degree together. We shared many days, nights, and weekends together in frustration, contemplation, collaboration, and eventually joy.

Most importantly, I want to thank Dr. Rachell E. Booth. For the past three years, she has been my number one source for troubleshooting, proofreading, idea pitches, advice, inspiration, equipment borrowing, and extracurricular opportunities.

Lastly, I want to thank Dr. Stacie A. Brown and Jimmy D. Gollihar for being the initial sparks that drove me towards a career in research science.

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 substitutions
Alb	Chicken egg albumin
amp^r	Gene that codes for β-lactamase (ampicillin-resistance)
BD-DNP-Asp	Blue dextran-DNP-Aspartate
CA	Bovine carbonic anhydrase
CD	Circular dichroism
DE3	Designation given to cells carrying an IPTG-inducible chromosomal copy of T7 RNA polymerase
DEAE	Diethylaminoethyl
FRET	Förster resonance energy transfer
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 μg/mL ampicillin
MRE	Mean residue ellipticity
MRW	Mean residue weight
Myo	Horse heart myoglobin
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance

OD_{600}	Optical density (absorbance) at 600 nm
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)
pLysS	Plasmid containing genes for chloramphenicol-resistance and T7 lysozyme
PP_{II}	Polyproline II
Pro^-	p53(1-93) with 22 proline to glycine substitutions
PTM	Post-translational modification
R_g	Radius of gyration
R_h	Hydrodynamic radius
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SN	*Staphylococcal* nuclease
SOC	Super optimal broth with catabolite repression
TBS	Tris-buffered saline (20 mM tris, 100 mM NaCl, pH 8.0)
V_0	Void volume
V_E	Elution volume
V_T	Total volume
WT	p53(1-93) wildtype

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
DNA 2.0	Menlo Park, CA	www.dna20.com
EMD Millipore	Billerica, MA	www.emdmillipore.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.hellma-analytics.com
Hirayama Manufacturing	Kasukabe-Shi, Japan	www.hirayama-hmc.co.jp
JASCO Analytical Instruments	Easton, MD	www.jascoinc.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
Sigma-Aldrich	St. Louis, MO	www.sigmaaldrich.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

ABSTRACT

Highly flexible and dynamic proteins with negligible tertiary structure are known as intrinsically disordered proteins (IDPs). Current literature suggests the hydrodynamic dimensions of IDPs are dominated by electrostatic effects. However, other studies indicate polyproline II (PP_{II}) content, a non-classical secondary structure, also greatly impacts the structure of IDPs. To test these suggestions in conjunction, we used the disordered N-terminal region (residues 1-93) of the human tumor suppressor protein p53 as our model and hydrodynamic radius (R_h) as a size reporter. We used size exclusion chromatography to measure the R_h and circular dichroism to measure PP_{II} content at a pH range of 4.5-7.0. Previous research indicates proline and alanine as two of the residues with the highest PP_{II} propensities. Hence, we repeated these measurements with two mutants: one in which all proline residues were replaced with glycines (Pro^-) and one with all alanines replaced with glycines (Ala^-). All three polypeptides have a net charge of -14.99 at pH 7.0 and should be significantly compacted at lower pHs due to a decrease in charge repulsion. As pH was decreased from 7.0 to 4.5, the R_h of the wildtype decreased by 15%, Ala^- by 14%, and Pro^- by 27%. These results show that prolines increase the spacing between charge groups, weakening electrostatic repulsion. Thus, charge effects are important to IDP structure, but are heavily modulated by PP_{II} content.

I. INTRODUCTION

Protein folding has traditionally been viewed as a two-state transition from a random coil to an ordered, functioning protein[1,2]. However, there exists a class of proteins that are critical to all forms of life and transcend this paradigm altogether: intrinsically disordered proteins (IDPs) are biologically active proteins with negligible tertiary structure[3]. IDPs are most common in eukaryotes where they are involved in many functions including signal transduction[4], small-molecule storage[5], post-translational modification (PTM)[6], and regulation of transcription[7], translation[8], and self-assembly[9]. The scope of IDPs extends to non-eukaryotes where recent studies have found disorder in viral proteins[10], protozoal protein-protein interactions[11], and bacterial secretion systems[12]. IDPs are extremely prevalent as well. Current literature estimates that 7-35% of prokaryotic proteins and 15-50% of eukaryotic proteins contain intrinsically disordered regions[8,10].

Due to their wide range of functionality, characterization of IDP structure is crucial. However, their unconventional nature makes it difficult to use classical structural biology tools. X-ray crystallography and nuclear magnetic resonance (NMR) can be used to detect disorder[13], but have high levels of uncertainty when it comes to structure prediction. The high flexibility found in large IDPs inhibits crystal formation[2] and causes too much spectral overlap for high-resolution NMR studies[14]. As a result, alternative approaches such as size exclusion chromatography (SEC)[15] and circular dichroism (CD)[6]

are being used to gain structural insight. SEC is a separatory technique that can also be used quantitatively to determine a molecule's hydrodynamic radius (R_h)[16,17] while CD spectroscopy is commonly used in protein studies to assess secondary structure and intrinsic backbone propensities[18].

1.1 A Common Motif in Biological Systems

Before the discovery of IDPs, a precise and ordered 3D structure was viewed as a necessity for protein function[19]. However, many of the early crystal structures of biologically active proteins contained areas with no apparent electron density[2]. This was reconciled in 1978 when the first small disordered region was discerned in the histone H5 protein by NMR[2]. The use of NMR in an interactome study in 1997 led to the discovery of the first entirely disordered protein[20].

After the discovery of IDPs, researchers turned their attention to the characterization and prediction of IDP primary structures. The idea of order-promoting and disorder-promoting residues was proposed by comparing a database of disordered proteins to one of folded proteins[21]. Researchers then began to predict disorder based on primary sequence with one study successfully forecasting disorder in 95% of large peptides[17].

Disorder in biological systems is not only predictable, but extraordinarily common. Keith Dunker's group constructed a database of 110 disordered regions containing 28 distinct functions. These functions include regulation, nucleic acid-binding, membrane-association, signaling, and PTMs[22,23].

There are many examples of IDPs involved in regulation. A classic example is that of disordered polyanion tail of calsequestrin which regulates sarcoplasmic reticulum

calcium levels[24]. A more recent study identified a disordered region in the inositol pyrophosphate PPIP5K1 that interacts with exocysts to regulate cell migration[25].

Nucleic acid binding is one of the most common functionalities for disordered peptides, generally leading to an induced folding event[26]. An example of disorder in DNA-binding can be seen in the transcription factor alcohol dehydrogenase regulatory gene 1. The disordered zinc-finger region of this protein becomes ordered upon binding to DNA, leading to protein dimerization[27]. An example of disorder in RNA-binding can be seen in the ribosomal protein S15. Upon binding to ribosomal RNA (rRNA), S15 induces the RNA in to the correct tertiary structure for further ribosomal protein binding[28].

Disordered membrane-associating proteins are common in both eukaryotes and prokaryotes. In animal cytoskeletons, Type V intermediate filaments contain both intrinsically disordered head and tail domains which function to support the nuclear membrane[9,29]. An example in the microbial world occurs during fd Phage infection of *Escherichia coli* cells. The helical coat proteins of the fd Phage transition from an ordered to disordered state as a means to penetrate the *E. coli* cell membrane[30].

Cell signaling, generally involving PTMs, are perhaps the most studied function of IDPs. Due to their high level of flexibility and solvent exposure, disordered regions tend to be more accessible targets for PTMs than their folded counterparts[4]. This can be seen

in the disordered region of transcription factor Hypoxia-Inducible Factor 1 in which hydroxylation of a proline residue starts a cascade leading to proteolytic degredation[31] Perhaps the predominant example of cell signaling proteins is the tumor suppressor p53. Often referred to as the guardian of the genome, p53 is a hub for PTMs and coordinates a multitude of signaling pathways. The disordered Transactivation Domain of p53 is only 60 amino acids, but contains an astounding seven phosphorylation sites[32]. Phosphorylation at all of these sites regulates affinity for HDM2, p300, and a multitude of other proteins leading to p53 upregulation, downregulation, degradation, or even cell cycle arrest[32,33].

1.2 Polyproline II content modulates charge effects

Structural biologists have traditionally considered electrostatic interactions as a primary determinant of IDP structure. Because peptides with high charge density can avoid a hydrophobic collapse and obtain an extended conformation, some researchers have attempted to predict disorder based solely on net charge and hydrophobicity[13,34]. However, there exists another component that could be as important as net charge density for determining IDP structure: polyproline II (PP_{II}) content. In fact, many studies indicate that PP_{II} may be a preferred conformation in the disordered and unfolded states of proteins[35,36].

The polyproline II helix (PP_{II}) is a non-classical secondary structure characterized by left-handedness and dihedral angles of -75°, 145°[37]. The structure consists of 3.0 residues per turn with a rise-per-residue of 3.1 Å[37]. It lacks intrachain hydrogen bonding and, due to its extended conformation, the backbone and side chains are exposed to the solvent[13]. A peptide with 100% PP_{II} content would consist entirely of proline residues in the *trans* position however, all amino acids have a certain PP_{II} propensity that contribute to the overall structure[38,39].

Creamer et al. was the first to publish a scale for calculating PP_{II} propensities by performing CD on short peptides with the structure Ac-P_3XP_3GY-NH_2 for each amino acid excluding tryptophan and tyrosine[39]. Since PP_{II} helices have a characteristic local

maxima around 228 nm, Creamer's group calculated PP_{II} propensities for each amino acid using the height of this local maxima for each respective spectrum[39]. They found proline to have the highest PP_{II} propensity (67%) with valine at the other end of the spectrum (49%)[39].

Kallenbach et al. expanded on Creamer's work by adding NMR J-coupling constants to the PP_{II} propensity calculations[38]. They utilized pentapeptides with the structure Ac-GGXGG-NH_2 for each amino acid excluding glycine and proline[38]. Through the use of far-UV CD spectroscopy and NMR, they calculated a scale ranking alanine with the highest PP_{II} content (81.8%) and histidine with the lowest (42.8%)[38].

Hilser et al. contributed to PP_{II} propensity scales by adding calorimetry as a component[40]. The group performed isothermal calorimetry titration on a series of small peptides with the formula Ac-$VP_2XVP_2R_3Y$-NH_2 for each amino acid[40]. PP_{II} propensity was then calculated resulting in proline having the largest PP_{II} content (100%) while glycine has the smallest (13%)[40].

Due to the number of amino acids with high PP_{II} propensities in all three of the aforementioned scales, we hypothesized that PP_{II} content would increase the spacing between charge groups thus weakening and modulating charge effects in intrinsically disordered proteins. To test this, we used the N-terminal region (1-93) of human p53 as a

model peptide. This region is ideal for our studies due to the large proline and alanine content, 22 and 12 residues respectively.

1.3 Project Goals

The goal of this project was to experimentally determine the effects of electrostatic interactions on IDP structure in the presence and absence of high PP_{II} content. CD was used to verify secondary structure and size exclusion chromatography was used to measure R_h for the wildtype protein, a mutant in which all 22 prolines were mutated to glycines (Pro^-), and a mutant in which all 12 alanines were mutated to glycines (Ala^-). Each peptide and each experiment was measured at four pHs: 7.0, 6.0, 5.0, and 4.5.

At pH 7.0, all three of our peptides have a theoretical net charge of 14.99. This drops to 14.83 at pH 6.0, 13.49 at pH 5.0, and 11.10 at pH 4.5. If we assume electrostatic effects are directly correlated with IDP structure, we should see very little change in CD spectra or R_h moving from pH 7.0 to 6.0 for all three peptides, with a larger change as we move to pH 4.5. If our hypothesis that PP_{II} content modulates charge effects is correct, we should see a dramatically larger shift in R_h and CD spectra as we decrease the pH for the two mutants compared to the wildtype.

Our results matched our hypothesis. As we decreased pH, the WT decreased in size by 4.7 Å (15%) whereas the Ala^- mutant decreased by 4.3 Å (14%) and the Pro^- mutant decreased by 7.4 Å (27%).

CD spectroscopy was mostly used qualitatively to assess secondary structure. The CD spectra for the WT as a function of temperature had a peak at 221 nm, characteristic

of the PP_{II} helix[41–43]. As pH was decreased, this peak remained but was flattened. Similar results were observed for the Ala^- mutant. Since the Pro^- mutant is devoid of much of the PP_{II} content seen in the other two, there wasn't much of a characteristic peak at 221 nm. As pH was decreased from 7.0 to 5.0, the spectra flattened as expected. However, at pH 4.5, the spectra appeared random with no temperature dependence. By reverting the same sample back to pH 7.0, a normal temperature-dependent spectra was obtained.

In sum, our data from size exclusion chromatography and CD support our hypothesis that PP_{II} content modulates charge effects in intrinsically disordered proteins. While net charge certainly plays a role, it appears to be secondary to PP_{II} content. It also appears that alanine, while calculated to be a prominent disorder-promoting residue in two of the three published PP_{II} propensity scales, has less of an effect on PP_{II} structure than expected.

II. MATERIALS AND METHODS

2.1 Materials

Water used for reagents and growth media was filtered and deionized using an EMD Millipore Milli-Q Integral 3 water purification system (Billerica, MA). Equipment and growth media were 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). Nickel affinity chromatography, ion exchange chromatography, and size exclusion chromatography were all performed using a Bio-Rad Biologic LP low pressure chromatography system (Hercules, CA). A Beckman Coulter Φ 510 pH meter was used to adjust the pH of all solutions (Brea, CA). A&D GH-200 analytical balance was used to weigh all materials and reagents (Tokyo, Japan). Degassing and vacuum filtration were performed using a Welch DryFast 2032 Ultra Diaphragm Pump (Niles, IL). CD spectroscopy was performed using a Jasco J-710 spectropolarimeter equipped with a Jasco spectropolarimeter power supply, and Jasco PFD-425S Peltier (Easton, MD).

All chemicals and reagents used were ACS grade or higher. Culture tubes used in this project were VWR disposable 16×100-mm borosilicate. Wooden applicators were

from Medline Industries (Mundelein, IL). All dialysis was performed using Spectrum Labs Spectra/Por tubing with a 12-14 kD molecular weight cut-off (Rancho Dominguez, CA). Molecular weights for p53(1-93) variants and folded protein standards were calculated from the primary sequence using the ExPASy ProtParm tool[44].

All centrifugation steps were carried out using a Thermo Fisher Scientific Sorvall LYNX 6000 Superspeed centrifuge unless otherwise noted.

2.2 Sequences

The intrinsically disordered N-terminus of the human tumor suppressor protein p53, containing a transactivation domain (TAD; residues 1-60) and an adjacent proline-rich region (PRR; residues 61-93)[32], was used as a model peptide for this project (WT). The amino acid sequence was taken from GenBank entry AAA59989.1[45] as shown in Figure 2.1.

```
  1  MEEPQSDPSV   EPPLSQETFS   DLWKLLPENN   VLSPLPSQAM   DDLMLSPDDI
 51  EQWFTEDPGP   DEAPRMPEAA   PPVAPAPAAP   TPAAPAPAPS   WPL
```

Figure 2.1. Sequence for the intrinsically disordered N-terminus of the human tumor suppressor protein p53(1-93) WT. Green lettering represents the transactivation domain while orange lettering represents the proline-rich region.

Two mutant p53(1-93) peptides were used as well. One in which all 12 alanines were replaced with glycines (Ala^-) and one in which all 22 prolines were replaced with glycines (Pro^-). These sequences are shown in Figures 2.2 and 2.3, respectively.

Figure 2.2. Sequence for p53(1-93) Ala$^-$. Green lettering represents the transactivation domain while orange lettering represents the proline-rich region. Mutation sites are bolded and underlined.

Figure 2.3. Sequence for p53(1-93) Pro$^-$. Green lettering represents the transactivation domain while orange lettering represents the proline-rich region. Mutation sites are bolded and underlined.

2.3 Expression and Purification of Recombinant p53(1-93)

2.3.1 Cloning and Transformation

The following methods are for the WT peptide. The cloning and transformation procedures were identical for the Ala$^-$ and Pro$^-$ mutants.

The sequence in Figure 2.1 was synthesized and cloned into a pJ404 expression vector by DNA 2.0 (Menlo Park, 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 p53(1-93) WT plus an N-terminal 6x histidine-tag and thrombin cut site was placed under the control of an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible T5 promoter. Nucleotide sequences for all three plasmids are shown in Figures 2.4 – 2.6 at the end of this chapter.

Upon receipt, the vial containing 2-5 μg plasmid DNA was centrifuged for 2 minutes at 2,000 × g using a Beckman Coulter Microfuge 16 centrifuge. The plasmid DNA was then solubilized in 200 μL DNA grade sterile water (final concentration 10 ng/μL) and allowed to incubate at room temperature for 10 minutes. The plasmid DNA was gently mixed by pipetting, then transferred to 5-μL aliquots and stored at -80°C.

A 50-μL aliquot of *E. coli* BL21 (DE3) pLysS competent cells by EMD Millipore were thawed on ice for 10 minutes. Transformation was achieved by adding 2.5 μL

plasmid DNA and gently flicking 1-3 times to mix. The mixture was allowed to incubate on ice for at least 5 minutes. The cells were then heat shocked for exactly 30 seconds in a 42°C water bath, then immediately returned to ice for an addition 2 minute incubation. After ice incubation, 250 μL of super optimal broth with catabolite repression (SOC) media was added to the cells. The cells were then incubated at 37°C for 60 minutes at 225 RPM.

After the outgrowth incubation, the cells were spread on Lysogeny Broth (LB) agar plates containing 100 μg/mL ampicillin (LB+Amp). Three plates were used to maximize transformant recovery with 50 μL pipetted on the first, 100 μL on the second, and the remainder (~150 μL) on the third. A glass spreader was dipped in 100% ethanol and flamed to sterilize. Upon cooling, the spreader was quenched on the agar, and the broth was gently spread on all three plates. Each plate was incubated 15-24 hours at 37°C.

Using a sterile wooden applicator stick, a single colony from one of the plates was transferred to a culture tube containing 3 mL LB+Amp. The tube was incubated 15-24 hours at 30°C with 170 RPM shaking. Glycerol stocks were made by combining 750 μL of the overnight culture with 250 μL of 50% (v/v) glycerol in a 2-mL cryovial. The cryovial was then briefly vortexed and stored at -80°C.

2.3.2 Protein Expression

The following methods are for p53(1-93) WT. The expression procedures were identical for the Ala^- and Pro^- mutants.

Using the previously created glycerol stock, an LB+Amp agar plate was streaked for isolation using a sterile wooden applicator stick. The plate was incubated at 37°C for 15-24 hours.

The next day, a single colony was transferred to a culture tube containing 3 mL LB+Amp. The tube was incubated 25-24 hours at 30°C with 170 RPM shaking.

The following day, 2 mL of the overnight broth culture was subcultured into a 2-L flask containing 1 L LB+Amp. The cells were incubated at 37°C with 170 RPM shaking until mid-exponential phase ($0.6 \leq OD_{600} \leq 0.7$).

Upon reaching the correct optical density, 1 mL 0.5 M IPTG was added to the cell culture (final concentration 0.5 mM) to induce protein expression. The cells were incubated at 37°C with 170 RPM shaking for four hours. The cells were then centrifuged at 30,240 × g for 15 minutes at 4°C using a F14-6x250y rotor. The supernatant was discarded and cell pellets were stored at -80°C.

2.3.3 Nickel Affinity Chromatography

Cell pellets were thawed and solubilized in Lysis Buffer (6 M guanidine, 10 mM tris, 100 mM Na_2PO_4, pH 8.0) then sonicated on ice at 80% duty, 50% output. Sonication was done using four cycles of 90 seconds on and 90 seconds off. The cell lysate was centrifuged at max speed (33,746 × g) for 60 minutes at 4°C.

While the cells were centrifuging, a 1.5×15-cm Bio-Rad Econo-column was filled with 13 mL HIS-Select nickel affinity gel from Sigma-Aldrich (St. Louis, MO). The column was attached to a low pressure chromatography system and equilibrated with Equilibration Buffer (6 M guanidine, 10 mM tris, 100 mM Na_2PO_4, pH 8.0) until the conductivity and absorbance at 280 nm (A_{280}) leveled off.

After centrifugation, the supernatant was loaded on the nickel affinity column, then washed with 45 mL of Wash Buffer #1 (6 M guanidine, 10 mM tris, 100 mM Na_2PO_4, pH 8.0). The column was then washed with 40 mL of Wash Buffer #2 (10 mM tris, 100 mM Na_2PO_4, pH 8.0) to remove all traces of guanidine. The column was then washed with 40 mL of Wash Buffer #3 (10 mM tris, 100 mM Na_2PO_4, imidazole 10 mM, pH 8.0) to remove any proteins with weak Ni^{2+} affinity. The target protein was then eluted with 40 mL Elution Buffer (10 mM tris, 100 mM Na_2PO_4, imidazole 10 mM, pH 4.3). Collection of the eluate began as soon as the conductivity began spiking. The eluate was collected in a sterile 15-mL conical until the A_{280} reading hit a local minimum. The eluate

was then dialyzed overnight at 4°C against tris-buffered saline (TBS; 20 mM tris, 100 mM NaCl, pH 8.0).

An example chromatograph for each p53(1-93) variant is shown in Figures 2.7 – 2.9 at the end of this chapter.

2.3.4 Ion Exchange Chromatography

The day before running the ion exchange column, 13-mL of DEAE Sephacel from GE Healthcare (Chicago, IL) was poured into a sterile 50-mL conical tube and allowed to settle overnight.

The next day, the eluate was removed from dialysis and placed in a sterile 15-mL conical tube. A 30 μL aliquot (50 U) of recombinant human thrombin from Sigma-Aldrich was rapidly thawed and added to the protein sample. The conical was wrapped with tin foil and taped to a Thermo Fisher Scientific Large 3-D rotator. Thrombin digestion of the HIS-tag proceeded for 4 hours at room temperature with gentle rotation (speed 2).

During thrombin digestion, the ethanol from the DEAE media was decanted and replaced with 30 mL Column Media Buffer (20 mM sodium acetate, 25 mM NaCl, pH 3.5). The media was placed into a slurry and poured into a 50-mL beaker. The beaker was loosely covered with paraffin film and degassed for at least 30 minutes. After degassing, the DEAE media was gently poured into a 1.5×15-cm Bio-Rad Econo-column.

The column was attached to a low pressure chromatography system and equilibrated with Equilibration Buffer (20 mM sodium acetate, 25 mM NaCl, pH 4.8) until the conductivity and A_{280} leveled off.

Once digestion was completed, the protein sample was loaded on the DEAE column, then washed with 50 mL Wash Buffer 50 (20 mM sodium acetate, 50 mM NaCl, pH 4.8) to remove any weakly binding proteins. The column was then washed with 50 mL Wash Buffer 150 (20 mM sodium acetate, 50 mM NaCl, pH 4.8). The target protein was then eluted with 50 mL Elution Buffer (20 mM sodium acetate, 400 mM NaCl, pH 4.8). Collection of the eluate began as soon as the conductivity began spiking. The eluate was collected in a sterile 15-mL conical until the A_{280} reading began to level out. The eluate was then dialyzed overnight at 4°C against phosphate-buffered saline (10 mM sodium phosphate, 100 mM NaCl, pH 7.0).

An example chromatograph for each p53(1-93) variant is shown in Figures 2.10 – 2.12 at the end of this chapter.

2.3.5 Protein Concentration and Storage

Due to the high concentration requirements of downstream assays, the protein needed to be concentrated before moving on. The eluate was removed from dialysis and evenly aliquoted into 2-mL microcentrifuge tubes. The tubes were placed in a Thermo Fisher Scientific Savant DNA 120 SpeedVac and run without heat until the volume of

each microcentrifuge tube was ~0.2 mL. The samples were combined and dialyzed at 4°C overnight against phosphate-buffered saline (10 mM sodium phosphate, 100 mM NaCl, pH 7.0).

2.3.6 Gel Electrophoresis

Purify of each protein purification was assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). A mixture of 20 μL protein, 19 μL Laemmli Sample Buffer from Bio-Rad, and 1 μL β-mercaptoethanol was created in a 500-μL microcentrifuge tube. The mixture was heated for 5 minutes at 95°C using an Eppendorf Thermomixer R (Hamburg, Germany). During heating, a Bio-Rad 4-20% Criterion Tris-HCl gel was placed into a Bio-Rad Criterion Cell. The unit was filled with 700 μL Running Buffer (250 mM tris, 192 mM glycine, 0.1% SDS, pH 8.3). After heating, 7 μL of protein was loaded into one of the wells. Four microliters of Bio-Rad Precision Plus Protein Standards were added to one or two other wells. The remainder of the wells were loaded with 4 μL of Laemmli Sample Buffer to ensure even electrophoresis.

The Criterion Cell was attached to a Bio-Rad PowerPac HV power supply. The gel was electrophoresed at 200 V for approximately 50 minutes at 4°C.

2.3.7 Silver Staining

Due to the high negative charge of both Coomassie Blue and p53(1-93), gels were silver stained instead of Coomassie stained.

Upon completion of electrophoresis, the gel was soaked in Fixation Solution (30% ethanol, 10% acetic acid) for 30 minutes. The gel was then rinsed twice with 20% ethanol for 10 minutes each. The gel was then rinsed in water for 10 minutes each.

