INTRINSIC DISORDER IN PROTEINS: FOLDING OF THE YEAST PROTEINASE INHIBITOR IA_3

By

OMJOY KUMAR GANESH

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

To God through whom all things are possible,
my wife, father and mother, who constantly provided care and patience along my journey,
and to the Edison, Hagen and Long labs who have provided a scientific community
in which to grow

ACKNOWLEDGMENTS

My journey through graduate school has not been an easy one. Through the years, several people have not only watched or commented on the struggle, but helped me in various ways. I genuinely feel that I would not have made it through this ordeal without this assistance.

My first acknowledgement (which is more than simple thanks) is to God. Without his guidance, my life would not be what it is today. Any achievements I have