The gel was soaked in Sensitizing Solution (0.02% w/v sodium thiosulfate) for exactly 1 minute. The gel was then rinsed twice with water for 1 minute each.

The gel was soaked in Impregnation Solution (12 mM silver nitrate) for 20 minutes. The gel was then rinsed with water for exactly 10 seconds.

The gel was soaked in Basic Developer (3% w/v sodium carbonate, 0.025% formalin v/v, 0.00125% sodium thiosulfate) for 3-10 minutes with careful monitoring. Once the correct contrast was met, the Basic Developer was discarded and the gel was soaked in Stop Solution (2% acetic acid, 264 mM Tris) for 30 minutes. The gel was then rinsed twice in water for at least 30 minutes each.

The gel was imaged with a Bio-Rad Molecular Imager ChemiDoc XRS+ imaging system. Example gels for each p53(1-93) variant are shown in Figure 2.13 at the end of this chapter.

2.4 Circular Dichroism Spectroscopy

CD spectroscopy was used to gain insight to the secondary structure, namely the PP_{II} propensity, of each p53(1-93) variant.

A Hellma Analytics High Precision quartz cell with 1-mm path length was filled with 300 μL PBS and placed in a JASCO J-710 spectropolarimeter (Easton, MD). The spectropolarimeter was purged with compressed N_2 gas provided by Airgas (Radnor, PA) at a pressure of 22,000 PSI for 15 minutes. After the spectropolarimeter warmed up, a Geoglobal Partners FP155 water pump (Palm Beach, FL) was engaged and a JASCO PFD-425S Peltier was set to 20°C. The N^2 pressure was decreased to 17,000 PSI and a blank file was created covering an absorption range of 195-250 nm. The water pump, Peltier, and spectropolarimeter were turned off in that order. The N^2 pressure was decreased to 12,000 PSI and the machine was purged for an additional 15 minutes.

The same quartz cell was cleaned and filled with 300 μL p53(1-93) WT at a concentration of 0.170 mg/mL. The cell was placed in the spectropolarimeter, which was then purged with N^2 gas for 15 minutes at 22,000 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 197 nm to 245 nm with 8 accumulations. Temperature was set to 5°C.

The temperature within the spectropolarimeter was allowed to equilibrate for 10 minutes before measurements began. Before the first temperature scan (5°C), the N^2 pressure was decreased to 17,000 PSI. After the first scan, the N^2 pressure was decreased to 12,000 PSI where it remained for the remainder of the experiment. CD spectra was measured from 5°C – 85°C in 5° 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).

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

2.4.1 Measuring Mutants and Charge Effects

A 1-mL aliquot of p53(1-93) WT was dialyzed at 4°C overnight against 10 mM sodium phosphate, 100 mM NaCl, pH 6.0. The above procedure for CD spectroscopy was repeated to obtain a spectra at pH 6.0. This entire procedure was repeated until CD spectra was obtained for the WT, Ala^-, and Pro^- at pHs of 7.0, 6,0, 5.0, and 4.5 each.

2.5 Size Exclusion Chromatography

SEC was used to extrapolate the R_h of each p53(1-93) variant from by comparing their partition coefficients (K_D) to a set of folded protein standards. Chicken egg albumin, bovine carbonic anhydrase, *Staphylococcal* nuclease, and horse heart myoglobin were chosen as folded standards due to their wide R_h range. The R_h for each folded protein standard was estimated from the crystallographic structure by calculating the maximum distance between any two alpha carbons on a single subunit (Table 2.1).

Table 2.1. Properties of folded protein standards.

Folded Protein Standard	PDB	Residues	M.W. (Da)	$R_{h,crys}$ (Å)
Chicken egg albumin	1ova	386	42,861.3	35.8
Bovine carbonic anhydrase	1v93	259	28,982.5	27.3
Staphylococcal nuclease	1stn	149	16,811.2	21.2
Horse heart myoglobin	2o58	153	16,951.4	21.8

2.5.1 Preparation of Running Buffer

Eight liters of PBS was 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 10-L beaker and placed on an elevated benchtop.

2.5.2 Preparation of Size Exclusion Media

Eight 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 100-mL beaker, and degassed for 3 hours.

After thorough degassing, the media was placed in suspension and poured in one continuous motion into a 1.5×30-cm Bio-Rad Econo-column. The bottom of the column was attached to a low pressure 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 10-L beaker containing filtered and degassed PBS to create a syphon. Both stopcocks were opened and the column was allowed to pack overnight.

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

Albumin, carbonic anhydrase, Staphylococcal nuclease, and myoglobin were chosen as protein standards

A vial of bovine carbonic anhydrase (Sigma-Aldrich C5024) was centrifuged for 2 minutes at 2,000 × g to remove any protein in or around the cap. The lyophilized protein

was hydrolyzed with 5 mL PBS and filtered with a Pall Acrodisc LC 25 mm syringe filter with 0.2 μm polyvinylidene fluoride (PVDF) membrane (New York, NY). The filtered protein was stored as 200 μL aliquots at -80°C.

Chicken egg albumin (Thermo-Fisher Scientific Acros 40045) was hydrated with PBS to a concentration of 2 mg/mL. The hydrated protein was filtered with a 0.2 μm PVDF membrane and stored for short-term use at 4°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 *Staphylococcal* nuclease was purified by David Engelhardt as previously published[46].

2.5.4 Size Exclusion Measurements

Each p53(1-93) variant and each folded protein standard, with the exception of albumin, was measured individually in the following manner. Immediately before running a sample, 90 μL BD-DNP-Asp (3 mg/mL Blue Dextran D5751, 0.75 mg/mL DNP-Aspartate) was mixed with 10 μL of protein. The mixture was then pipetted directly to the top of the matrix bed, then covered with 3 mL PBS. The column was reconnected to the PBS reservoir to start the syphon. The start time was noted. Each p53(1-93) variant and each folded protein standard was run a minimum of three times.

Due to the large R_h of both albumin and blue dextran, albumin was run separately. A 90 μL sample of BD-DNP-Asp was measured, followed immediately by a 10 μL sample of albumin. This was repeated a minimum of three times.

An example chromatograph is shown in Figure 2.13 at the end of this chapter.

2.5.5 Measuring Charge Effects

The entire procedure was repeated with running buffer at pH 6.0, 5.0, and 4.5 with the following changes. *Staphylococcal* nuclease was used as a control as its structure is stable from pH 7.0 to pH 4.5; it doesn't begin acid denaturation until pH 4.0 (Figure 2.15)[46]. Similar results were found in the literature for chicken egg albumin and bovine carbonic anhydrase[47,48]. Wong et al. reported that bovine carbonic anhydrase does not begin denaturing until pH 4.3[47]. Ahmad and Salahuddin reported that chicken egg albumin begins acid denaturation at pH 3.5[48]. Horse heart myoglobin, however, undergoes significant unfolding at pH 4.5[49]. Thus, myoglobin was not used as a standard for pH 4.5 measurements.

2.5.6 K_D and R_h Determination

K_D for each IDP and folded protein standard was calculated according to the equation below.

$$K_D = \frac{V_E - V_0}{V_T - V_0}$$

V_0 is the void volume, the volume in which blue dextran 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-Aspartate elutes from the column.

Our 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 p53(1-93) variant at each pH.

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 each p53(1-93) variant.

$$R_h = m_{reg} K_D + b_{reg}$$

Lin's concordance correlation (ρ_c) was calculated for the set of folded protein standards and the IDPs using the R programming language and the epiR package (available: cran.r-project.org/web/packages/epiR)[50]. A 95% confidence interval was chosen.

2.6 Additional Figures

```
   1  AATTGTGAGC  GGATAACAAT  TACGAGCTTC  ATGCACAGTG  AAATCATGAA  AAATTTATTT
  61  GCTTTGTGAG  CGGATAACAA  TTATAATATG  TGGAATTGTG  AGCGCTCACA  ATTCCACAAC
 121  GGTTTCCCTC  TAGAAATAAT  TTTGTTTAAC  TTTTAGGAGG  TAAAACATAT  GCGCGGTTCC
 181  CATCATCACC  ATCATCACAG  CAGCGGTTTG  GTCCCACGTG  GTTCTATGGA  GGAACCGCAA
 241  AGCGACCCGA  GCGTGGAACC  GCCGCTGAGC  CAGGAAACCT  TCTCGGATCT  GTGGAAACTG
 301  CTGCCGGAGA  ACAATGTTTT  GTCCCCGCTG  CCGAGCCAAG  CGATGGATGA  CCTGATGCTG
 361  AGCCCGGATG  ACATCGAGCA  GTGGTTTACC  GAAGATCCGG  GTCCGGACGA  GGCGCCACGT
 421  ATGCCGGAGG  CCGCACCGCC  GGTTGCACCG  GCACCAGCTG  CGCCGACGCC  TGCGGCACCG
 481  GCTCCTGCGC  CGAGCTGGCC  GCTGTAACTC  GAGCCCCAAG  GGCGACACAA  AATTTATTCT
 541  AAATGATAAT  AAATACTGAT  AACATCTTAT  AGTTTGTATT  ATATTTTGTA  TTATCGTTGA
 601  CATGTATAAT  TTTGATATCA  AAAACTGATT  TTCCCTTTAT  TATTTTCGAG  ATTTATTTTC
 661  TTAATTCTCT  TTAACAAACT  AGAAATATTG  TATATACAAA  AAATCATAAA  TAATAGATGA
 721  ATAGTTTAAT  TATAGGTGTT  CATCAATCGA  AAAAGCAACG  TATCTTATTT  AAAGTGCGTT
 781  GCTTTTTTCT  CATTTATAAG  GTTAAATAAT  TCTCATATAT  CAAGCAAAGT  GACAGGCGCC
 841  CTTAAATATT  CTGACAAATG  CTCTTTCCCT  AAACTCCCCC  CATAAAAAAA  CCCGCCGAAG
 901  CGGGTTTTTA  CGTTATTTGC  GGATTAACGA  TTACTCGTTA  TCAGAACCGC  CCAGGGGGCC
 961  CGAGCTTAAG  ACTGGCCGTC  GTTTTACAAC  ACAGAAAGAG  TTTGTAGAAA  CGCAAAAAGG
1021  CCATCCGTCA  GGGGCCTTCT  GCTTAGTTTG  ATGCCTGGCA  GTTCCCTACT  CTCGCCTTCC
1081  GCTTCCTCGC  TCACTGACTC  GCTGCGCTCG  GTCGTTCGGC  TGCGGCGAGC  GGTATCAGCT
1141  CACTCAAAGG  CGGTAATACG  GTTATCCACA  GAATCAGGGG  ATAACGCAGG  AAAGAACATG
1201  TGAGCAAAAG  GCCAGCAAAA  GGCCAGGAAC  CGTAAAAAGG  CCGCGTTGCT  GGCGTTTTTC
1261  CATAGGCTCC  GCCCCCCTGA  CGAGCATCAC  AAAAATCGAC  GCTCAAGTCA  GAGGTGGCGA
1321  AACCCGACAG  GACTATAAAG  ATACCAGGCG  TTTCCCCCTG  GAAGCTCCCT  CGTGCGCTCT
1381  CCTGTTCCGA  CCCTGCCGCT  TACCGGATAC  CTGTCCGCCT  TTCTCCCTTC  GGGAAGCGTG
1441  GCGCTTTCTC  ATAGCTCACG  CTGTAGGTAT  CTCAGTTCGG  TGTAGGTCGT  TCGCTCCAAG
1501  CTGGGCTGTG  TGCACGAACC  CCCCGTTCAG  CCCGACCGCT  GCGCCTTATC  CGGTAACTAT
1561  CGTCTTGAGT  CCAACCCGGT  AAGACACGAC  TTATCGCCAC  TGGCAGCAGC  CACTGGTAAC
1621  AGGATTAGCA  GAGCGAGGTA  TGTAGGCGGT  GCTACAGAGT  TCTTGAAGTG  GTGGGCTAAC
1681  TACGGCTACA  CTAGAAGAAC  AGTATTTGGT  ATCTGCGCTC  TGCTGAAGCC  AGTTACCTTC
1741  GGAAAAAGAG  TTGGTAGCTC  TTGATCCGGC  AAACAAACCA  CCGCTGGTAG  CGGTGGTTTT
1801  TTTGTTTGCA  AGCAGCAGAT  TACGCGCAGA  AAAAAAGGAT  CTCAAGAAGA  TCCTTTGATC
1861  TTTTCTACGG  GGTCTGACGC  TCAGTGGAAC  GACGCGCGCG  TAACTCACGT  TAAGGGATTT
1921  TGGTCATGAG  TCACTGCCCG  CTTTCCAGTC  GGGAAACCTG  TCGTGCCAGC  TGCATTAATG
1981  AATCGGCCAA  CGCGCGGGGA  GAGGCGGTTT  GCGTATTGGG  CGCCAGGGTG  GTTTTTCTTT
2041  TCACCAGTGA  GACTGGCAAC  AGCTGATTGC  CCTTCACCGC  CTGGCCCTGA  GAGAGTTGCA
```

Figure 2.4. Nucleotide sequence for p53(1-93) WT plasmid. The red region (151-517) is the gene for p53(1-93) WT plus N-terminal 6x histidine-tag and thrombin cut site. The blue region (1126-1929) is the high-copy origin of replication. The green region (1940-3010) represents *lacI*. The gold region (3105-3953) represents *amp*^{r}.

```
2101  GCAAGCGGTC  CACGCTGGTT  TGCCCCAGCA  GGCGAAAATC  CTGTTTGATG  GTGGTTAACG
2161  GCGGGATATA  ACATGAGCTA  TCTTCGGTAT  CGTCGTATCC  CACTACCGAG  ATATCCGCAC
2221  CAACGCGCAG  CCCGGACTCG  GTAATGGCGC  GCATTGCGCC  CAGCGCCATC  TGATCGTTGG
2281  CAACCAGCAT  CGCAGTGGGA  ACGATGCCCT  CATTCAGCAT  TTGCATGGTT  TGTTGAAAAC
2341  CGGACATGGC  ACTCCAGTCG  CCTTCCCGTT  CCGCTATCGG  CTGAATTTGA  TTGCGAGTGA
2401  GATATTTATG  CCAGCCAGCC  AGACGCAGAC  GCGCCGAGAC  AGAACTTAAT  GGGCCCGCTA
2461  ACAGCGCGAT  TTGCTGGTGA  CCCAATGCGA  CCAGATGCTC  CACGCCCAGT  CGCGTACCGT
2521  CCTCATGGGA  GAAAATAATA  CTGTTGATGG  GTGTCTGGTC  AGAGACATCA  AGAAATAACG
2581  CCGGAACATT  AGTGCAGGCA  GCTTCCACAG  CAATGGCATC  CTGGTCATCC  AGCGGATAGT
2641  TAATGATCAG  CCCACTGACG  CGTTGCGCGA  GAAGATTGTG  CACCGCCGCT  TTACAGGCTT
2701  CGACGCCGCT  TCGTTCTACC  ATCGACACCA  CCACGCTGGC  ACCCAGTTGA  TCGGCGCGAG
2761  ATTTAATCGC  CGCGACAATT  TGCGACGGCG  CGTGCAGGGC  CAGACTGGAG  GTGGCAACGC
2821  CAATCAGCAA  CGACTGTTTG  CCCGCCAGTT  GTTGTGCCAC  GCGGTTGGGA  ATGTAATTCA
2881  GCTCCGCCAT  CGCCGCTTCC  ACTTTTTCCC  GCGTTTTCGC  AGAAACGTGG  CTGGCCTGGT
2941  TCACCACGCG  GGAAACGGTC  TGATAAGAGA  CACCGGCATA  CTCTGCGACA  TCGTATAACG
3001  TTACTGGTTT  CATATTCACC  ACCCTGAATT  GACTCTCTTC  CGGGCGCTAT  CATGCCATAC
3061  CGCGAAAGGT  TTTGCGCCAT  TCGATGGCGC  GCCGCTTACC  AATGCTTAAT  CAGTGAGGCA
3121  CCTATCTCAG  CGATCTGTCT  ATTTCGTTCA  TCCATAGTTG  CCTGACTCCC  CGTCGTGTAG
3181  ATAACTACGA  TACGGGAGGG  CTTACCATCT  GGCCCCAGCG  CTGCGATGAT  ACCGCGAGAA
3241  CCACGCTCAC  CGGCTCCGGA  TTTATCAGCA  ATAAACCAGC  CAGCCGGAAG  GGCCGAGCGC
3301  AGAAGTGGTC  CTGCAACTTT  ATCCGCCTCC  ATCCAGTCTA  TTAATTGTTG  CCGGGAAGCT
3361  AGAGTAAGTA  GTTCGCCAGT  TAATAGTTTG  CGCAACGTTG  TTGCCATCGC  TACAGGCATC
3421  GTGGTGTCAC  GCTCGTCGTT  TGGTATGGCT  TCATTCAGCT  CCGGTTCCCA  ACGATCAAGG
3481  CGAGTTACAT  GATCCCCCAT  GTTGTGCAAA  AAAGCGGTTA  GCTCCTTCGG  TCCTCCGATC
3541  GTTGTCAGAA  GTAAGTTGGC  CGCAGTGTTA  TCACTCATGG  TTATGGCAGC  ACTGCATAAT
3601  TCTCTTACTG  TCATGCCATC  CGTAAGATGC  TTTTCTGTGA  CTGGTGAGTA  CTCAACCAAG
3661  TCATTCTGAG  AATAGTGTAT  GCGGCGACCG  AGTTGCTCTT  GCCCGGCGTC  AATACGGGAT
3721  AATACCGCGC  CACATAGCAG  AACTTTAAAA  GTGCTCATCA  TTGGAAAACG  TTCTTCGGGG
3781  CGAAAACTCT  CAAGGATCTT  ACCGCTGTTG  AGATCCAGTT  CGATGTAACC  CACTCGTGCA
3841  CCCAACTGAT  CTTCAGCATC  TTTTACTTTC  ACCAGCGTTT  CTGGGTGAGC  AAAAACAGGA
3901  AGGCAAAATG  CCGCAAAAAA  GGGAATAAGG  GCGACACGGA  AATGTTGAAT  ACTCATATTC
3961  TTCCTTTTTC  AATATTATTG  AAGCATTTAT  CAGGGTTATT  GTCTCATGAG  CGGATACATA
4021  TTTGAATGTA  TTTAGAAAAA  TAAACAAATA  GGGGTCAGTG  TTACAACCAA  TTAACCAATT
4081  CTGAACATTA  TCGCGAGCCC  ATTTATACCT  GAATATGGCT  CATAACACCC  CTTGTTTGCC
4141  TGGCGGCAGT  AGCGCGGTGG  TCCCACCTGA  CCCCATGCCG  AACTCAGAAG  TGAAACGCCG
4201  TAGCGCCGAT  GGTAGTGTGG  GGACTCCCCA  TGCGAGAGTA  GGGAACTGCC  AGGCATCAAA
4261  TAAAACGAAA  GGCTCAGTCG  AAAGACTGGG  CCTTTCGCCC  GGGCTAATTA  TGGGGTGTCG
4321  CCCTT
```

Figure 2.4, continued. Nucleotide sequence for p53(1-93) WT plasmid. The red region (151-517) is the gene for p53(1-93) WT plus N-terminal 6x histidine-tag and thrombin cut site. The blue region (1126-1929) is the high-copy origin of replication. The green region (1940-3010) represents *lacI*. The gold region (3105-3953) represents *amp*^{r}.

```
   1  TGTCCGGTTT  TCAACAAACC  ATGCAAATGC  TGAATGAGGG  CATCGTTCCC  ACTGCGATGC
  61  TGGTTGCCAA  CGATCAGATG  GCGCTGGGCG  CAATGCGCGC  CATTACCGAG  TCCGGGCTGC
 121  GCGTTGGTGC  GGATATCTCG  GTAGTGGGAT  ACGACGATAC  CGAAGATAGC  TCATGTTATA
 181  TCCCGCCGTT  AACCACCATC  AAACAGGATT  TTCGCCTGCT  GGGGCAAACC  AGCGTGGACC
 241  GCTTGCTGCA  ACTCTCTCAG  GGCCAGGCGG  TGAAGGGCAA  TCAGCTGTTG  CCAGTCTCAC
 301  TGGTGAAAAG  AAAAACCACC  CTGGCGCCCA  ATACGCAAAC  CGCCTCTCCC  CGCGCGTTGG
 361  CCGATTCATT  AATGCAGCTG  GCACGACAGG  TTTCCCGACT  GGAAAGCGGG  CAGTGACTCA
 421  TGACCAAAAT  CCCTTAACGT  GAGTTACGCG  CGCGTCGTTC  CACTGAGCGT  CAGACCCCGT
 481  AGAAAAGATC  AAAGGATCTT  CTTGAGATCC  TTTTTTTCTG  CGCGTAATCT  GCTGCTTGCA
 541  AACAAAAAAA  CCACCGCTAC  CAGCGGTGGT  TTGTTTGCCG  GATCAAGAGC  TACCAACTCT
 601  TTTTCCGAAG  GTAACTGGCT  TCAGCAGAGC  GCAGATACCA  AATACTGTTC  TTCTAGTGTA
 661  GCCGTAGTTA  GCCCACCACT  TCAAGAACTC  TGTAGCACCG  CCTACATACC  TCGCTCTGCT
 721  AATCCTGTTA  CCAGTGGCTG  CTGCCAGTGG  CGATAAGTCG  TGTCTTACCG  GGTTGGACTC
 781  AAGACGATAG  TTACCGGATA  AGGCGCAGCG  GTCGGGCTGA  ACGGGGGGTT  CGTGCACACA
 841  GCCCAGCTTG  GAGCGAACGA  CCTACACCGA  ACTGAGATAC  CTACAGCGTG  AGCTATGAGA
 901  AAGCGCCACG  CTTCCCGAAG  GGAGAAAGGC  GGACAGGTAT  CCGGTAAGCG  GCAGGGTCGG
 961  AACAGGAGAG  CGCACGAGGG  AGCTTCCAGG  GGGAAACGCC  TGGTATCTTT  ATAGTCCTGT
1021  CGGGTTTCGC  CACCTCTGAC  TTGAGCGTCG  ATTTTTGTGA  TGCTCGTCAG  GGGGGCGGAG
1081  CCTATGGAAA  AACGCCAGCA  ACGCGGCCTT  TTTACGGTTC  CTGGCCTTTT  GCTGGCCTTT
1141  TGCTCACATG  TTCTTTCCTG  CGTTATCCCC  TGATTCTGTG  GATAACCGTA  TTACCGCCTT
1201  TGAGTGAGCT  GATACCGCTC  GCCGCAGCCG  AACGACCGAG  CGCAGCGAGT  CAGTGAGCGA
1261  GGAAGCGGAA  GGCGAGAGTA  GGGAACTGCC  AGGCATCAAA  CTAAGCAGAA  GGCCCCTGAC
1321  GGATGGCCTT  TTTGCGTTTC  TACAAACTCT  TTCTGTGTTG  TAAAACGACG  GCCAGTCTTA
1381  AGCTCGGGCC  CCCTGGGCGG  TTCTGATAAC  GAGTAATCGT  TAATCCGCAA  ATAACGTAAA
1441  AACCCGCTTC  GGCGGGTTTT  TTTATGGGGG  GAGTTTAGGG  AAAGAGCATT  TGTCAGAATA
1501  TTTAAGGGCG  CCTGTCACTT  TGCTTGATAT  ATGAGAATTA  TTTAACCTTA  TAAATGAGAA
1561  AAAAGCAACG  CACTTTAAAT  AAGATACGTT  GCTTTTTCGA  TTGATGAACA  CCTATAATTA
1621  AACTATTCAT  CTATTATTTA  TGATTTTTTG  TATATACAAT  ATTTCTAGTT  TGTTAAAGAG
1681  AATTAAGAAA  ATAAATCTCG  AAAATAATAA  AGGGAAAATC  AGTTTTTGAT  ATCAAAATTA
1741  TACATGTCAA  CGATAATACA  AAATATAATA  CAAACTATAA  GATGTTATCA  GTATTTATTA
1801  TCATTTAGAA  TAAATTTTGT  GTCGCCCTTA  ATTGTGAGCG  GATAACAATT  ACGAGCTTCA
1861  TGCACAGTGA  AATCATGAAA  AATTTATTTG  CTTTGTGAGC  GGATAACAAT  TATAATATGT
1921  GGAATTGTGA  GCGCTCACAA  TTCCACAACG  GTTTCCCTCT  AGAAATAATT  TTGTTTAACT
1981  TTTAGGAGGT  AAAACATATG  CGCGGTAGCC  ACCACCATCA  CCATCACAGC  AGCGGTTTAG
2041  TCCCACGTGG  TTCTATGGAA  GAACCGCAGA  GCGATCCGTC  GGTGGAGCCG  CCGTTGTCCC
```

Figure 2.5. Nucleotide sequence for p53(1-93) Ala$^-$ plasmid. The green region (3659-413) represents *lacI*. The red region (1980-2346) is the gene for p53(1-93) Ala$^-$ plus N-terminal 6x histidine-tag and thrombin cut site. The blue region (417-1220) is the high-copy origin of replication. The gold region (2716-3573) represents *amp*^{r}.

```
2101  AAGAAACGTT  CTCCGACCTG  TGGAAACTGC  TGCCGGAGAA  TAACGTTCTG  AGCCCGCTGC
2161  CGAGCCAGGG  CATGGACGAC  CTGATGCTGA  GCCCGGATGA  TATCGAGCAA  TGGTTTACCG
2221  AGGATCCGGG  TCCGGACGAA  GGCCCGCGTA  TGCCTGAGGG  TGGCCCACCG  GTTGGCCCTG
2281  GTCCGGGTGG  CCCGACCCCG  GGTGGTCCGG  GTCCAGGCCC  GAGCTGGCCG  CTGTAACTCG
2341  AGCCCCAAGG  GCGACACCCC  ATAATTAGCC  CGGGCGAAAG  GCCCAGTCTT  TCGACTGAGC
2401  CTTTCGTTTT  ATTTGATGCC  TGGCAGTTCC  CTACTCTCGC  ATGGGGAGTC  CCCACACTAC
2461  CATCGGCGCT  ACGGCGTTTC  ACTTCTGAGT  TCGGCATGGG  GTCAGGTGGG  ACCACCGCGC
2521  TACTGCCGCC  AGGCAAACAA  GGGGTGTTAT  GAGCCATATT  CAGGTATAAA  TGGGCTCGCG
2581  ATAATGTTCA  GAATTGGTTA  ATTGGTTGTA  ACACTGACCC  CTATTTGTTT  ATTTTTCTAA
2641  ATACATTCAA  ATATGTATCC  GCTCATGAGA  CAATAACCCT  GATAAATGCT  TCAATAATAT
2701  TGAAAAAGGA  AGAATATGAG  TATTCAACAT  TTCCGTGTCG  CCCTTATTCC  CTTTTTTGCG
2761  GCATTTTGCC  TTCCTGTTTT  TGCTCACCCA  GAAACGCTGG  TGAAAGTAAA  AGATGCTGAA
2821  GATCAGTTGG  GTGCACGAGT  GGGTTACATC  GAACTGGATC  TCAACAGCGG  TAAGATCCTT
2881  GAGAGTTTTC  GCCCCGAAGA  ACGTTTTCCA  ATGATGAGCA  CTTTTAAAGT  TCTGCTATGT
2941  GGCGCGGTAT  TATCCCGTAT  TGACGCCGGG  CAAGAGCAAC  TCGGTCGCCG  CATACACTAT
3001  TCTCAGAATG  ACTTGGTTGA  GTACTCACCA  GTCACAGAAA  AGCATCTTAC  GGATGGCATG
3061  ACAGTAAGAG  AATTATGCAG  TGCTGCCATA  ACCATGAGTG  ATAACACTGC  GGCCAACTTA
3121  CTTCTGACAA  CGATCGGAGG  ACCGAAGGAG  CTAACCGCTT  TTTTGCACAA  CATGGGGGAT
3181  CATGTAACTC  GCCTTGATCG  TTGGGAACCG  GAGCTGAATG  AAGCCATACC  AAACGACGAG
3241  CGTGACACCA  CGATGCCTGT  AGCGATGGCA  ACAACGTTGC  GCAAACTATT  AACTGGCGAA
3301  CTACTTACTC  TAGCTTCCCG  GCAACAATTA  ATAGACTGGA  TGGAGGCGGA  TAAAGTTGCA
3361  GGACCACTTC  TGCGCTCGGC  CCTTCCGGCT  GGCTGGTTTA  TTGCTGATAA  ATCCGGAGCC
3421  GGTGAGCGTG  GTTCTCGCGG  TATCATCGCA  GCGCTGGGGC  CAGATGGTAA  GCCCTCCCGT
3481  ATCGTAGTTA  TCTACACGAC  GGGGAGTCAG  GCAACTATGG  ATGAACGAAA  TAGACAGATC
3541  GCTGAGATAG  GTGCCTCACT  GATTAAGCAT  TGGTAAGCGG  CGCGCCATCG  AATGGCGCAA
3601  AACCTTTCGC  GGTATGGCAT  GATAGCGCCC  GGAAGAGAGT  CAATTCAGGG  TGGTGAATAT
3661  GAAACCAGTA  ACGTTATACG  ATGTCGCAGA  GTATGCCGGT  GTCTCTTATC  AGACCGTTTC
3721  CCGCGTGGTG  AACCAGGCCA  GCCACGTTTC  TGCGAAAACG  CGGGAAAAAG  TGGAAGCGGC
3781  GATGGCGGAG  CTGAATTACA  TTCCCAACCG  CGTGGCACAA  CAACTGGCGG  GCAAACAGTC
3841  GTTGCTGATT  GGCGTTGCCA  CCTCCAGTCT  GGCCCTGCAC  GCGCCGTCGC  AAATTGTCGC
3901  GGCGATTAAA  TCTCGCGCCG  ATCAACTGGG  TGCCAGCGTG  GTGGTGTCGA  TGGTAGAACG
3961  AAGCGGCGTC  GAAGCCTGTA  AAGCGGCGGT  GCACAATCTT  CTCGCGCAAC  GCGTCAGTGG
4021  GCTGATCATT  AACTATCCGC  TGGATGACCA  GGATGCCATT  GCTGTGGAAG  CTGCCTGCAC
4081  TAATGTTCCG  GCGTTATTTC  TTGATGTCTC  TGACCAGACA  CCCATCAACA  GTATTATTTT
4141  CTCCCATGAG  GACGGTACGC  GACTGGGCGT  GGAGCATCTG  GTCGCATTGG  GTCACCAGCA
4201  AATCGCGCTG  TTAGCGGGCC  CATTAAGTTC  TGTCTCGGCG  CGTCTGCGTC  TGGCTGGCTG
4261  GCATAAATAT  CTCACTCGCA  ATCAAATTCA  GCCGATAGCG  GAACGGGAAG  GCGACTGGAG
4321  TGCCA
```

Figure 2.5, continued. Nucleotide sequence for p53(1-93) Ala$^-$ plasmid. The green region (3659-413) represents *lacI*. The red region (1980-2346) is the gene for p53(1-93) Ala$^-$ plus N-terminal 6x histidine-tag and thrombin cut site. The blue region (417-1220) is the high-copy origin of replication. The gold region (2716-3573) represents *amp^r*.

```
   1  TGTCCGGTTT  TCAACAAACC  ATGCAAATGC  TGAATGAGGG  CATCGTTCCC  ACTGCGATGC
  61  TGGTTGCCAA  CGATCAGATG  GCGCTGGGCG  CAATGCGCGC  CATTACCGAG  TCCGGGCTGC
 121  GCGTTGGTGC  GGATATCTCG  GTAGTGGGAT  ACGACGATAC  CGAAGATAGC  TCATGTTATA
 181  TCCCGCCGTT  AACCACCATC  AAACAGGATT  TTCGCCTGCT  GGGGCAAACC  AGCGTGGACC
 241  GCTTGCTGCA  ACTCTCTCAG  GGCCAGGCGG  TGAAGGGCAA  TCAGCTGTTG  CCAGTCTCAC
 301  TGGTGAAAAG  AAAAACCACC  CTGGCGCCCA  ATACGCAAAC  CGCCTCTCCC  CGCGCGTTGG
 361  CCGATTCATT  AATGCAGCTG  GCACGACAGG  TTTCCCGACT  GGAAAGCGGG  CAGTGACTCA
 421  TGACCAAAAT  CCCTTAACGT  GAGTTACGCG  CGCGTCGTTC  CACTGAGCGT  CAGACCCCGT
 481  AGAAAAGATC  AAAGGATCTT  CTTGAGATCC  TTTTTTTCTG  CGCGTAATCT  GCTGCTTGCA
 541  AACAAAAAAA  CCACCGCTAC  CAGCGGTGGT  TTGTTTGCCG  GATCAAGAGC  TACCAACTCT
 601  TTTTCCGAAG  GTAACTGGCT  TCAGCAGAGC  GCAGATACCA  AATACTGTTC  TTCTAGTGTA
 661  GCCGTAGTTA  GCCCACCACT  TCAAGAACTC  TGTAGCACCG  CCTACATACC  TCGCTCTGCT
 721  AATCCTGTTA  CCAGTGGCTG  CTGCCAGTGG  CGATAAGTCG  TGTCTTACCG  GGTTGGACTC
 781  AAGACGATAG  TTACCGGATA  AGGCGCAGCG  GTCGGGCTGA  ACGGGGGGTT  CGTGCACACA
 841  GCCCAGCTTG  GAGCGAACGA  CCTACACCGA  ACTGAGATAC  CTACAGCGTG  AGCTATGAGA
 901  AAGCGCCACG  CTTCCCGAAG  GGAGAAAGGC  GGACAGGTAT  CCGGTAAGCG  GCAGGGTCGG
 961  AACAGGAGAG  CGCACGAGGG  AGCTTCCAGG  GGGAAACGCC  TGGTATCTTT  ATAGTCCTGT
1021  CGGGTTTCGC  CACCTCTGAC  TTGAGCGTCG  ATTTTTGTGA  TGCTCGTCAG  GGGGGCGGAG
1081  CCTATGGAAA  AACGCCAGCA  ACGCGGCCTT  TTTACGGTTC  CTGGCCTTTT  GCTGGCCTTT
1141  TGCTCACATG  TTCTTTCCTG  CGTTATCCCC  TGATTCTGTG  GATAACCGTA  TTACCGCCTT
1201  TGAGTGAGCT  GATACCGCTC  GCCGCAGCCG  AACGACCGAG  CGCAGCGAGT  CAGTGAGCGA
1261  GGAAGCGGAA  GGCGAGAGTA  GGGAACTGCC  AGGCATCAAA  CTAAGCAGAA  GGCCCCTGAC
1321  GGATGGCCTT  TTTGCGTTTC  TACAAACTCT  TTCTGTGTTG  TAAAACGACG  GCCAGTCTTA
1381  AGCTCGGGCC  CCCTGGGCGG  TTCTGATAAC  GAGTAATCGT  TAATCCGCAA  ATAACGTAAA
1441  AACCCGCTTC  GGCGGGTTTT  TTTATGGGGG  GAGTTTAGGG  AAAGAGCATT  TGTCAGAATA
1501  TTTAAGGGCG  CCTGTCACTT  TGCTTGATAT  ATGAGAATTA  TTTAACCTTA  TAAATGAGAA
1561  AAAAGCAACG  CACTTTAAAT  AAGATACGTT  GCTTTTTCGA  TTGATGAACA  CCTATAATTA
1621  AACTATTCAT  CTATTATTTA  TGATTTTTTG  TATATACAAT  ATTTCTAGTT  TGTTAAAGAG
1681  AATTAAGAAA  ATAAATCTCG  AAAATAATAA  AGGGAAAATC  AGTTTTTGAT  ATCAAAATTA
1741  TACATGTCAA  CGATAATACA  AAATATAATA  CAAACTATAA  GATGTTATCA  GTATTTATTA
1801  TCATTTAGAA  TAAATTTTGT  GTCGCCCTTA  ATTGTGAGCG  GATAACAATT  ACGAGCTTCA
1861  TGCACAGTGA  AATCATGAAA  AATTTATTTG  CTTTGTGAGC  GGATAACAAT  TATAATATGT
1921  GGAATTGTGA  GCGCTCACAA  TTCCACAACG  GTTTCCCTCT  AGAAATAATT  TTGTTTAACT
1981  TTTAGGAGGT  AAAACATATG  CGCGGTTCCC  ATCATCACCA  CCATCACTCT  AGCGGCTTGG
2041  TCCCACGTGG  TAGCATGGAA  GAGGGCCAAA  GCGATGGTAG  CGTTGAGGGT  GGCCTGTCCC
```

Figure 2.6. Nucleotide sequence for p53(1-93) Pro$^-$ plasmid. The green region (3659-413) represents *lacI*. The red region (1980-2346) is the gene for p53(1-93) Pro$^-$ plus N-terminal 6x histidine-tag and thrombin cut site. The blue region (417-1220) is the high-copy origin of replication. The gold region (2716-3573) represents *amp^r*.

```
2101  AAGAAACGTT  TAGCGACTTG  TGGAAACTGC  TGGGCGAGAA  CAATGTTCTG  TCGGGTCTGG
2161  GTAGCCAGGC  GATGGATGAC  CTGATGCTGA  GCGGTGACGA  TATCGAGCAG  TGGTTCACCG
2221  AAGATGGCGG  CGGCGACGAA  GCTGGCCGTA  TGGGCGAGGC  AGCGGGTGGT  GTGGCGGGTG
2281  CGGGTGCGGC  AGGCACCGGT  GCAGCTGGTG  CCGGTGCCGG  TAGCTGGGGT  CTGTAACTCG
2341  AGCCCCAAGG  GCGACACCCC  ATAATTAGCC  CGGGCGAAAG  GCCCAGTCTT  TCGACTGAGC
2401  CTTTCGTTTT  ATTTGATGCC  TGGCAGTTCC  CTACTCTCGC  ATGGGGAGTC  CCCACACTAC
2461  CATCGGCGCT  ACGGCGTTTC  ACTTCTGAGT  TCGGCATGGG  GTCAGGTGGG  ACCACCGCGC
2521  TACTGCCGCC  AGGCAAACAA  GGGGTGTTAT  GAGCCATATT  CAGGTATAAA  TGGGCTCGCG
2581  ATAATGTTCA  GAATTGGTTA  ATTGGTTGTA  ACACTGACCC  CTATTTGTTT  ATTTTTCTAA
2641  ATACATTCAA  ATATGTATCC  GCTCATGAGA  CAATAACCCT  GATAAATGCT  TCAATAATAT
2701  TGAAAAAGGA  AGAATATGAG  TATTCAACAT  TTCCGTGTCG  CCCTTATTCC  CTTTTTTGCG
2761  GCATTTTGCC  TTCCTGTTTT  TGCTCACCCA  GAAACGCTGG  TGAAAGTAAA  AGATGCTGAA
2821  GATCAGTTGG  GTGCACGAGT  GGGTTACATC  GAACTGGATC  TCAACAGCGG  TAAGATCCTT
2881  GAGAGTTTTC  GCCCCGAAGA  ACGTTTTCCA  ATGATGAGCA  CTTTTAAAGT  TCTGCTATGT
2941  GGCGCGGTAT  TATCCCGTAT  TGACGCCGGG  CAAGAGCAAC  TCGGTCGCCG  CATACACTAT
3001  TCTCAGAATG  ACTTGGTTGA  GTACTCACCA  GTCACAGAAA  AGCATCTTAC  GGATGGCATG
3061  ACAGTAAGAG  AATTATGCAG  TGCTGCCATA  ACCATGAGTG  ATAACACTGC  GGCCAACTTA
3121  CTTCTGACAA  CGATCGGAGG  ACCGAAGGAG  CTAACCGCTT  TTTTGCACAA  CATGGGGGAT
3181  CATGTAACTC  GCCTTGATCG  TTGGGAACCG  GAGCTGAATG  AAGCCATACC  AAACGACGAG
3241  CGTGACACCA  CGATGCCTGT  AGCGATGGCA  ACAACGTTGC  GCAAACTATT  AACTGGCGAA
3301  CTACTTACTC  TAGCTTCCCG  GCAACAATTA  ATAGACTGGA  TGGAGGCGGA  TAAAGTTGCA
3361  GGACCACTTC  TGCGCTCGGC  CCTTCCGGCT  GGCTGGTTTA  TTGCTGATAA  ATCCGGAGCC
3421  GGTGAGCGTG  GTTCTCGCGG  TATCATCGCA  GCGCTGGGGC  CAGATGGTAA  GCCCTCCCGT
3481  ATCGTAGTTA  TCTACACGAC  GGGGAGTCAG  GCAACTATGG  ATGAACGAAA  TAGACAGATC
3541  GCTGAGATAG  GTGCCTCACT  GATTAAGCAT  TGGTAAGCGG  CGCGCCATCG  AATGGCGCAA
3601  AACCTTTCGC  GGTATGGCAT  GATAGCGCCC  GGAAGAGAGT  CAATTCAGGG  TGGTGAATAT
3661  GAAACCAGTA  ACGTTATACG  ATGTCGCAGA  GTATGCCGGT  GTCTCTTATC  AGACCGTTTC
3721  CCGCGTGGTG  AACCAGGCCA  GCCACGTTTC  TGCGAAAACG  CGGGAAAAAG  TGGAAGCGGC
3781  GATGGCGGAG  CTGAATTACA  TTCCCAACCG  CGTGGCACAA  CAACTGGCGG  GCAAACAGTC
3841  GTTGCTGATT  GGCGTTGCCA  CCTCCAGTCT  GGCCCTGCAC  GCGCCGTCGC  AAATTGTCGC
3901  GGCGATTAAA  TCTCGCGCCG  ATCAACTGGG  TGCCAGCGTG  GTGGTGTCGA  TGGTAGAACG
3961  AAGCGGCGTC  GAAGCCTGTA  AAGCGGCGGT  GCACAATCTT  CTCGCGCAAC  GCGTCAGTGG
4021  GCTGATCATT  AACTATCCGC  TGGATGACCA  GGATGCCATT  GCTGTGGAAG  CTGCCTGCAC
4081  TAATGTTCCG  GCGTTATTTC  TTGATGTCTC  TGACCAGACA  CCCATCAACA  GTATTATTTT
4141  CTCCCATGAG  GACGGTACGC  GACTGGGCGT  GGAGCATCTG  GTCGCATTGG  GTCACCAGCA
4201  AATCGCGCTG  TTAGCGGGCC  CATTAAGTTC  TGTCTCGGCG  CGTCTGCGTC  TGGCTGGCTG
4261  GCATAAATAT  CTCACTCGCA  ATCAAATTCA  GCCGATAGCG  GAACGGGAAG  GCGACTGGAG
4321  TGCCA
```

Figure 2.6, continued. Nucleotide sequence for p53(1-93) Pro$^-$ plasmid. The green region (3659-413) represents *lacI*. The red region (1980-2346) is the gene for p53(1-93) Pro$^-$ plus N-terminal 6x histidine-tag and thrombin cut site. The blue region (417-1220) is the high-copy origin of replication. The gold region (2716-3573) represents *amp^r*.

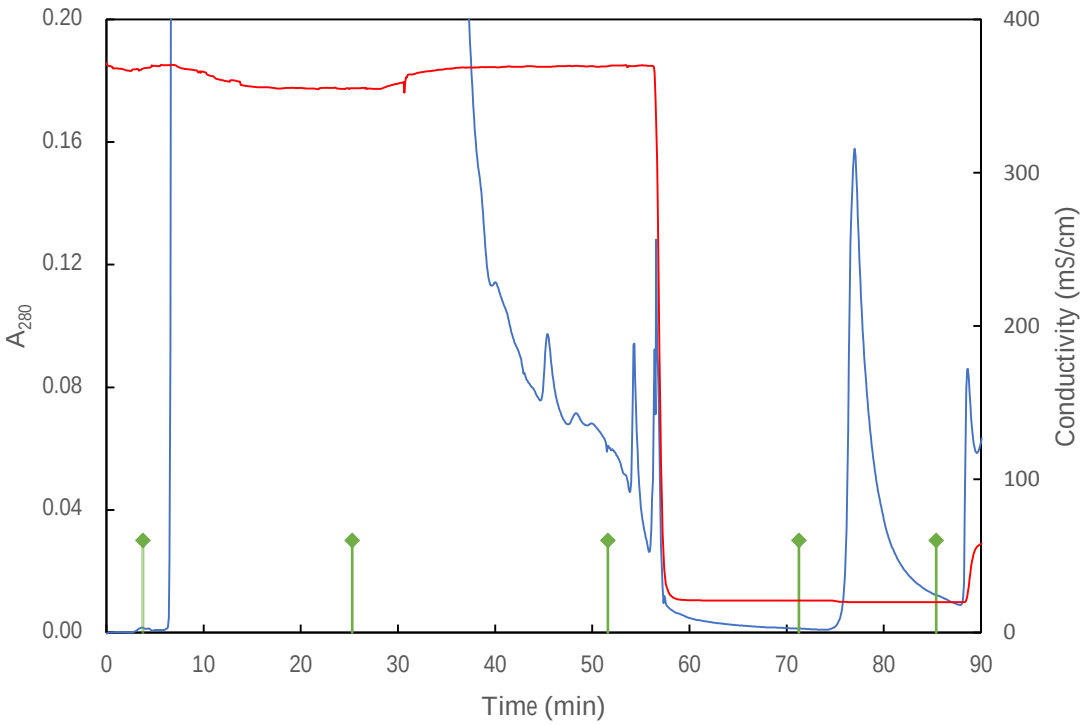

Figure 2.7. Chromatograph for nickel affinity purification of p53(1-93) WT. The blue line represents absorbance at 280 nm. The red line represents conductivity. The green diamonds are event markers. In chronological order: sample load (3.75 min), addition of Wash Buffer 1 (25.28 min), addition of Wash Buffer 2 (51.61 min), addition of Wash Buffer 3 (71.28 min), and addition of Elution Buffer (85.39 min). The peak at 86.63 min is the target protein.

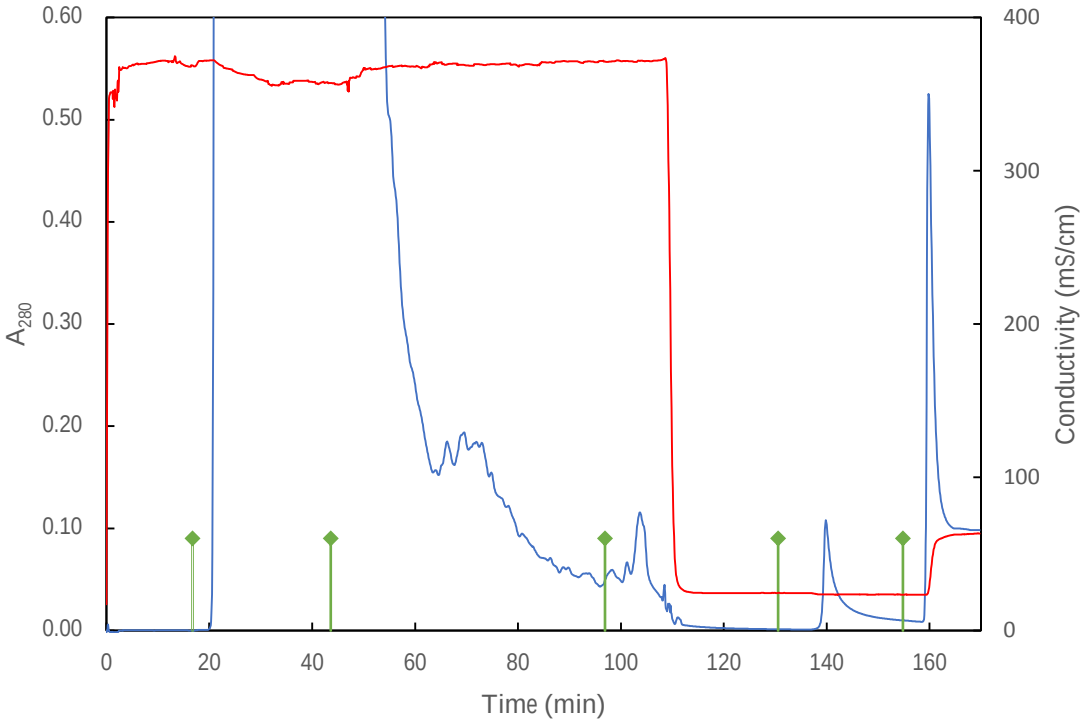

Figure 2.8. Chromatograph for nickel affinity purification of p53(1-93) Ala^-. The blue line represents absorbance at 280 nm. The red line represents conductivity. The green diamonds are event markers. In chronological order: sample load (16.75 min), addition of Wash Buffer 1 (43.61 min), addition of Wash Buffer 2 (96.89 min), addition of Wash Buffer 3 (130.60 min), and addition of Elution Buffer (154.80 min). The peak at 159.84 min is the target protein.

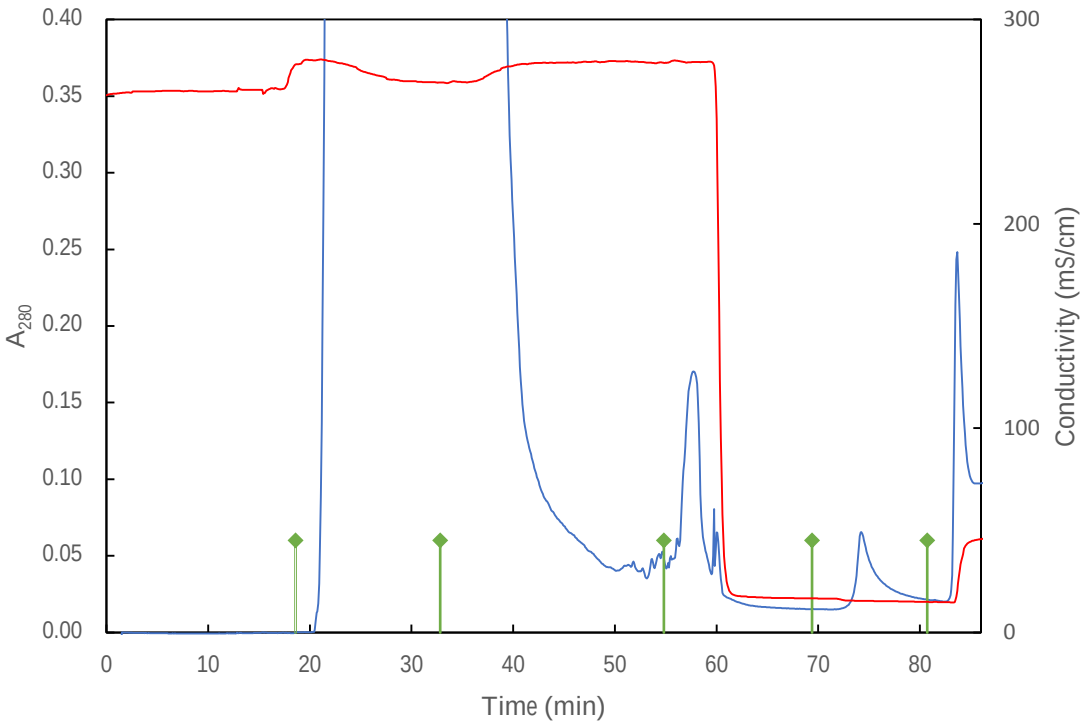

Figure 2.9. Chromatograph for nickel affinity purification of p53(1-93) Pro⁻. The blue line represents absorbance at 280 nm. The red line represents conductivity. The green diamonds are event markers. In chronological order: sample load (18.59 min), addition of Wash Buffer 1 (32.83 min), addition of Wash Buffer 2 (54.81 min), addition of Wash Buffer 3 (69.38 min), and addition of Elution Buffer (80.73 min). The peak at 83.64 min is the target protein.

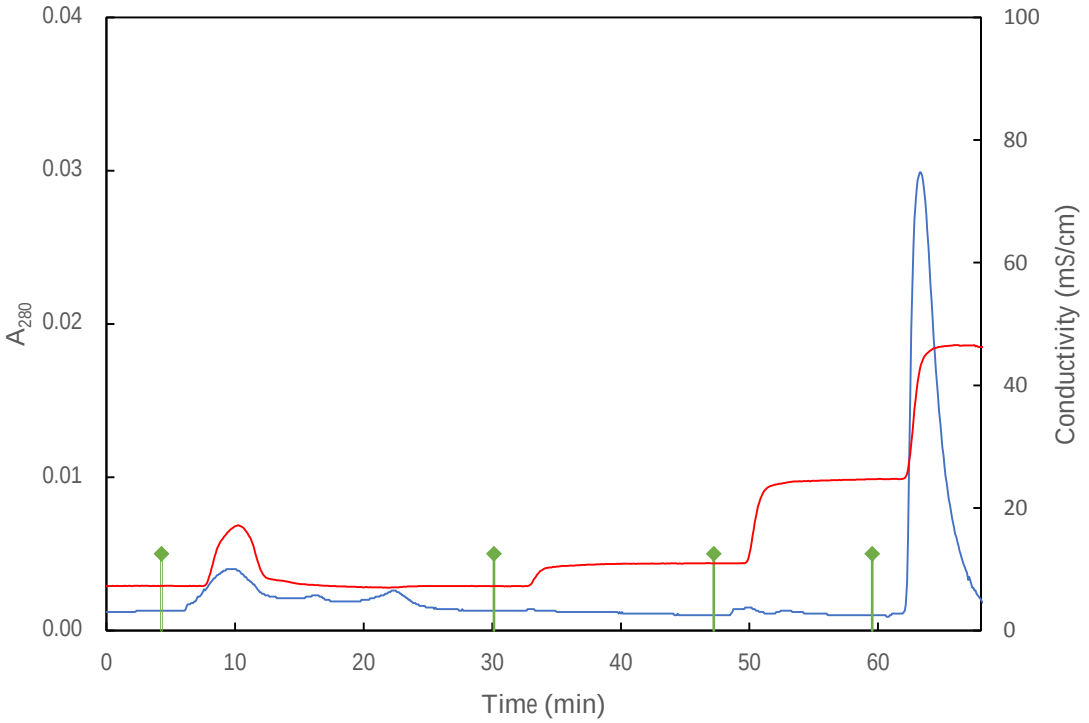

Figure 2.10. Chromatograph for anion exchange purification of p53(1-93) WT. The blue line represents absorbance at 280 nm. The red line represents conductivity. The green diamonds are event markers. In chronological order: sample load (4.29 min), addition of Wash Buffer 50 (30.14 min), addition of Wash Buffer 150 (47.22 min), and addition of Elution Buffer (59.54 min). The peak at 93.29 min is the target protein.

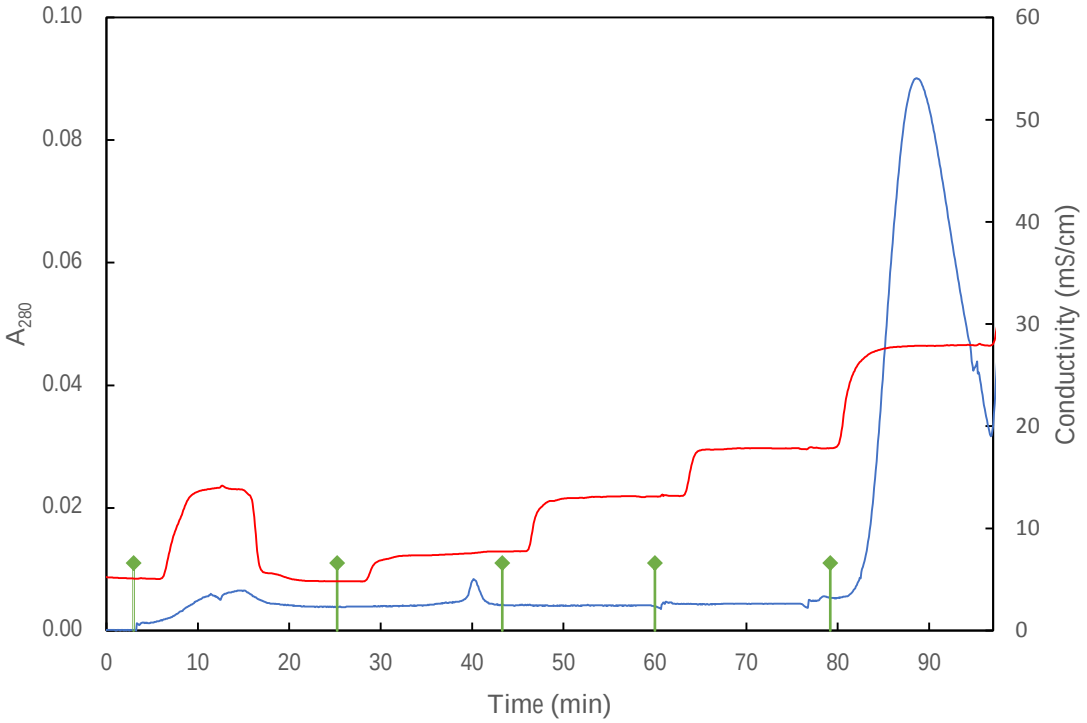

Figure 2.11. Chromatograph for anion exchange purification of p53(1-93) Ala$^-$. The blue line represents absorbance at 280 nm. The red line represents conductivity. The green diamonds are event markers. In chronological order: sample load (2.96 min), addition of Wash Buffer 50 (25.25 min), addition of Wash Buffer 150 (43.29 min), addition of Wash Buffer 250 (20 mM sodium acetate, 250 mM NaCl, pH 4.8; 60.00 min), and addition of Elution Buffer (59.54 min). The peak at 88.63 min is the target protein. An additional wash step was used as this was the first purification of p53(1-93) Ala$^-$ using DEAE media and I was unsure where the protein would elute.

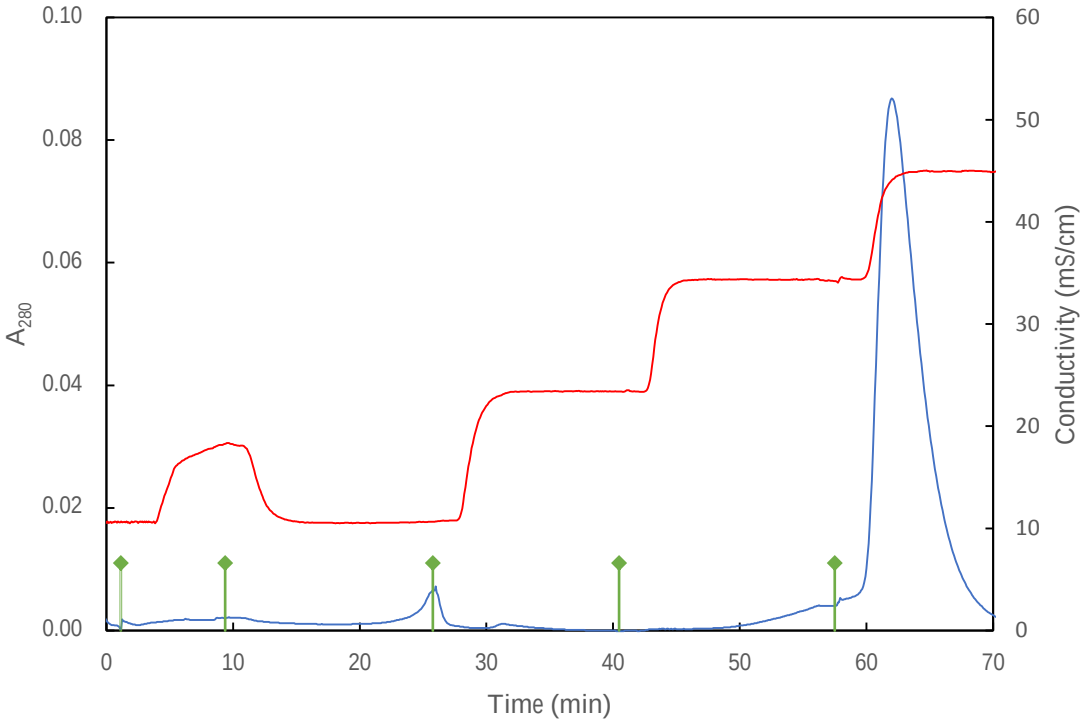

Figure 2.12. Chromatograph for anion exchange purification of p53(1-93) Pro$^-$. The blue line represents absorbance at 280 nm. The red line represents conductivity. The green diamonds are event markers. In chronological order: sample load (1.14 min), addition of Wash Buffer 50 (9.37 min), addition of Wash Buffer 150 (25.77 min), addition of Wash Buffer 250 (20 mM sodium acetate, 250 mM NaCl, pH 4.8; 40.48 min), and addition of Elution Buffer (57.50 min). The peak at 61.99 min is the target protein. An additional wash step was used as this was the first purification of p53(1-93) Pro$^-$ using DEAE media and I was unsure where the protein would elute.

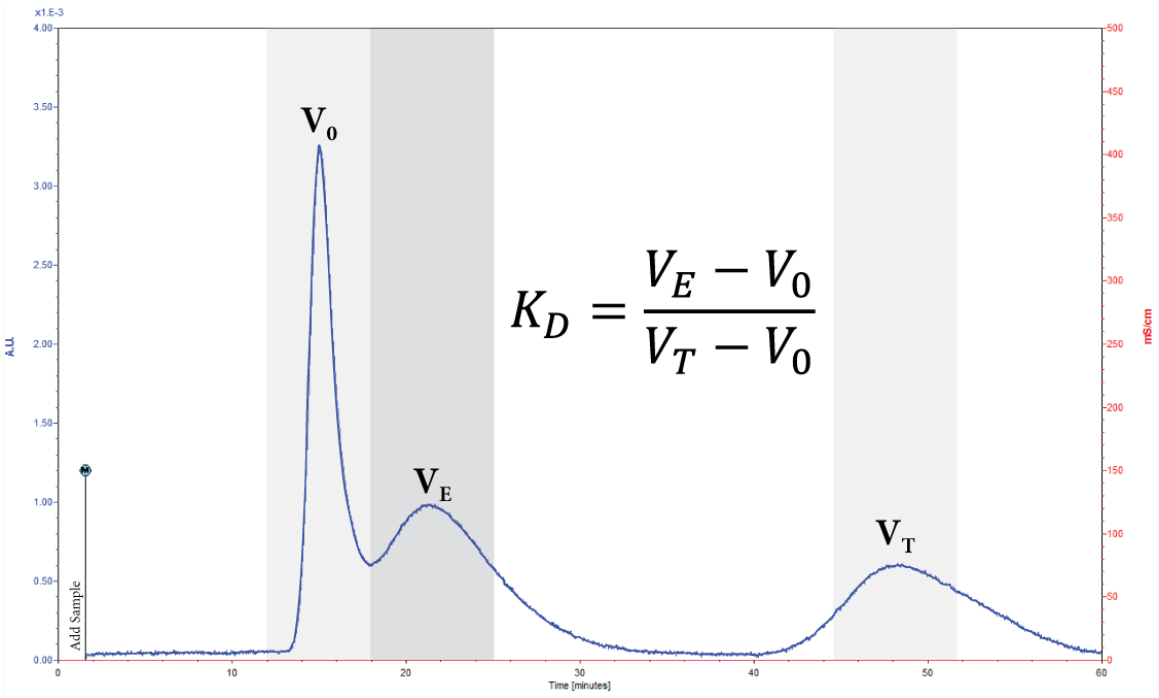

Figure 2.13. Example SEC chromatograph. Biologic LP Data View v1.03 display showing peaks for void volume (blue dextran; V_0), elution volume (WT; V_E), and total volume (DNP-aspartate; V_T). K_D is then calculated according to the equation in the figure. Shaded areas and text added using Adobe Illustrator CS5 for visualization purposes.

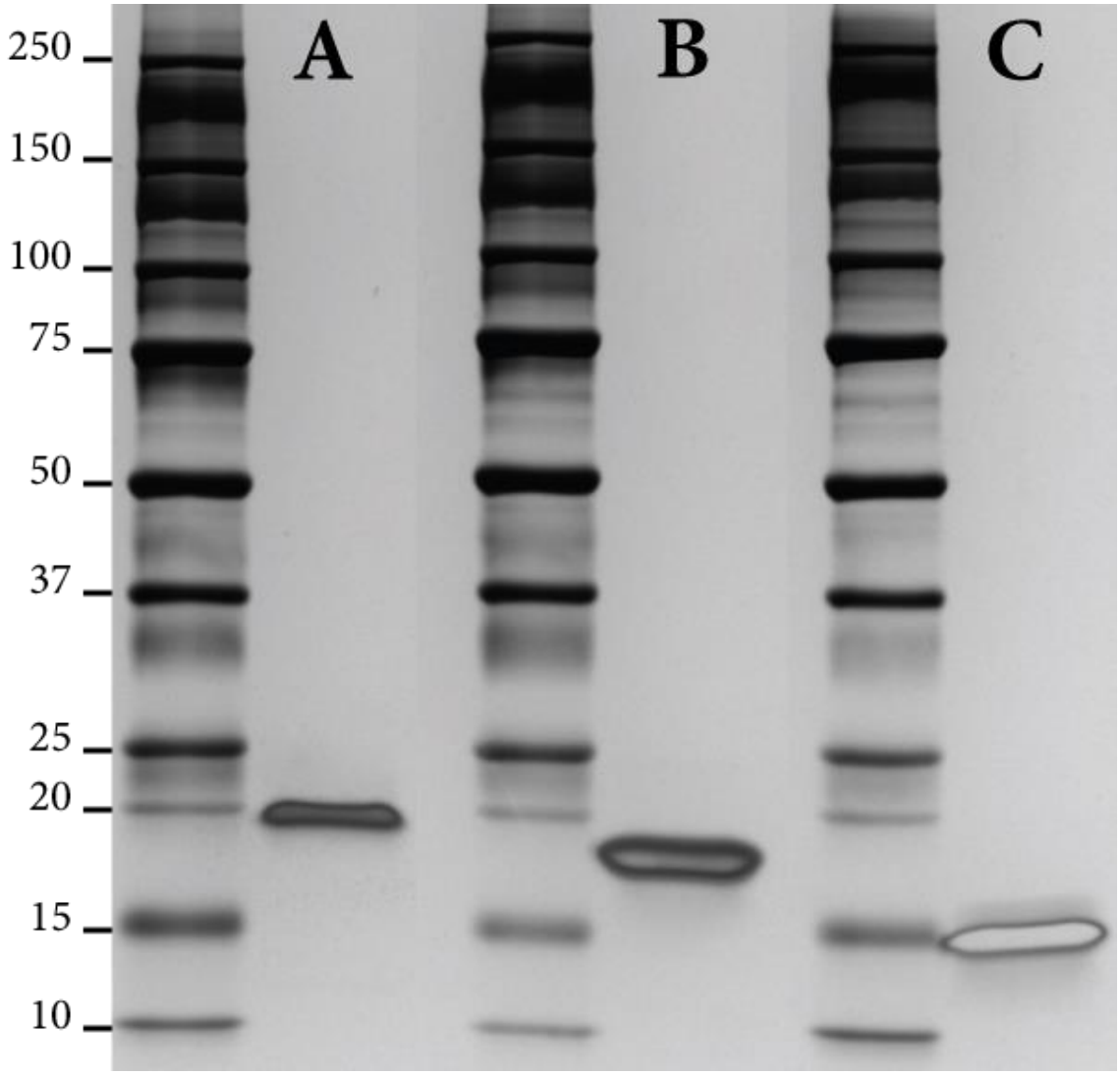

Figure 2.14. Silver stained polyacrylamide gels for each p53(1-93) variant. Each p53(1-93) variant after Ni-NTA and anion exchange purification. The first lane in each gel is Precision Plus protein standards by Bio-Rad. The second lane in each gel is following the p53(1-93) variant: WT (A), Ala^- (B), Pro^- (C). Band distortion and outlining is due to high protein concentrations.

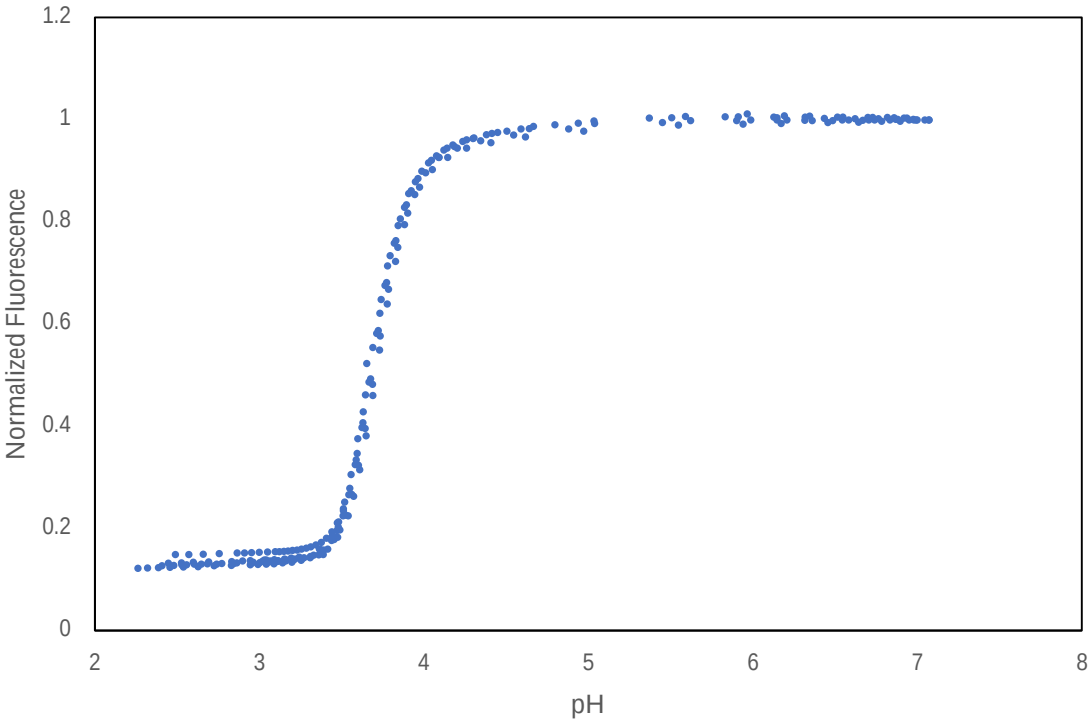

Figure 2.15 Acid/base titration of *Staphylococcal* nuclease by normalized fluorescence. Acid/base titration in 100 mM KCl monitored by fluorescence shows *Staphylococcal* nuclease is stable and folded at pH 4.5. This experiment was performed as previously published[46].

III. RESULTS AND DISCUSSION

3.1 Introduction

Current literature suggests that IDP structures are primarily governed by charge effects[51,52]. Vitalis and Pappu created a continuum solvation model for use in Monte Carlo simulations specifically aimed at studying IDPs[53]. This model, known as ABSINTH for self-Assembly of Biomolecules Studied by an Implicit, Novel, and Tunable Hamiltonian, was used in conjunction with fluorescence to study the charge effects on short disordered peptides (27-49 residues)[51]. The researchers calculated the radius of gyration (R_g) for a set of 21 arginine-rich IDPs with variable net charge per residue (0.21-1.00)[51]. They normalized the data by dividing each calculated R_g by the expected R_g of a random coil, then plotted that ratio vs. net charge per residue[51]. The data showed a positive logarithmic relationship between ΔR_g and net charge per residue.

Schuler et al. took a different approach to showing the prevalence of charge effects on IDP structure by using single-molecule Förster resonance energy transfer (FRET) to measure R_g[52]. Schuler's group chose three model proteins with vastly different charge states: the stably folded CspTm (net charge -2), the intrinsically disordered N-terminal region of HIV-1 integrase (net charge -4), and the entirely disordered prothymosin α (net charge -27)[52]. They measured R_g via FRET at increasing concentrations of urea and guanidinium for each protein and found that the denaturant-induced expansion of both

IDPs correlated with net charge[52]. Prothymosin α, which already has an expanded structure due to high net charge, experienced less denaturant-induced expansion than the more neutral HIV-1 integrase.

While both of these studies show a correlation between net charge and hydrodynamic properties, neither of them factored in PP_{II} propensity. As mentioned in Chapter I, we designed this project to examine if and how charge effects are modulated by PP_{II} structure. To test this, we used pH as a mechanism to control net charge and proline-to-glycine mutations as a way to control PP_{II} structure.

3.2 Size Exclusion Chromatography Results

3.2.1 SEC Results at pH 7.0

Going in to this study, we expected all three peptides to undergo significant decreases in size as pH was decreased. At a pH of 7.0, most glutamate and aspartate residues are in the deprotonated state. As the pH is lowered, free protons bind to the carboxylic acid side chains neutralizing their negative charges. As more acid residues are protonated, the polypeptide should experience significantly less charge repulsion and the average ensemble R_h should decrease.

SEC results at pH 7.0 showed p53(1-93) WT had an R_h of 31.8 Å with Ala^- at 30.0 Å and Pro^- at 27.4 Å. These results are summarized in Table 3.1 and Figure 3.1 below.

These results were as expected and matched those previously reported by Perez et al.[35] (Table 3.2). Since all three peptides have the same 17 charged side chains, N-terminal residue, and C-terminal residue, they have the same theoretical pK_a. Thus all three peptides have a theoretical charge of -14.99 at pH 7.0 based on the Henderson-Hasselbalch equation[54,55] (Figure 3.2) and the amino acid pK_as as reported by the Handbook of Chemistry and Physics[56].

Table 3.1. SEC results at pH 7.0.

Protein	$K_{D, mean}$	$K_{D, sd}$	M.W. (Da)	R_h (Å)
Chicken egg albumin	0.112	0.002	42,861.3	-
p53(1-93) WT	0.182	0.007	10,123.2	31.8
p53(1-93) Ala^-	0.218	0.002	9,954.9	30.0
Bovine carbonic anhydrase	0.265	0.002	28,982.5	-
p53(1-93) Pro^-	0.267	0.009	9,241.8	27.4
Staphylococcal nuclease	0.359	0.002	16,811.2	-
Horse heart myoglobin	0.398	0.003	16,951.4	-

Table 3.2. Comparison of SEC results to previously published work.

Protein	$R_{h, actual}$ (Å)	$R_{h, published}$ (Å)	Pct. Diff.
p53(1-93) WT	31.8	31.95	0.47%
p53(1-93) Ala^-	30.0	29.65	1.17%
p53(1-93) Pro^-	27.4	27.36	0.15%

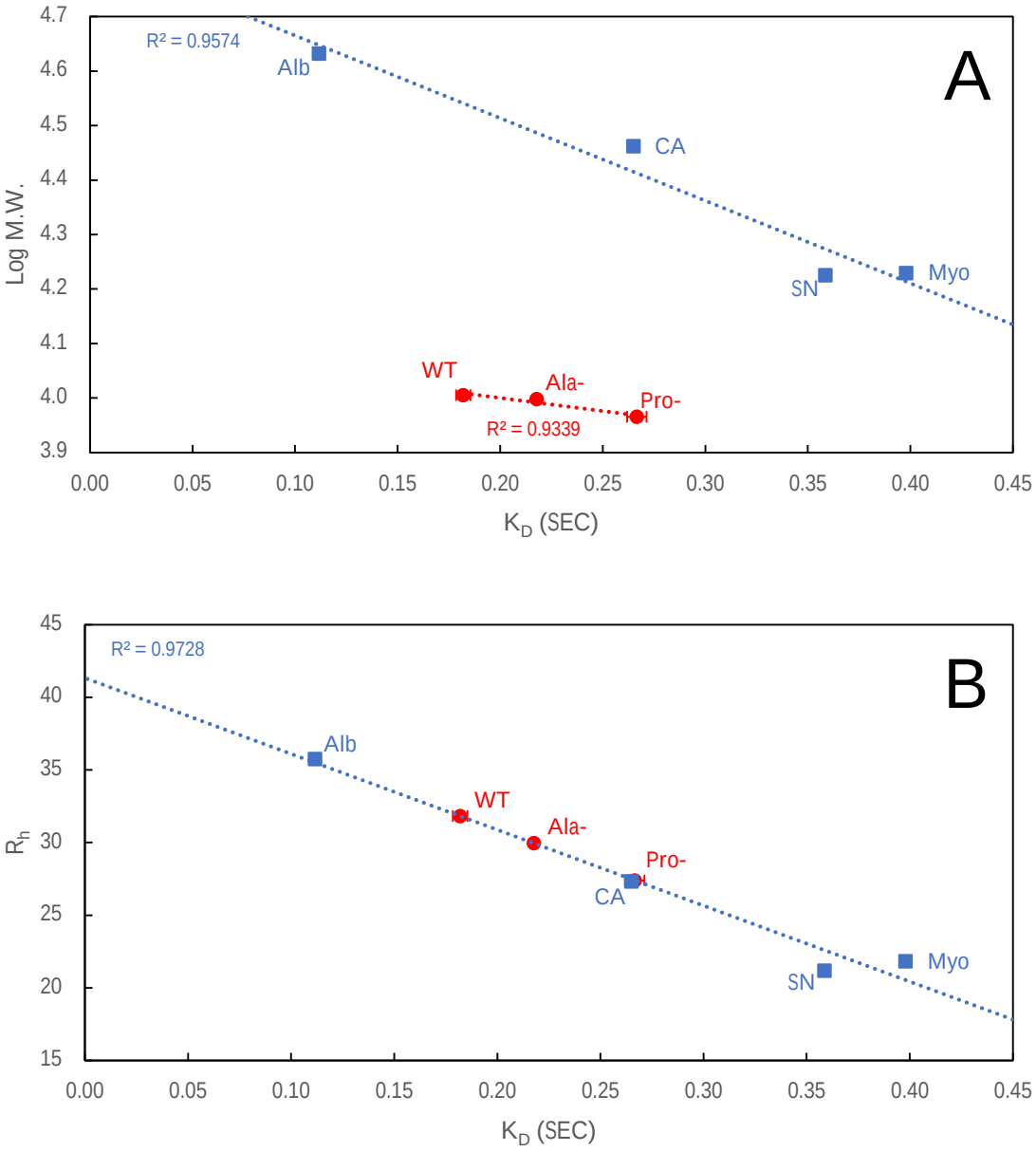

Figure 3.1. SEC results at pH 7.0. Blue squares are folded protein standards. Red circles are p53(1-93) variants. (A) Log M.W. based on primary sequence vs. K_D shows separate linear correlations for both folded and disordered proteins. (B) R_hs for disordered proteins were extrapolated from folded protein crystal structures and the folded protein regression line.

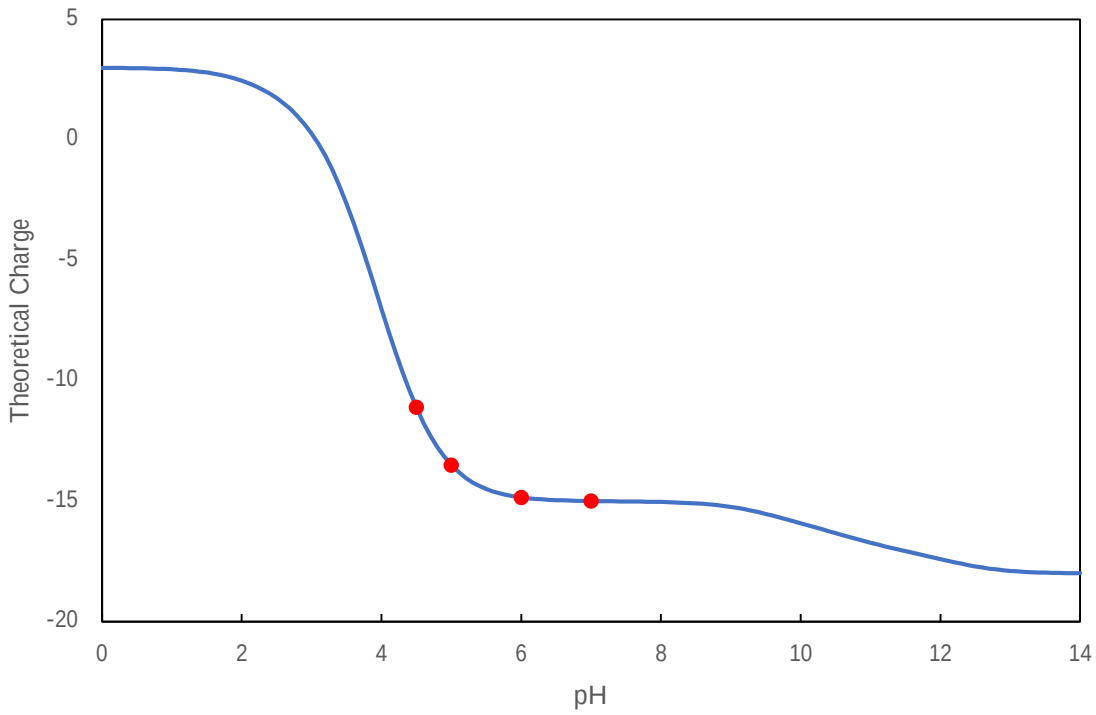

Figure 3.2. Theoretical charge for p53(1-93) variants at each pH. The blue line represents the theoretical pH of each p53(1-93) variant based on the Henderson-Hasselbalch equation and the pK_as from the Handbook of Chemistry and Physics. The red circles are the pHs used in this study.

3.2.2 SEC Results at pH 6.0

SEC results at pH 6.0 showed p53(1-93) WT had an R_h of 31.2 Å with Ala^- at 30.3 Å and Pro^- at 27.1 Å. These results are summarized in Table 3.3 and Figure 3.3 below.

At pH 6.0, the theoretical charge increases to -14.83. This minor charge in theoretical charge correlates well with the small change in R_h for each peptide. The 0.3 Å increase in R_h for Ala^- is most likely a function of small sample size.

Table 3.3. SEC results at pH 6.0.

Protein	$K_{D, mean}$	$K_{D, sd}$	M.W. (Da)	R_h (Å)
Chicken egg albumin	0.106	0.005	42,861.3	-
p53(1-93) WT	0.189	0.002	10,123.2	31.2
p53(1-93) Ala^-	0.206	0.003	9,954.9	30.3
Bovine carbonic anhydrase	0.262	0.004	28,982.5	-
p53(1-93) Pro^-	0.269	0.003	9,241.8	27.1
Staphylococcal nuclease	0.354	0.002	16,811.2	-
Horse heart myoglobin	0.396	0.007	16,951.4	-

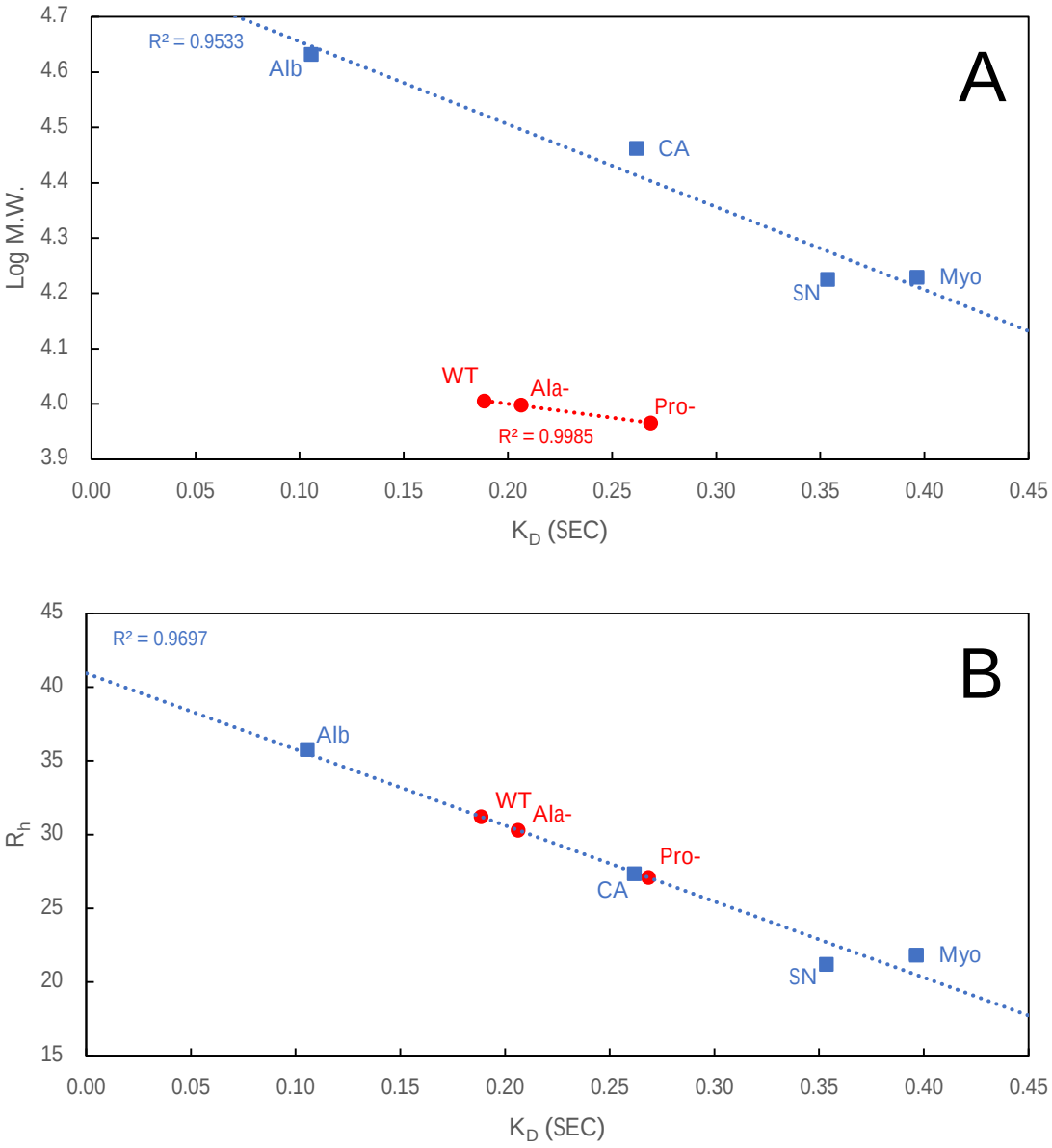

Figure 3.3. SEC results at pH 6.0. Blue squares are folded protein standards. Red circles are p53(1-93) variants. (A) Log M.W. based on primary sequence vs. K_D shows separate linear correlations for both folded and disordered proteins. (B) R_hs for disordered proteins were extrapolated from folded protein crystal structures and the folded protein regression line.

3.2.3 SEC Results at pH 5.0

SEC results at pH 5.0 showed p53(1-93) WT had an R_h of 30.0 Å with Ala^- at 28.2 Å and Pro^- at 24.5 Å. These results are summarized in Table 3.4 and Figure 3.4 below.

Table 3.4. SEC results at pH 5.0.

Protein	$K_{D, mean}$	$K_{D, sd}$	M.W. (Da)	R_h (Å)
Chicken egg albumin	0.123	0.009	42,861.3	-
p53(1-93) WT	0.217	0.003	10,123.2	30.0
p53(1-93) Ala^-	0.249	0.001	9,954.9	28.2
Bovine carbonic anhydrase	0.272	0.001	28,982.5	-
p53(1-93) Pro^-	0.319	0.001	9,241.8	24.5
Staphylococcal nuclease	0.324	0.001	16,811.2	-
Horse heart myoglobin	0.406	0.001	16,951.4	-

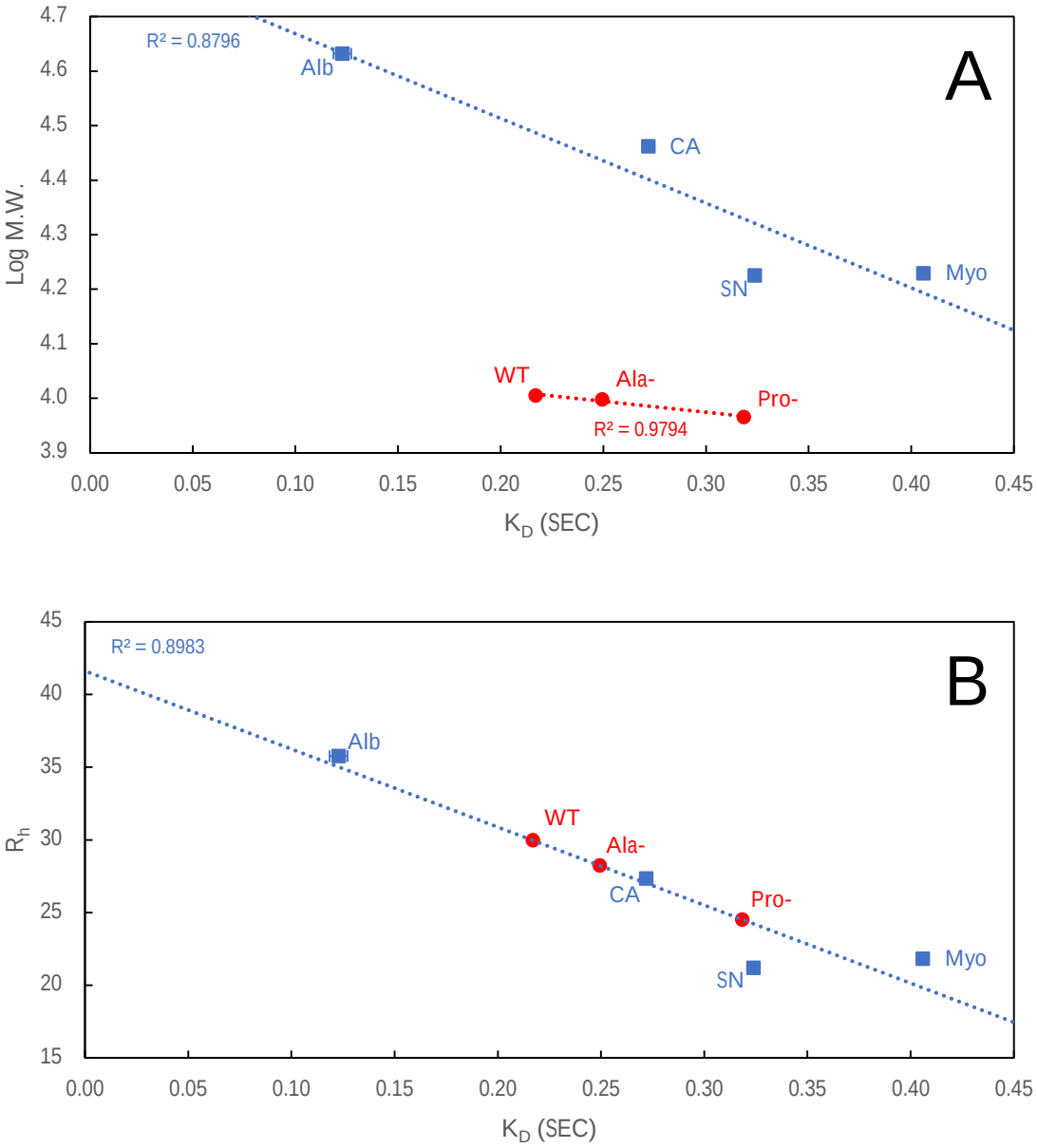

Figure 3.4. SEC results at pH 5.0. Blue squares are folded protein standards. Red circles are p53(1-93) variants. (A) Log M.W. based on primary sequence vs. K_D shows separate linear correlations for both folded and disordered proteins. (B) R_hs for disordered proteins were extrapolated from folded protein crystal structures and the folded protein regression line.

3.2.4 SEC Results at pH 4.5

SEC results at pH 4.5 showed p53(1-93) WT had an Rh of 27.1 Å with Ala^- at 25.7 Å and Pro^- at 20.0 Å. These results are summarized in Table 3.5 and Figure 3.5 below. As mentioned previously, horse heart myoglobin was excluded as a folded protein standard at this pH due to instability.

Table 3.5. SEC results at pH 4.5.

Protein	$K_{D, mean}$	$K_{D, sd}$	M.W. (Da)	R_h (Å)
Chicken egg albumin	0.137	0.005	42,861.3	-
p53(1-93) WT	0.28	0.004	10,123.2	27.1
Bovine carbonic anhydrase	0.300	0.002	28,982.5	-
p53(1-93) Ala^-	0.303	0.003	9,954.9	25.7
Staphylococcal nuclease	0.357	0.003	16,811.2	-
p53(1-93) Pro^-	0.393	0.004	9,241.8	20.0

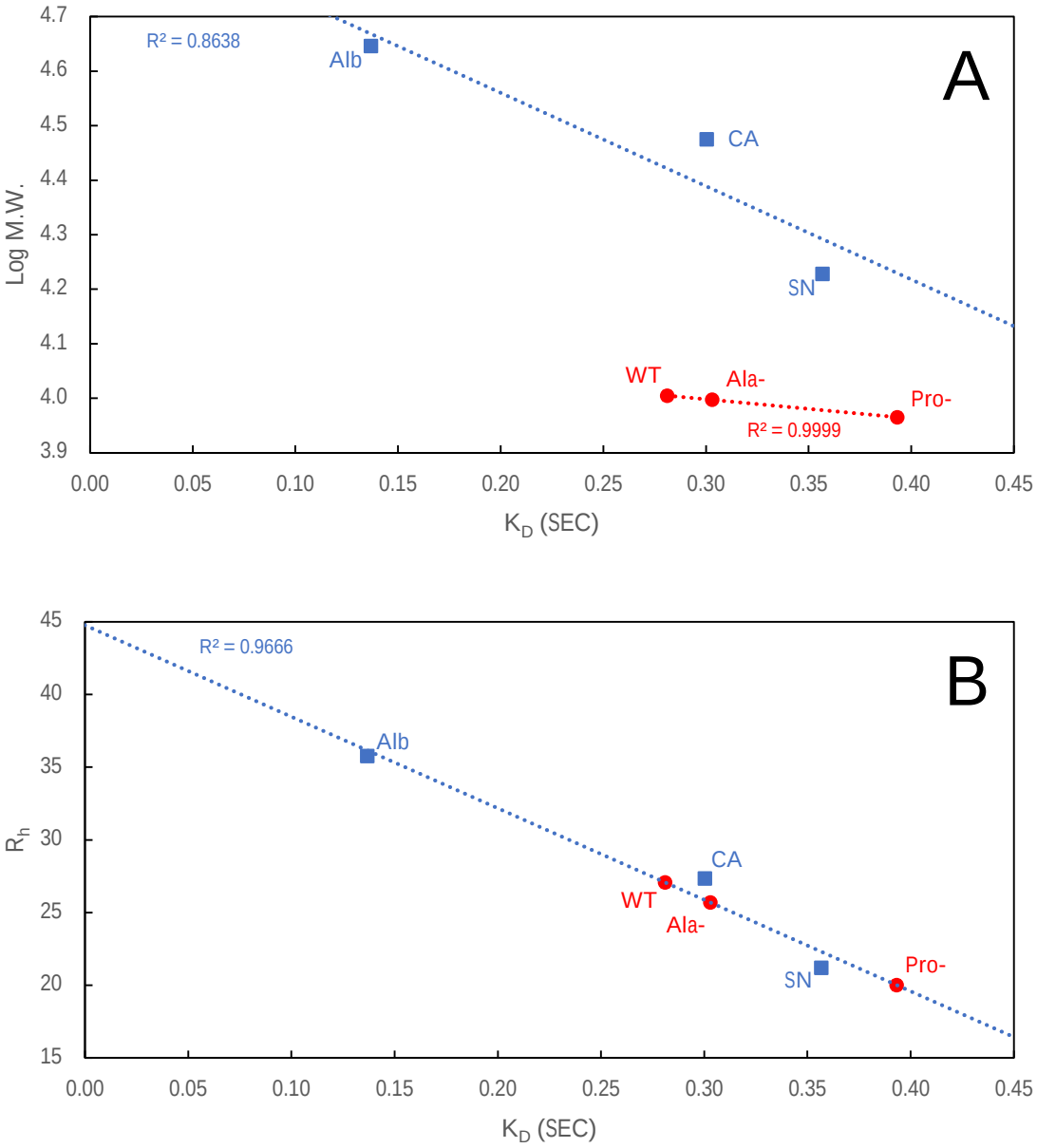

Figure 3.5. SEC results at pH 4.5. Blue squares are folded protein standards. Red circles are p53(1-93) variants. (A) Log M.W. based on primary sequence vs. K_D shows separate linear correlations for both folded and disordered proteins. (B) R_hs for disordered proteins were extrapolated from folded protein crystal structures and the folded protein regression line.

3.3 Circular Dichroism Spectroscopy Results

3.3.1 CD Results for p53(1-93) WT

CD results at pH 7.0 matched those previously published[3,35]. The WT peptide had a local maxima at 221 nm as is characteristic of PP_{II} secondary structure[41–43]. The CD spectra showed this peak was temperature dependent as reported by Schaub et al[3]. A clear isochromatic point is seen at 209.5 nm which is generally indicative of a non-two-state transition[57]. The linear temperature dependence of the 221 nm peak suggests that the thermal unfolding is noncooperative[3].

CD results at pH 6.0 were nearly identical to those at pH 7.0 due to the negligible change in theoretical charge. The isochromatic point shifted slightly to 210 nm.

At pH 5.0, a noticeable change in the spectra is seen due to a decrease in PP_{II} content. The peak at 221 nm is significantly reduced. The isochromatic point shifted slightly to 210.5 nm. A similar effect is seen at pH 4.5 although much more prominent. The 221 nm peak is even flatter and the isochromatic point once again is slightly redshifted to 211.5 nm.

The overall pH dependence can be seen by plotting the 221 nm peak height verses temperature. The height of the 221 nm peak relative to the CD value at 85°C decreases as a function of pH. This summary is shown in Figure 3.6 below. Complete CD spectra for p53(1-93) WT at all four pHs are shown in Figures 3.7 – 3.10 at the end of this section.

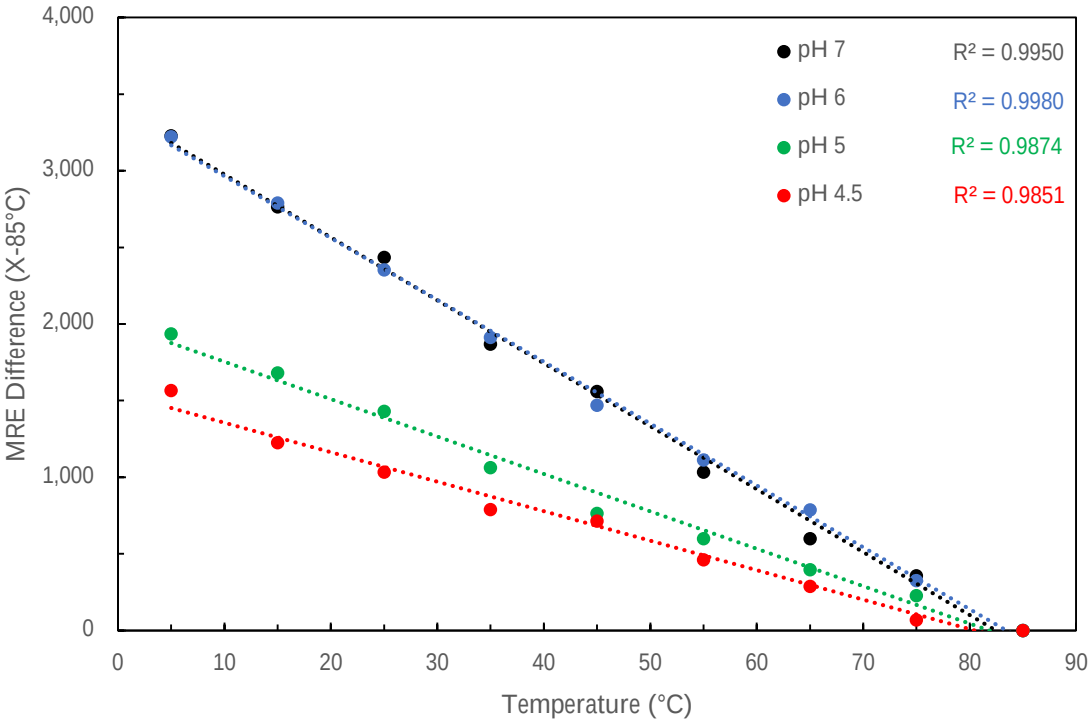

Figure 3.6. Summary of CD results for p53(1-93) WT . The molar residual ellipticity (MRE) difference was generated by averaging the 220-222 nm readings for each temperature, then subtracting the 85°C value from each reading. This was repeated for each pH: 7.0 (black), 6.0 (blue), 5.0 (green), 4.5 (red). The slope of the regression line flattens as a function of pH.

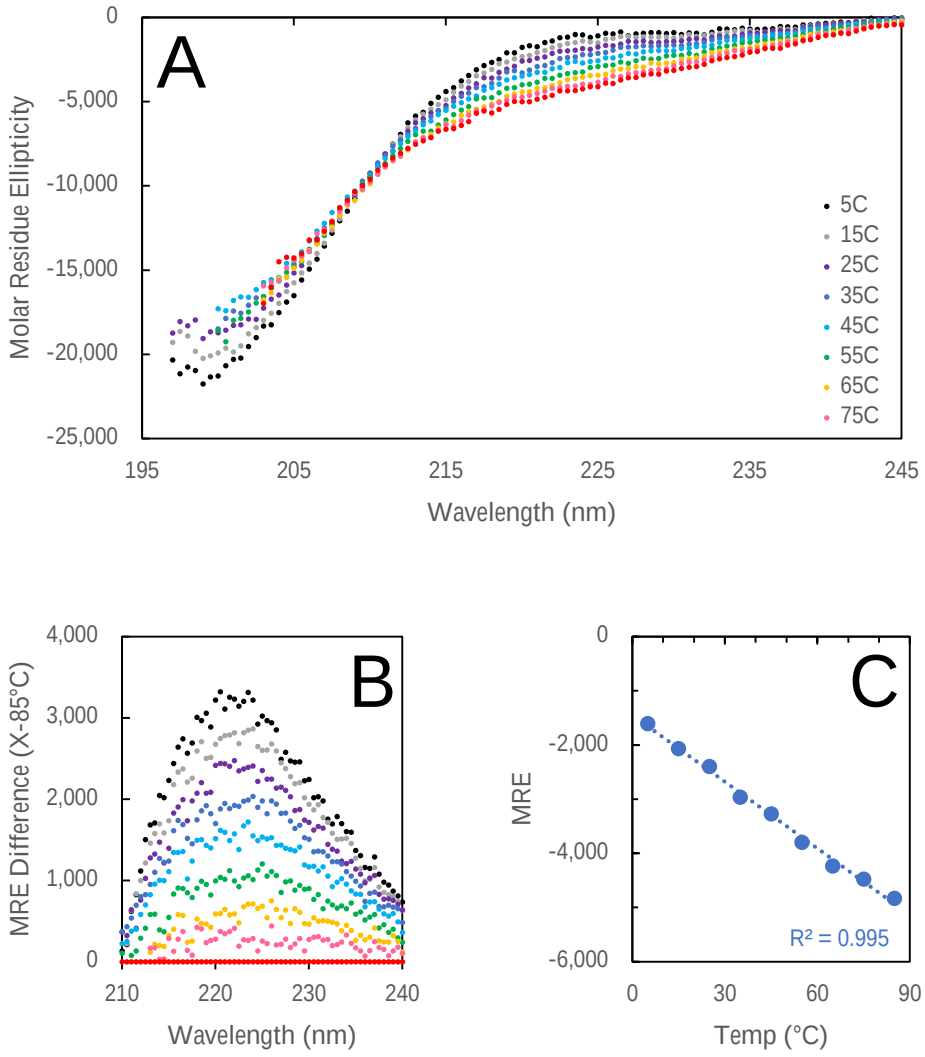

Figure 3.7. CD results for p53(1-93) WT at pH 7.0. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

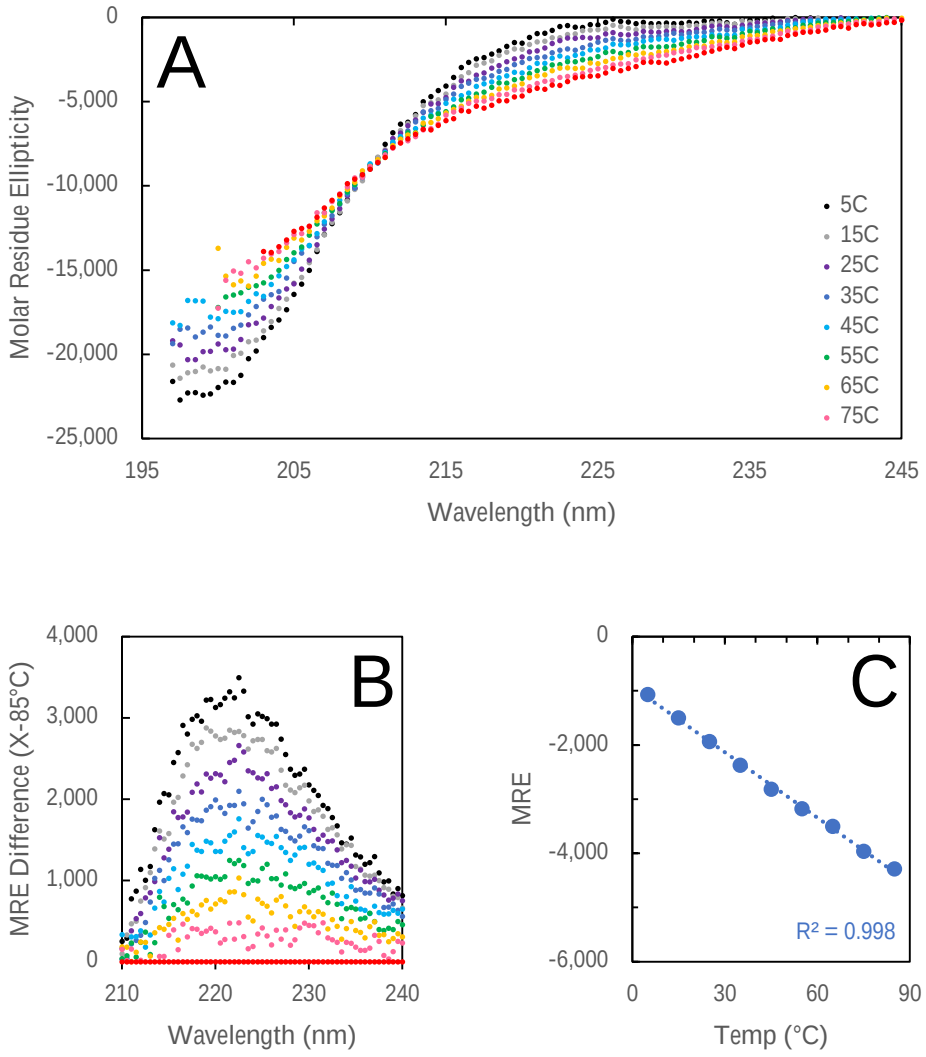

Figure 3.8. CD results for p53(1-93) WT at pH 6.0. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

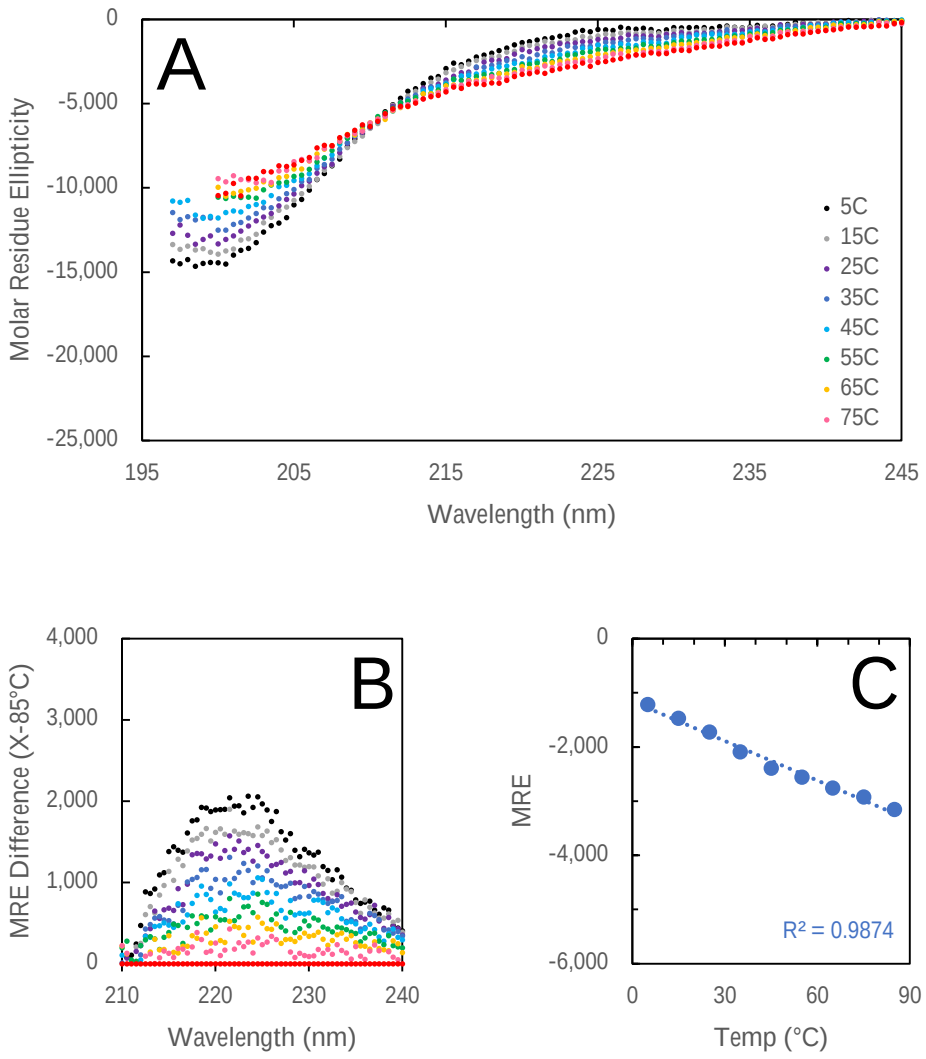

Figure 3.9. CD results for p53(1-93) WT at pH 5.0. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

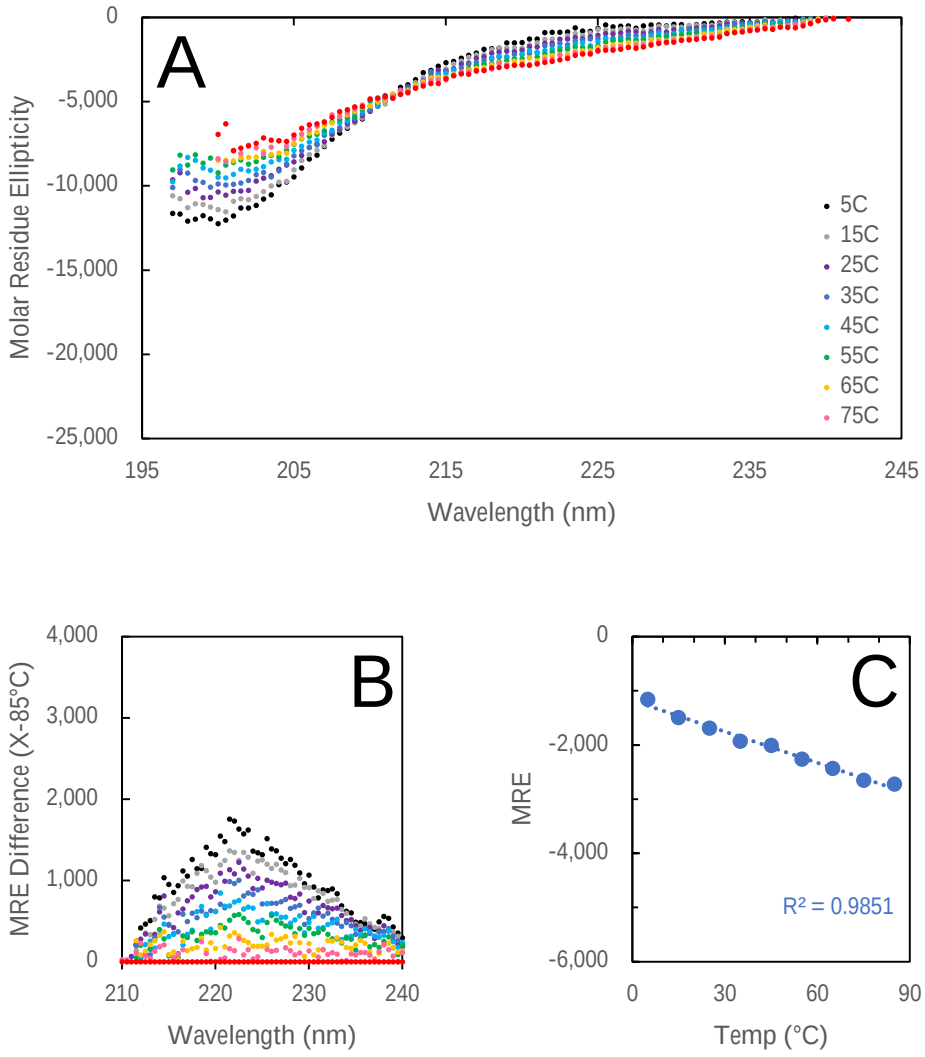

Figure 3.10. CD results for p53(1-93) WT at pH 4.5. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

3.3.2 CD Results for p53(1-93) Ala⁻

CD results at pH 7.0 matched those previously published[3,35]. The Ala⁻ protein had the same temperature dependent local maxima at 221 nm as seen in the WT but with decreased magnitude. The isochromatic point was 209.5 nm.

CD results at pH 6.0 were similar in shape to those at pH 7.0, although the magnitude of the 221 nm peak decreased more than expected. Due to theoretical charge, I expected the 221 nm peak to be virtually indistinguishable from that at pH 7.0. The isochromatic point was 210.5 nm.

The results at pH 5.0 were similar to pH 6.0, which was surprising. At pH 5.0, a much larger percentage of side chains should be protonated compared to pH 6.0, thus causing a larger decrease in PP_{II} content than observed. The isochromatic point was 210.5.

The trend continues at pH 4.5 with even more flattening of the 221 nm peak. The isochromatic point is one again slightly redshifted to 211 nm.

The overall pH dependence is seen by plotting the 221 nm peak height verses temperature. Just like the WT, the height of the 221 nm peak relative to the CD value at 85°C decreases as a function of pH. This summary is shown in Figure 3.11 below. Complete CD spectra for p53(1-93) Ala⁻ at all four pHs are shown in Figures 3.12 – 3.15 at the end of this section.

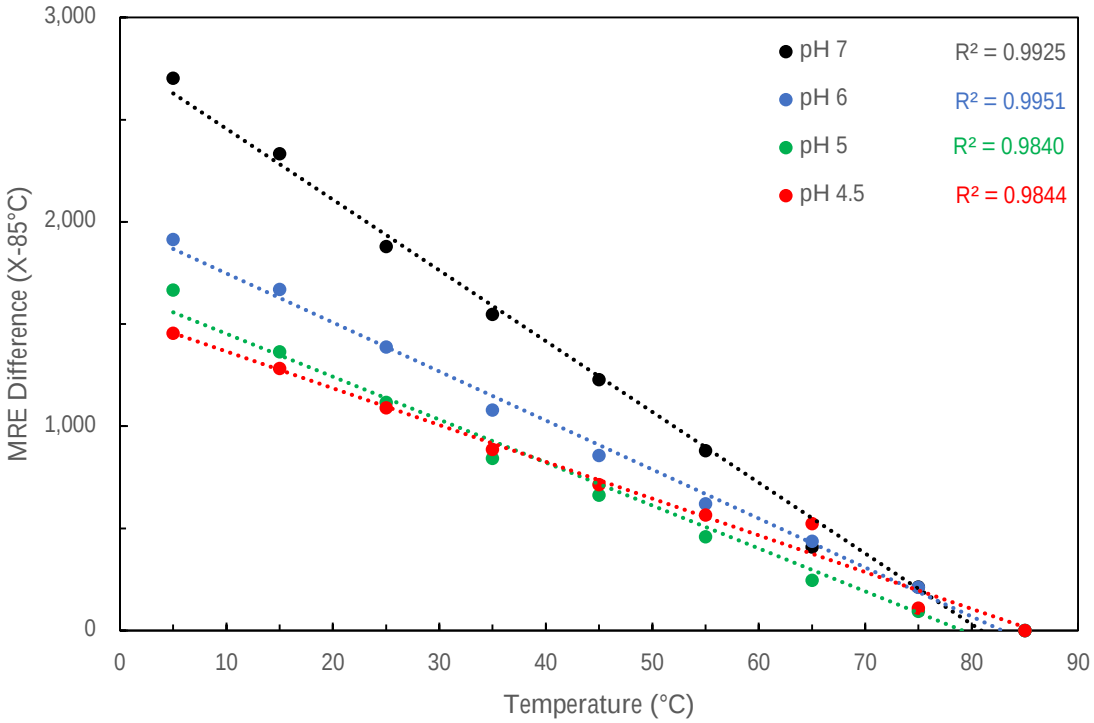

Figure 3.11. Summary of CD results for p53(1-93) Ala^{-}. The molar residual ellipticity (MRE) difference was generated by averaging the 220-222 nm readings for each temperature, then subtracting the 85°C value from each reading. This was repeated for each pH: 7.0 (black), 6.0 (blue), 5.0 (green), 4.5 (red). The slope of the regression line flattens as a function of pH.

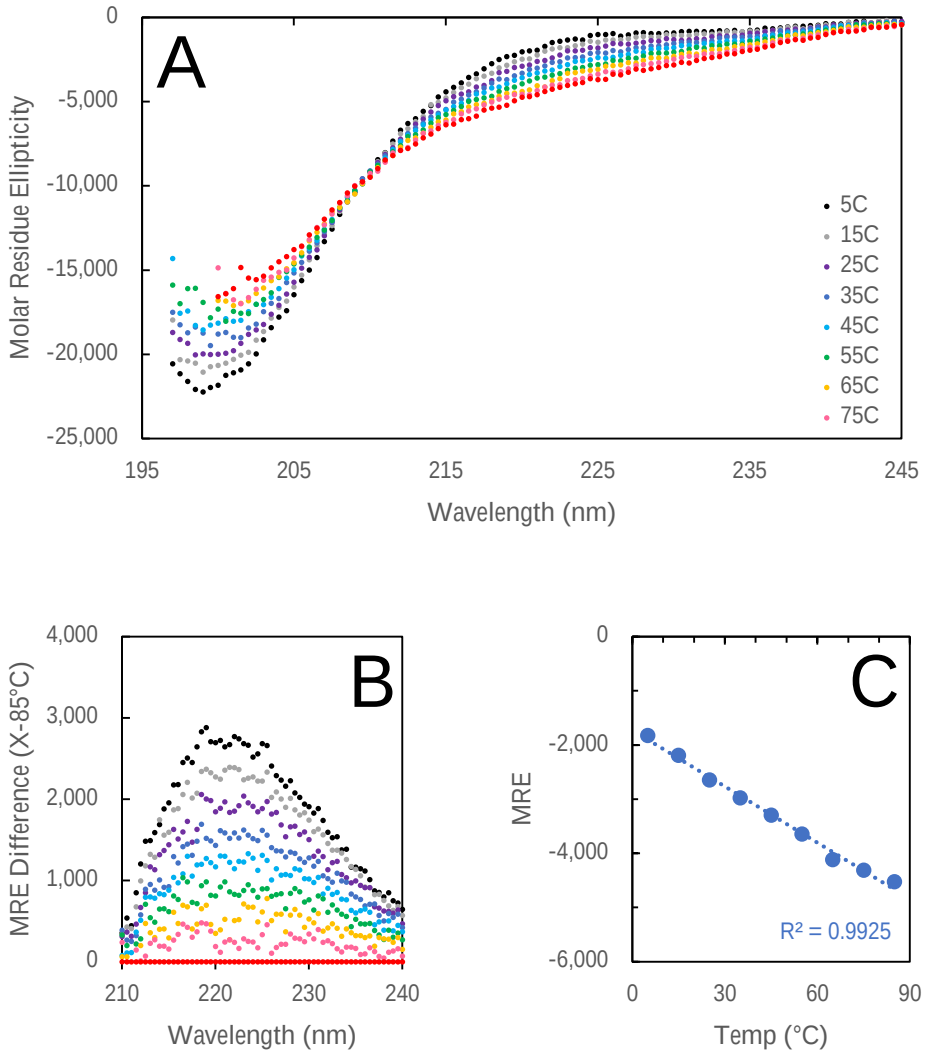

Figure 3.12. CD results for p53(1-93) Ala^- at pH 7.0. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

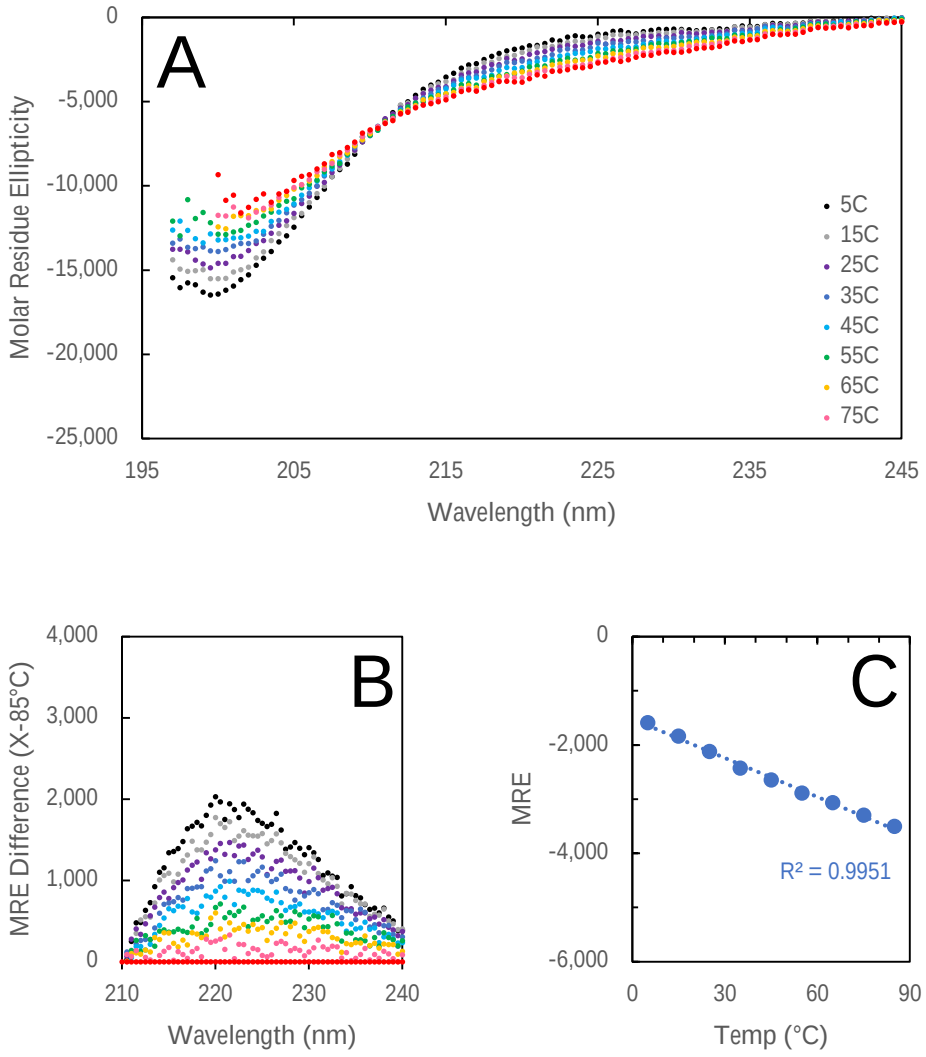

Figure 3.13. CD results for p53(1-93) Ala⁻ at pH 6.0. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

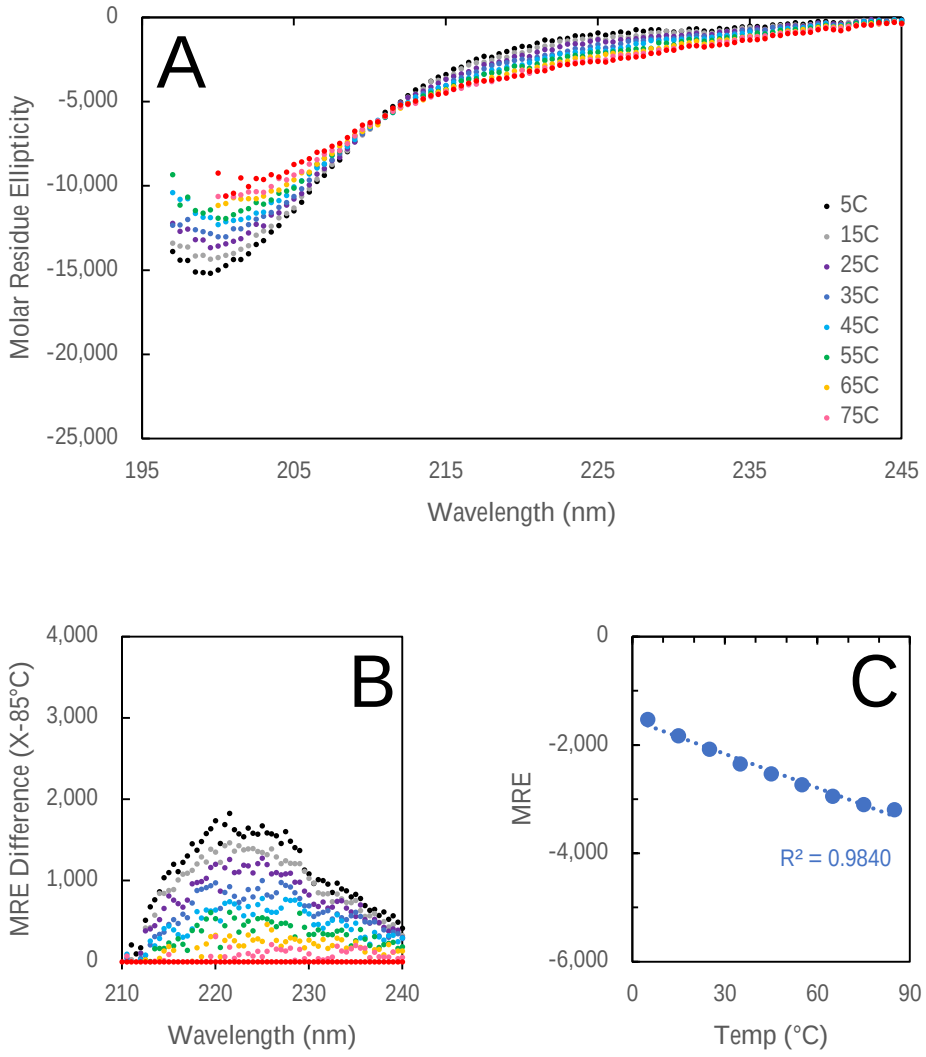

Figure 3.14. CD results for p53(1-93) Ala^- at pH 5.0. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

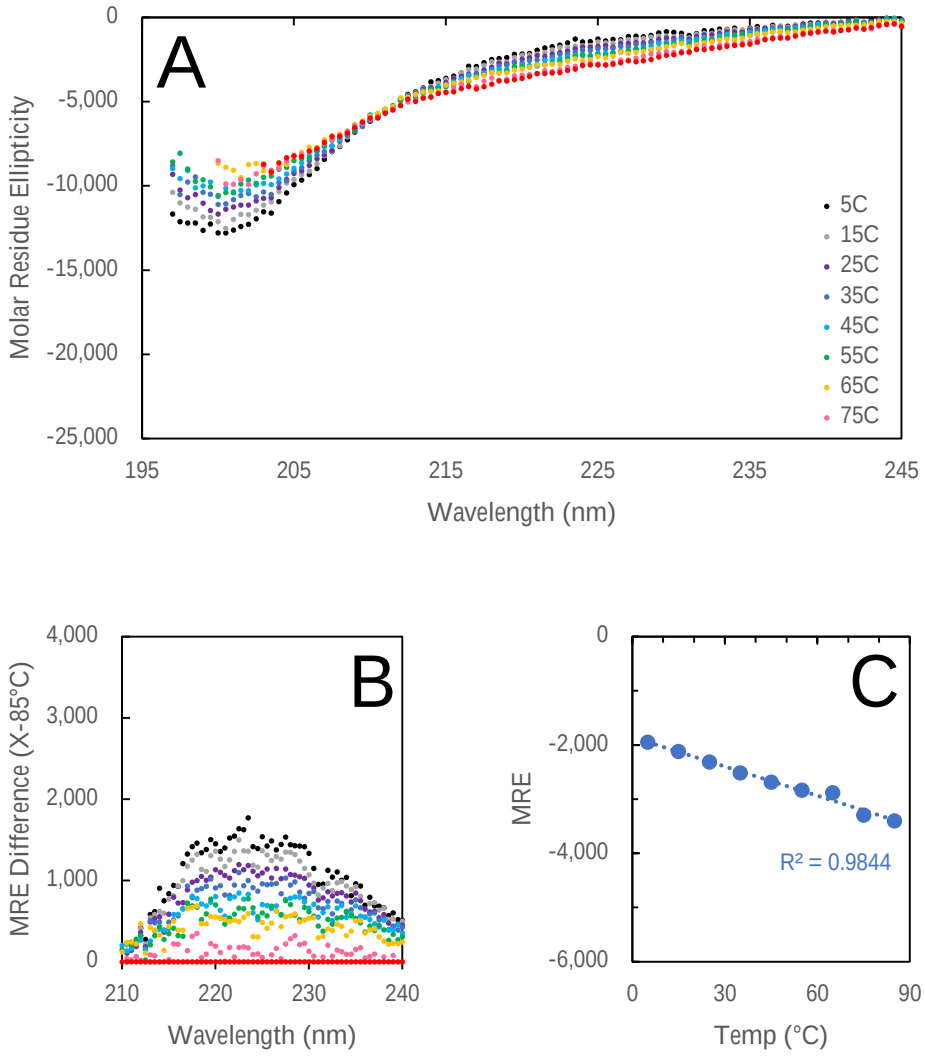

Figure 3.15. CD results for p53(1-93) Ala⁻ at pH 4.5. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

3.3.3 CD Results for p53(1-93) Pro⁻

CD results at pH 7.0 matched those previously published[3,35]. The Pro- protein had the same temperature dependent local maxima at 221 nm as seen in the WT and Ala⁻, but with much decreased magnitude. The isochromatic point was 211.0 nm.

CD results at pH 6.0 were nearly indistinguishable from those at pH 7.0 due to the negligible change in theoretical charge as we move from pH 7.0 to 6.0. The isochromatic point was 206.5 nm.

The results at pH 5.0 were as expected with a similar shape to those at pH 7.0 and 6.0, but with much smaller magnitude of the temperature-dependent 221 nm peak. The isochromatic point was 207.5 nm.

The CD results at pH 4.5 were the most surprising of any in this entire study. The overall shape was similar to those at high pHs, the spectra was completely devoid of temperature dependence. There was no clear isochromatic point.

The overall pH dependence is seen by plotting the 221 nm peak height verses temperature. Just like the WT Ala⁻, the height of the 221 nm peak relative to the CD value at 85°C decreases as a function of pH. However, unlike the other two polypeptides, this temperature dependent peak is completely missing at pH 4.5. This summary is shown in Figure 3.16 below. Complete CD spectra for p53(1-93) Ala⁻ at all four pHs are shown in Figures 3.17 – 3.21 at the end of this section.

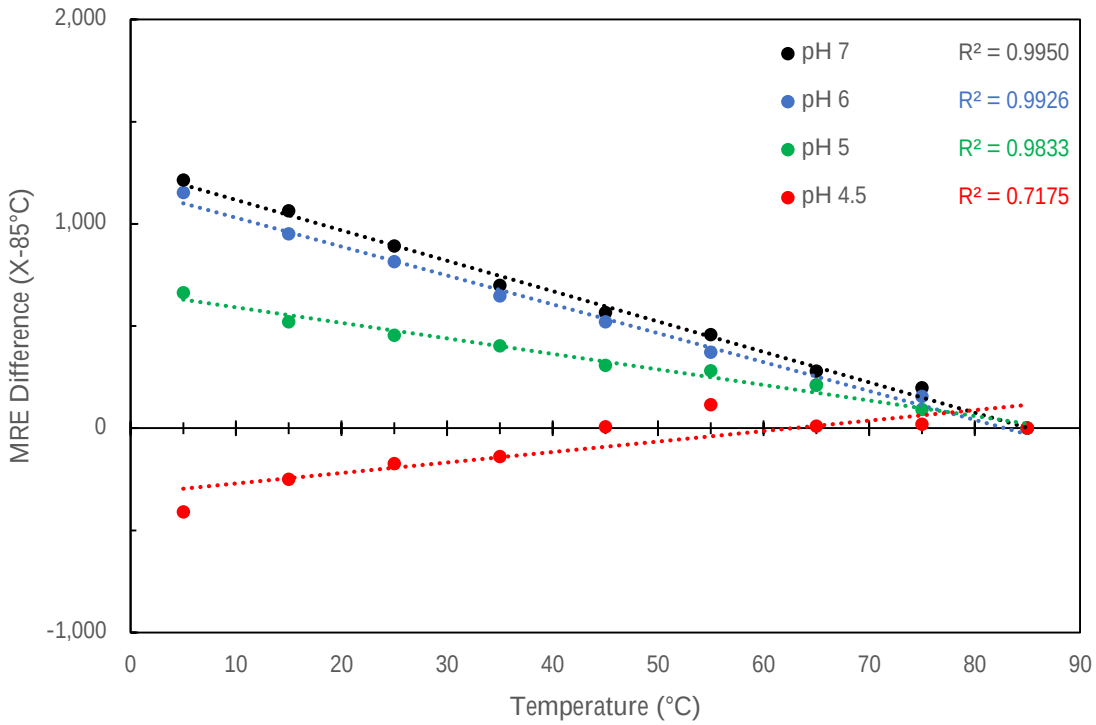

Figure 3.16. Summary of CD results for p53(1-93) Pro^-. The molar residual ellipticity (MRE) difference was generated by averaging the 220-222 nm readings for each temperature, then subtracting the 85°C value from each reading. This was repeated for each pH: 7.0 (black), 6.0 (blue), 5.0 (green), 4.5 (red). The slope of the regression line flattens as a function of pH.

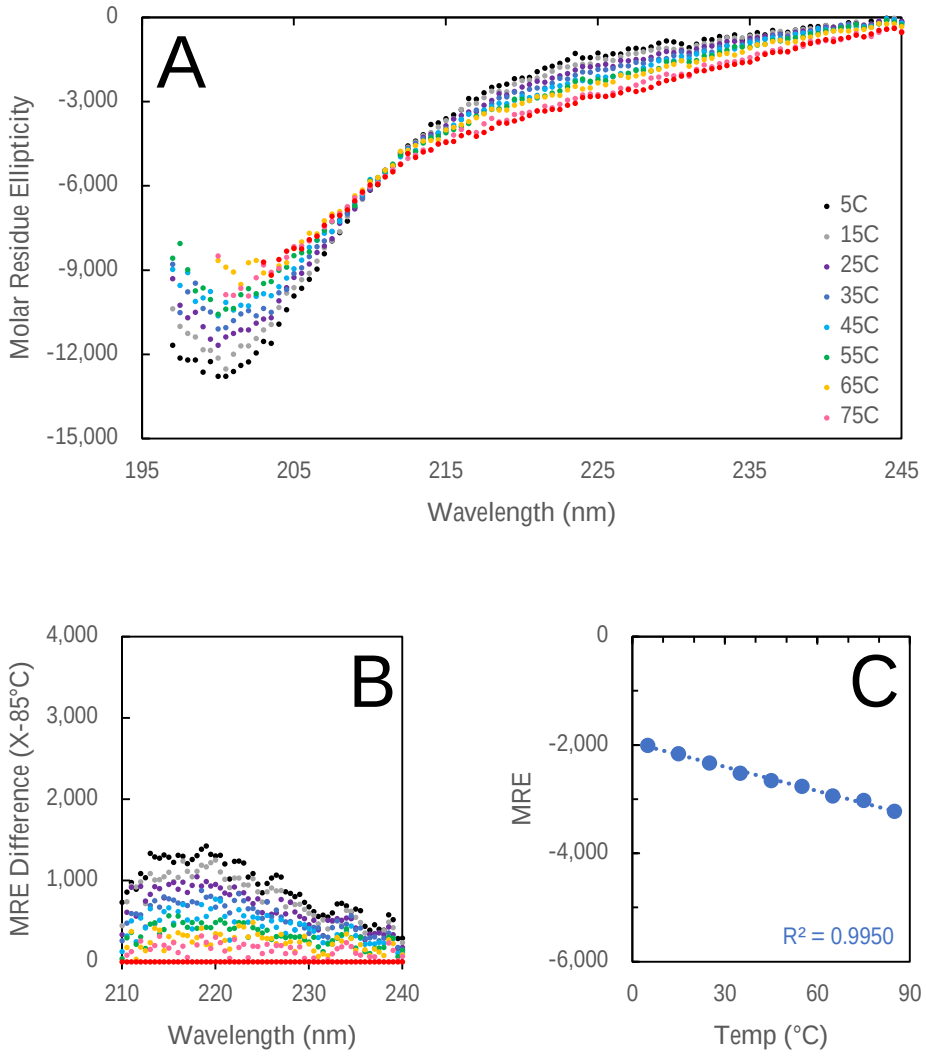

Figure 3.17. CD results for p53(1-93) Pro^- at pH 7.0. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

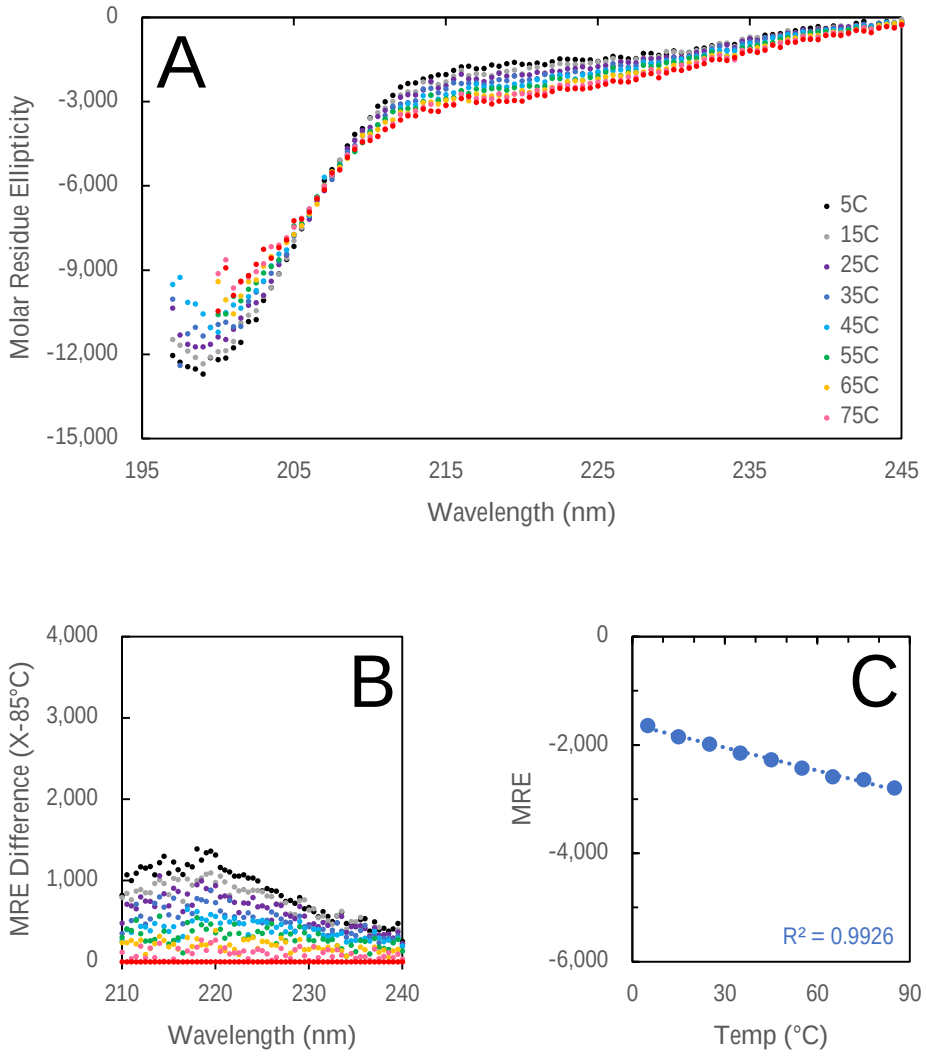

Figure 3.18. CD results for p53(1-93) Pro^- at pH 6.0. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

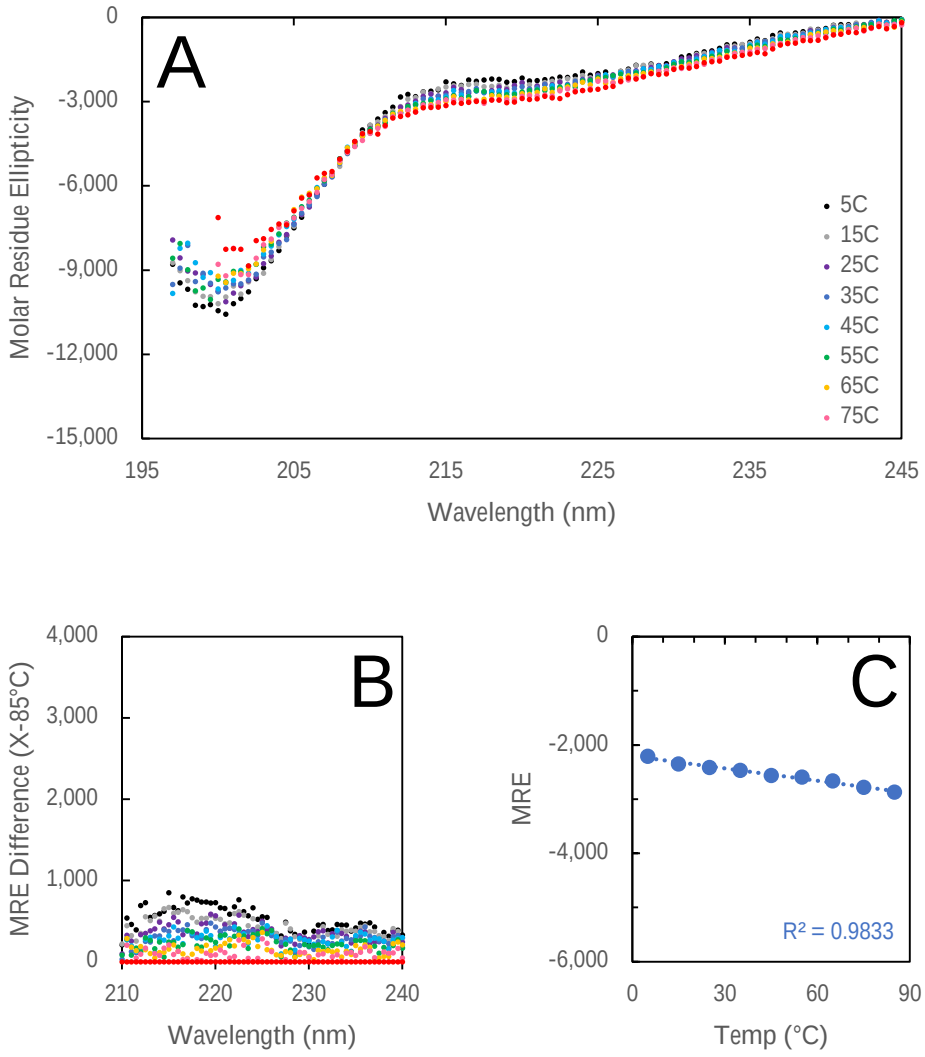

Figure 3.19. CD results for p53(1-93) Pro$^-$ at pH 5.0. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

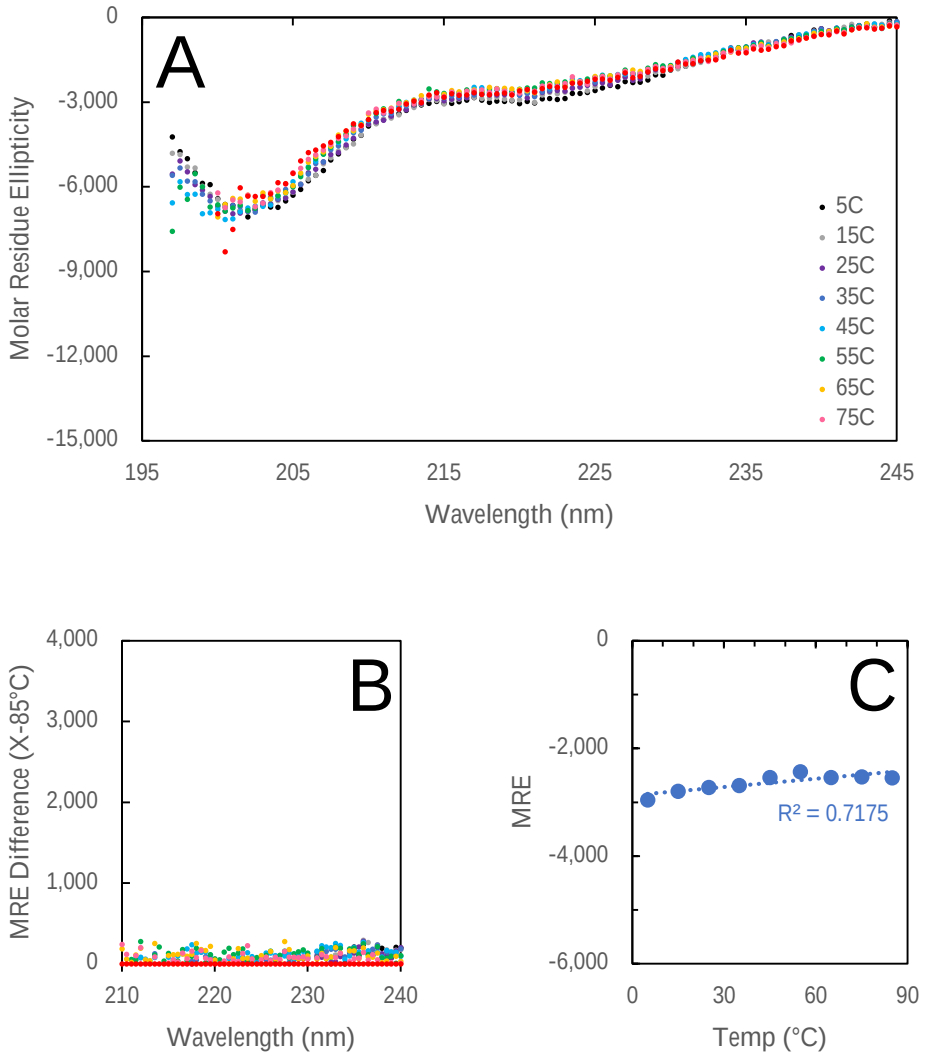

Figure 3.20. CD results for p53(1-93) Pro$^-$ at pH 4.5. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. Part (C) shows the temperature dependence of this local maximum by plotting the average MRE value from 220-222 nm vs. each temperature.

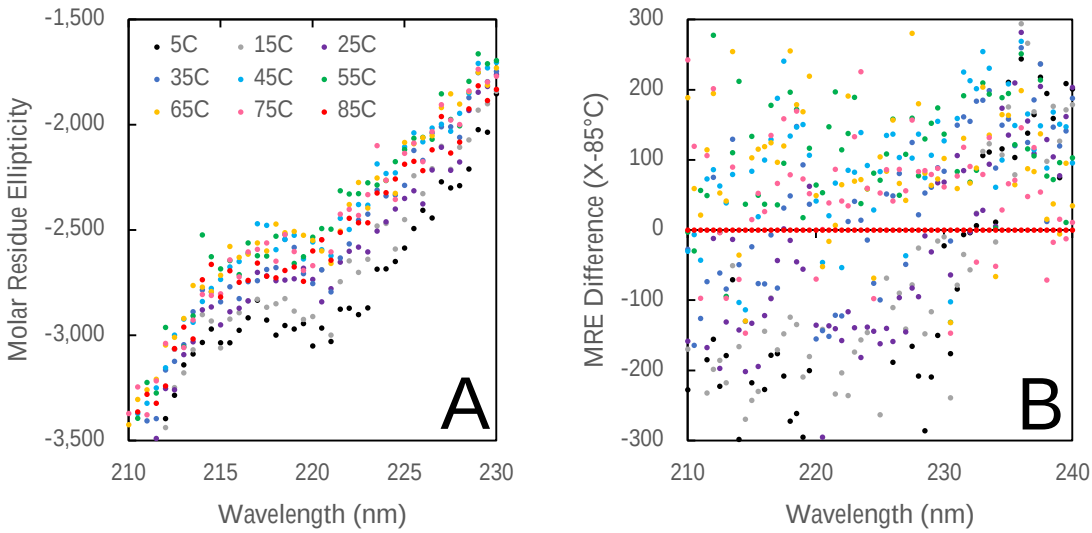

Figure 3.21. Zoomed in view of CD results for p53(1-93) Pro^- at pH 4.5. This figure shows the same information as Figure 3.20 but with a different scale. The spectra for (A) is shown at nine different temperatures 5°C (black), 15°C (gray), 25°C (purple), 35°C (blue), 45°C (cyan), 55°C (green), 65°C (orange), 75°C (pink), 85°C (red). Part (B) shows the MRE between the spectra at each temperature relative to the 85°C spectra. In both images, the characteristic temperature-dependence at and around 221 nm is completely gone. The protein appears to be near a random coil at this point.

3.4 Discussion

Phosphorylation is a key step in a number of biological processes from apoptosis[58,59], cell-cycle progression[60,61], cytoskeleton rearrangement[62,63], and cell-to-cell communication[64,65]. Phosphorylation was discovered in 1955 by Eddie Fischer and Ed Krebs[63,66] and by the late 1960s it was widely understood to be a ubiquitous biological event[63,67]. From a structural biology standpoint, phosphorylation is a critical factor determining the size, shape, and functionality of many proteins. The addition of a -2 charged phosphate group can lead to an expanded structure due to an increase in PP_{II}[40] or compaction due to structure stabilization through interactions with basic side chains[63]. These kind of structural changes can activate enzyme activity[68], inhibit enzyme activity[69], or create recognition sites for other proteins[70].

IDPs and disordered protein regions have a high density of phosphorylation sites and are, in fact, where most phosphorylation events happen[71]. Due to the fact that both charge repulsion and intrinsic backbone propensities contribute to the expanded structure of IDPs, and that phosphorylation has a direct impact on net charge, it is important to develop systems to thoroughly understand and characterize IDPs. Our system, based upon R_h as a reporter metric, does just that.

3.4.1 Folded Protein Standards Were Stable at Lower pHs Unlike IDPs

In Section 2.5.5, I showed that chicken egg albumin, *Staphylococcal* nuclease, and bovine carbonic anhydrase are stable at pH > 4.5. This is also evident by looking at the change in K_D values for each protein. While KDs are not repeatable from one prep to the next, they do fall within a certain range. Thus, KDs can be used qualitatively to show a greater variance among IDPs in comparison to the folded protein standards. By looking at the K_D standard deviation for each protein across the pH spectrum, we see much greater fluctuation in the IDPs. A similar trend is seen by looking at the K_D range of each protein across the pH spectrum (Figure 3.22, end of section).

Another way to show the concordance correlation (ρ_c) for the folded protein standards verses that of the IDPs. Lin's concordance correlation was created in 1989 by Lawrence Lin as a way to measure the agreement between two variables[72]. Unlike the Pearson correlation which simply measures a linear relationship, the Lin concordance correlation measures deviation from the identity line[72]. As seen in Figure 3.23 at the end of this section, the folded protein standards have a ρ_c score of 0.9682 whereas the IDPs only have a ρ_c score of 0.2298. This is just another metric to show the folded proteins retain their size throughout the pH range in this study but the IDPs do not.

3.4.2 Net Charge is Modulated by PP_{II}

Net charge is an important factor in determining the hydrodynamic radius of a disordered protein. This is apparent by looking at the SEC results for Pro- which rapidly decreased in size as the pH was lowered. Based on the pK_a values for each titratable residue, virtually every (>99%) Asp, Glu, Lys, and Arg, as well as the C- and N-termini should be charged. This gives the polypeptide a theoretical net charge of -14.99. This information is summarized in Table 3.6 below. N^+ and C^- are the percentage N- and C-termini charged at each pH, respectively. Lys^+, Arg^+, Asp^-, and Glu^- are the number each titratable residue charged at each pH.

Table 3.6. Percentage of charged residues for p53(1-93) at each pH.

pH	N^+	Lys^+	Arg^+	Asp^-	Glu^-	C^-	Charge
7.0	0.997	1.000	1.000	0.999	0.999	1.000	-14.99
6.0	1.000	1.000	1.000	0.995	0.986	1.000	-14.83
5.0	1.000	1.000	1.000	0.951	0.876	0.998	-13.49
4.5	1.000	1.000	1.000	0.860	0.691	0.993	-11.10

As the pH is lowered to 4.5, only 86.0% of Asp residues and 69.1% of Glu residues should be deprotonated giving a theoretical net charge of -11.10. A change of 3.89 spread over only 93 residues should cause significant compaction through the neutralization of

charge groups. Our observations confirmed this hypothesis as we saw a linearly charge-dependent 27% compaction of Pro^-.

If charge repulsion was the driving factor in IDP expansion, then we would expect a similar rate of compaction in Ala^- and WT. However, we saw much smaller compaction in both. The WT saw a 15% reduction in R_h as we moved from pH 7.0 to 4.5, while Ala^- saw a 14% decrease in R_h over the same range. This leads us to conclude that a high concentration of prolines, which cause an increase in rigidity due to the side-chain bonding to the backbone nitrogen, modulates charge repulsion.

3.4.2 Alanine is not as Disorder-Promoting as Expected

Two of the three published PP_{II} propensity scales have alanine as one of the most disorder-promoting residues. The Creamer scale gives alanine a PP_{II} propensity of 0.61, the fourth largest value[39]. The Kallenbach scale ranks alanine with the highest PP_{II} propensity at 0.818[38]. The Hilser scale ranks alanine 8th with a PP_{II} propensity of 0.37.

CD spectra of Ala^-, at all four pHs, more closely resembles that of the WT than Pro^-. As we reduced pH from 7.0 to 4.5, the WT decreased by 15% whereas Ala^- only decreased by 14%. If alanine had a similar PP_{II} propensity to that of proline, we would expect Ala^- to have significantly less spacing between charged groups, leading to a substantially smaller R_h at lower pHs. Our observations did not support this hypothesis,

thus substituting alanines for glycines might have less of an effect on PP_{II} structure than expected.

3.4.2 The Acid-based Unfolding of p53(1-93) is Noncooperative

As mentioned earlier, the size compaction of Pro^- as we moved from pH 7.0 to 4.5 was linear and charge-dependent. Figure 3.24 (end of section) shows this observation holds true for the WT and Ala^- as well. These results are similar to the thermal unfolding of p53(1-93) published by Schaub et al. in which the polypeptide prefers a denatured macrostate at low pH[3].

3.4.3 Pro^- is Close to a Random Coil at pH 4.5

CD spectra was used qualitatively to detect PP_{II} content in each p53(1-93) variant at each pH. However, some trends did emerge. The height of the 221 nm peak was directly correlated with net charge, for each variant, at each pH. While the magnitude of this peak decreased, noncooperative thermal unfolding was observed for 11 of the 12 peptides. The sole exception was Pro^- at pH 4.5 (Figures 3.20 – 3.21). Without prolines and with the overall charge decreased to -11.10, there is no noticeable peak at 221 nm or temperature dependence. The lack of detectable structure via CD spectroscopy indicates that Pro^- is, or is near, a random coil at pH 4.5.

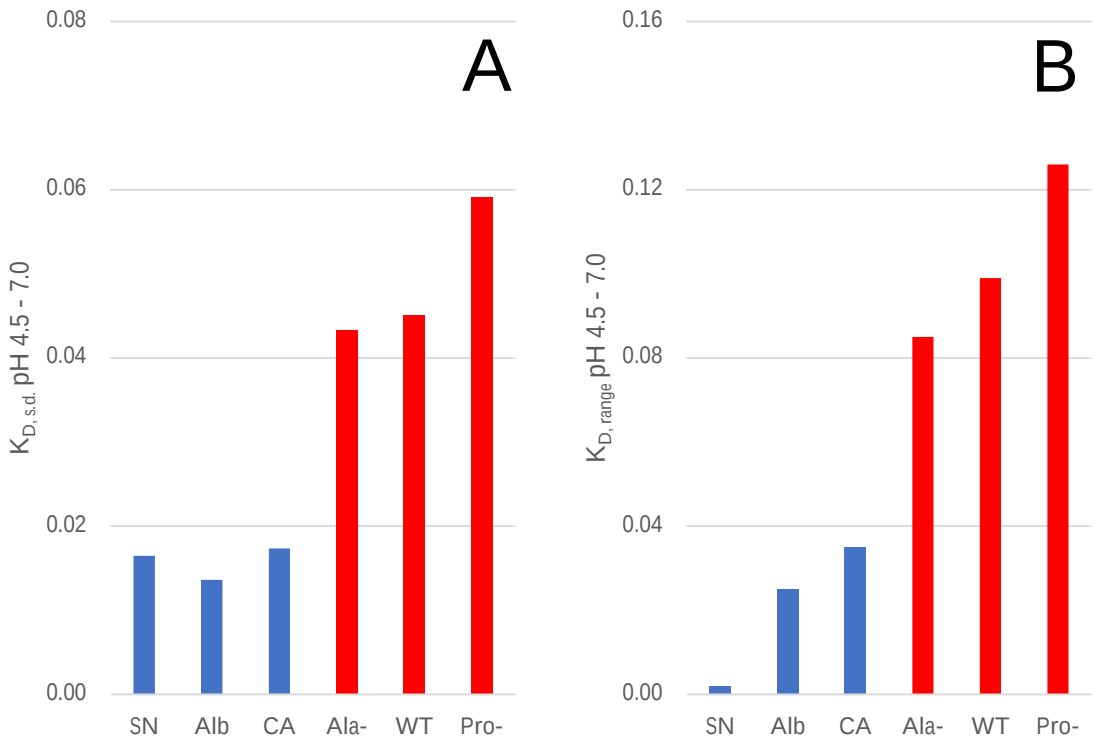

Figure 3.22. K_D range for folded protein standards in comparison to IDPs. Part (A) shows the K_D standard deviation for each folded protein standard is much lower than their IDP counterparts. Part (B) shows the K_D range (K_D 7.0 – K_D 4.5) for each folded protein standard is also much smaller than each IDP.

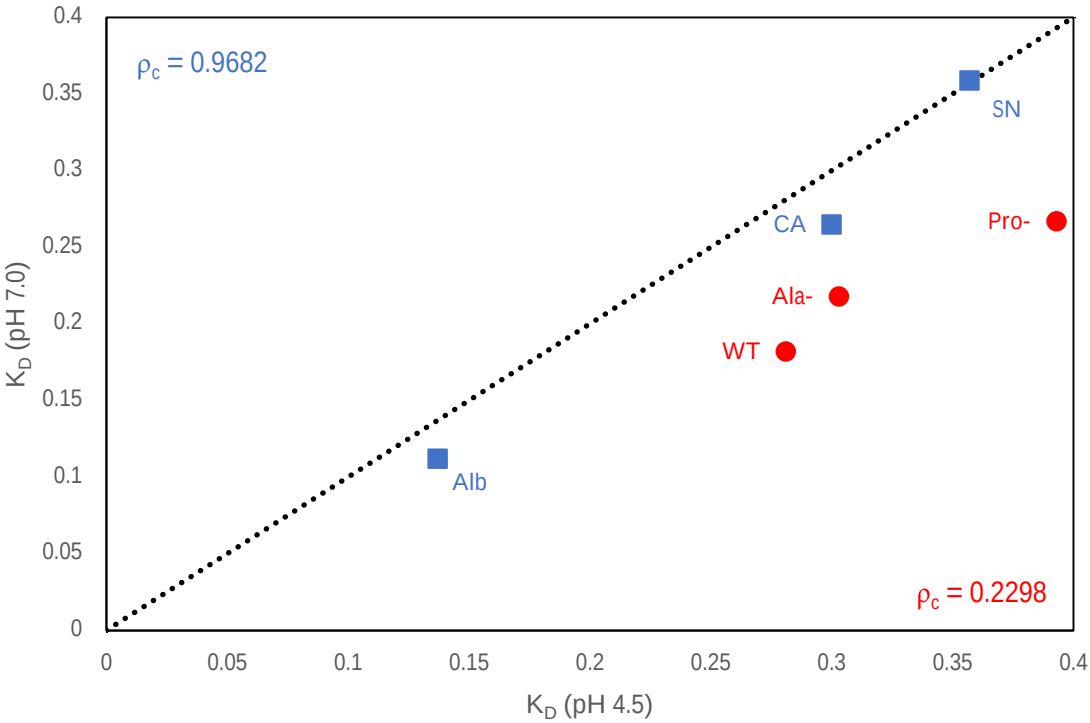

Figure 3.23. Lin's concordance correlation for the set of folded protein standards verses the set of IDPs. The concordance correlation shows a high agreeability (0.9682) for the folded protein standards at pH 7.0 compared to pH 4.5. The same metric shows a much lower agreeability (0.2298) for the set of IDPs at pH 7.0 compared to pH 4.5. These statistics were calculated using the R programming language and the EpiR[50] package.

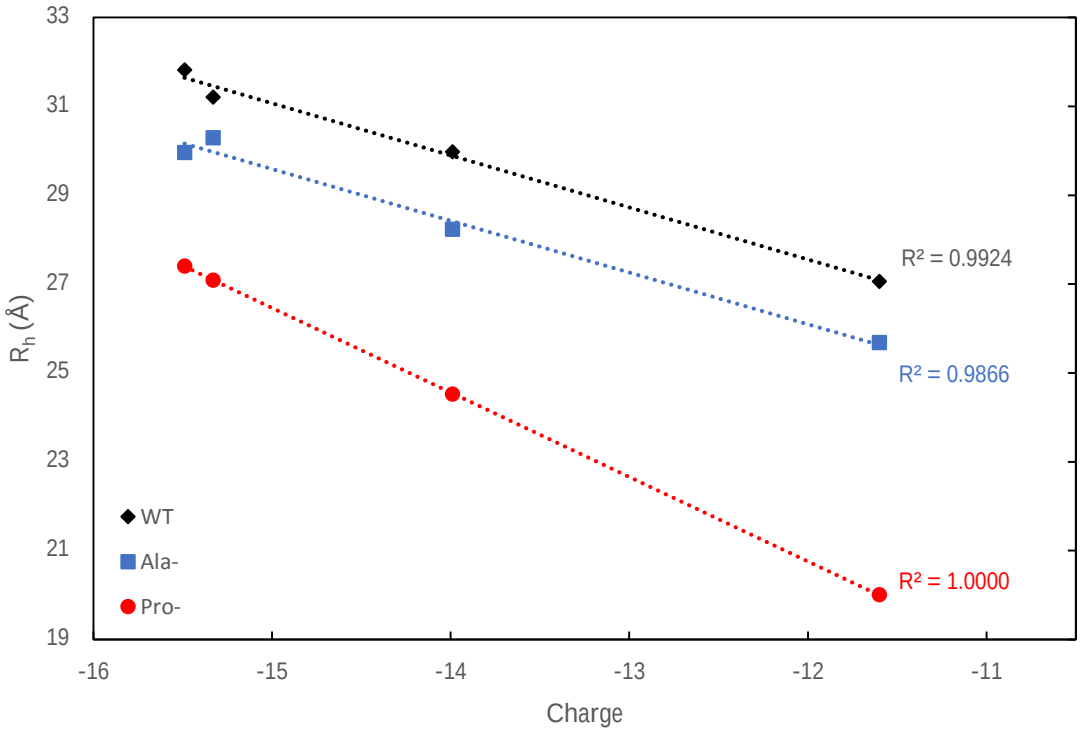

Figure 3.24. Charge-dependent unfolding of p53(1-93). The acid-unfolding of p53(1-93) WT (black diamond), Ala- (blue square), and Pro- (red circle) as pH decreases from 7.0 to 4.5.

3.5 Conclusions

The data in this project are consistent with our hypothesis that PP_{II} content causes an expansion in the R_h of IDPs. CD data suggests that proline has a higher PP_{II} propensity than alanine or glycine, which accounts for the larger temperature dependence in the WT and Ala^- compared to Pro^-. The CD spectra, namely the magnitude of the peak at 221 nm, is also sensitive to net charge. However, lowering the pH had a much greater effect on Pro^- than the WT and Ala^- demonstrating there is more at play here than charge repulsion; PP_{II} content is modulating electrostatic effects.

REFERENCES

(1) Anfinsen, C. B. *Science* 1973, *181* (4096), 223–230.

(2) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. *J. Mol. Graph. Model.* 2001, *19* (1), 26–59.

(3) Schaub, L. J.; Campbell, J. C.; Whitten, S. T. *Protein Sci. Publ. Protein Soc.* 2012, *21* (11), 1682–1688.

(4) Wright, P. E.; Dyson, H. J. *Nat. Rev. Mol. Cell Biol.* 2015, *16* (1), 18–29.

(5) Cuchillo, R.; Michel, J. *Biochem. Soc. Trans.* 2012, *40* (5), 1004–1008.

(6) Berlow, R. B.; Dyson, H. J.; Wright, P. E. *FEBS Lett.* 2015, *589* (19A), 2433–2440.

(7) Sharma, R.; Raduly, Z.; Miskei, M.; Fuxreiter, M. *FEBS Lett.* 2015, *589* (19A), 2533–2542.

(8) Tompa, P. *Trends Biochem. Sci.* 2012, *37* (12), 509–516.

(9) Kornreich, M.; Avinery, R.; Malka-Gibor, E.; Laser-Azogui, A.; Beck, R. *FEBS Lett.* 2015, *589* (19A), 2464–2476.

(10) Xue, B.; Blocquel, D.; Habchi, J.; Uversky, A. V.; Kurgan, L.; Uversky, V. N.; Longhi, S. *Chem. Rev.* 2014, *114* (13), 6880–6911.

(11) Mohan, A.; Sullivan, W. J.; Radivojac, P.; Dunker, A. K.; Uversky, V. N. *Mol. Biosyst.* 2008, *4* (4), 328–340.

(12) Gazi, A. D.; Bastaki, M.; Charova, S. N.; Gkougkoulia, E. A.; Kapellios, E. A.; Panopoulos, N. J.; Kokkinidis, M. *J. Biol. Chem.* 2008, *283* (49), 34062–34068.

(13) Tompa, P. *Trends Biochem. Sci.* 2002, *27* (10), 527–533.

(14) Sahu, D.; Bastidas, M.; Showalter, S. *Anal. Biochem.* 2014, *449*, 17–25.

(15) Langridge, T. D.; Tarver, M. J.; Whitten, S. T. *Proteins* 2014, *82* (4), 668–678.

(16) Knauf, M. J.; Bell, D. P.; Hirtzer, P.; Luo, Z. P.; Young, J. D.; Katre, N. V. *J. Biol. Chem.* 1988, *263* (29), 15064–15070.

(17) Obradovic, Z.; Peng, K.; Vucetic, S.; Radivojac, P.; Brown, C. J.; Dunker, A. K. *Proteins* 2003, *53 Suppl 6*, 566–572.

(18) Greenfield, N. J. *Nat. Protoc.* 2006, *1* (6), 2876–2890.

(19) Dunker, A. K.; Obradovic, Z.; Romero, P.; Garner, E. C.; Brown, C. J. *Genome Inform. Workshop Genome Inform.* 2000, *11*, 161–171.

(20) Fletcher, C. M.; McGuire, A. M.; Gingras, A. C.; Li, H.; Matsuo, H.; Sonenberg, N.; Wagner, G. *Biochemistry (Mosc.)* 1998, *37* (1), 9–15.

(21) Williams, R. M.; Obradovi, Z.; Mathura, V.; Braun, W.; Garner, E. C.; Young, J.; Takayama, S.; Brown, C. J.; Dunker, A. K. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.* 2001, 89–100.

(22) Dunker, A. K.; Brown, C. J.; Obradovic, Z. *Adv. Protein Chem.* 2002, *62*, 25–49.

(23) Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradović, Z. *Biochemistry (Mosc.)* 2002, *41* (21), 6573–6582.
(24) Wang, S.; Trumble, W. R.; Liao, H.; Wesson, C. R.; Dunker, A. K.; Kang, C. H. *Nat. Struct. Biol.* 1998, *5* (6), 476–483.
(25) Machkalyan, G.; Trieu, P.; Pétrin, D.; Hébert, T. E.; Miller, G. J. *Cell. Signal.* 2016, *28* (5), 401–411.
(26) Dyson, H. J.; Wright, P. E. *Curr. Opin. Struct. Biol.* 2002, *12* (1), 54–60.
(27) Holmbeck, S. M.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* 1998, *284* (3), 533–539.
(28) Agalarov, S. C.; Sridhar Prasad, G.; Funke, P. M.; Stout, C. D.; Williamson, J. R. *Science* 2000, *288* (5463), 107–113.
(29) Herrmann, H.; Strelkov, S. V.; Burkhard, P.; Aebi, U. *J. Clin. Invest.* 2009, *119* (7), 1772–1783.
(30) Dunker, A. K.; Ensign, L. D.; Arnold, G. E.; Roberts, L. M. *FEBS Lett.* 1991, *292* (1–2), 275–278.
(31) Ivan, M.; Kondo, K.; Yang, H.; Kim, W.; Valiando, J.; Ohh, M.; Salic, A.; Asara, J. M.; Lane, W. S.; Kaelin, W. G. *Science* 2001, *292* (5516), 464–468.
(32) Joerger, A. C.; Fersht, A. R. *Cold Spring Harb. Perspect. Biol.* 2010, *2* (6), 1–20.
(33) Teufel, D. P.; Bycroft, M.; Fersht, A. R. *Oncogene* 2009, *28* (20), 2112–2118.
(34) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. *Proteins* 2000, *41* (3), 415–427.
(35) Perez, R. B.; Tischer, A.; Auton, M.; Whitten, S. T. *Proteins* 2014.
(36) Tomasso, M. E.; Tarver, M. J.; Devarajan, D.; Whitten, S. T. *PLoS Comput. Biol.* 2016, *12* (1), 1–22.
(37) Adzhubei, A. A.; Sternberg, M. J. E. *J. Mol. Biol.* 1993, *229* (2), 472–493.
(38) Shi, Z.; Chen, K.; Liu, Z.; Ng, A.; Bracken, W. C.; Kallenbach, N. R. *Proc. Natl. Acad. Sci. U. S. A.* 2005, *102* (50), 17964–17968.
(39) Rucker, A. L.; Pager, C. T.; Campbell, M. N.; Qualls, J. E.; Creamer, T. P. *Proteins* 2003, *53* (1), 68–75.
(40) Elam, W. A.; Schrank, T. P.; Campagnolo, A. J.; Hilser, V. J. *Protein Sci. Publ. Protein Soc.* 2013, *22* (4), 405–417.
(41) Woody, R. W. *Adv Biophys Chem* 1992, *2*, 37–79.
(42) Rucker, A. L.; Creamer, T. P. *Protein Sci. Publ. Protein Soc.* 2002, *11* (4), 980–985.
(43) Whittington, S. J.; Chellgren, B. W.; Hermann, V. M.; Creamer, T. P. *Biochemistry (Mosc.)* 2005, *44* (16), 6269–6275.
(44) Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; Grosdidier, A.; Hernandez, C.; Ioannidis, V.; Kuznetsov, D.; Liechti, R.; Moretti, S.; Mostaguir, K.; Redaschi, N.; Rossier, G.; Xenarios, I.; Stockinger, H. *Nucleic Acids Res.* 2012, *40* (Web Server issue), W597-603.

(45) Harlow, E.; Williamson, N. M.; Ralston, R.; Helfman, D. M.; Adams, T. E. *Mol. Cell. Biol.* 1985, *5* (7), 1601–1610.
(46) Whitten, S. T.; García-Moreno E, B. *Biochemistry (Mosc.)* 2000, *39* (46), 14292–14304.
(47) Wong, K. P.; Hamlin, L. M. *Biochemistry (Mosc.)* 1974, *13* (13), 2678–2683.
(48) Ahmad, F.; Salahuddin, A. *Biochim. Biophys. Acta* 1979, *576* (2), 333–338.
(49) Gryczynski, Z.; Lubkowski, J.; Bucci, E. *J. Biol. Chem.* 1995, *270* (33), 19232–19237.
(50) Mark Stevenson. *epiR: Tools for the Analysis of Epidemiological Data*; 2016.
(51) Mao, A. H.; Crick, S. L.; Vitalis, A.; Chicoine, C. L.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* 2010, *107* (18), 8183–8188.
(52) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* 2010, *107* (33), 14609–14614.
(53) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* 2009, *30* (5), 673–699.
(54) Henderson, L. J. *Am. J. Physiol. -- Leg. Content* 1908, *21* (4), 427–448.
(55) Hasselbalch, K. A. *Die Berechnung der Wasserstoffzahl des Blutes aus der freien und gebundenen Kohlensäure desselben, und die Sauerstoffbindung des Blutes als Funktion der Wasserstoffzahl*; Julius Springer, 1916.
(56) *CRC Handbook of Chemistry and Physics, 96th Edition*, 96 edition.; Haynes, W. M., Ed.; CRC Press, 2015.
(57) Gursky, O.; Atkinson, D. *Proc. Natl. Acad. Sci. U. S. A.* 1996, *93* (7), 2991–2995.
(58) Jackson, N. M.; Ceresa, B. P. *Exp. Cell Res.* 2016.
(59) Yadav, N.; Kumar, S.; Marlowe, T.; Chaudhary, A. K.; Kumar, R.; Wang, J.; O'Malley, J.; Boland, P. M.; Jayanthi, S.; Kumar, T. K. S.; Yadava, N.; Chandra, D. *Cell Death Dis.* 2015, *6*, 1–13.
(60) Gheghiani, L.; Gavet, O. *Methods Mol. Biol. Clifton NJ* 2016, *1342*, 157–171.
(61) Truman, A. W.; Kristjansdottir, K.; Wolfgeher, D.; Hasin, N.; Polier, S.; Zhang, H.; Perrett, S.; Prodromou, C.; Jones, G. W.; Kron, S. J. *Cell* 2012, *151* (6), 1308–1318.
(62) Parisiadou, L.; Xie, C.; Cho, H. J.; Lin, X.; Gu, X.-L.; Long, C.-X.; Lobbestael, E.; Baekelandt, V.; Taymans, J.-M.; Sun, L.; Cai, H. *J. Neurosci. Off. J. Soc. Neurosci.* 2009, *29* (44), 13971–13980.
(63) Johnson, L. N. *Biochem. Soc. Trans.* 2009, *37* (4), 627–641.
(64) Sirnes, S.; Kjenseth, A.; Leithe, E.; Rivedal, E. *Biochem. Biophys. Res. Commun.* 2009, *382* (1), 41–45.
(65) Lee, B. K.; Chung, M.-Y.; Lee, K. W. *Mol. Carcinog.* 2015, *54* (5), 351–358.
(66) Fischer, E. H.; Krebs, E. G. *J. Biol. Chem.* 1955, *216* (1), 121–132.
(67) Walsh, D. A.; Perkins, J. P.; Krebs, E. G. *J. Biol. Chem.* 1968, *243* (13), 3763–3765.
(68) Johnson, L. N.; Barford, D. *Annu. Rev. Biophys. Biomol. Struct.* 1993, *22*, 199–232.

(69) Hurley, J. H.; Dean, A. M.; Thorsness, P. E.; Koshland, D. E.; Stroud, R. M. *J. Biol. Chem.* 1990, *265* (7), 3599–3602.
(70) Filippakopoulos, P.; Kofler, M.; Hantschel, O.; Gish, G. D.; Grebien, F.; Salah, E.; Neudecker, P.; Kay, L. E.; Turk, B. E.; Superti-Furga, G.; Pawson, T.; Knapp, S. *Cell* 2008, *134* (5), 793–803.
(71) Xiao, Q.; Miao, B.; Bi, J.; Wang, Z.; Li, Y. *Sci. Rep.* 2016, *6*, 28413.
(72) Lin, L. I. *Biometrics* 1989, *45* (1), 255–268.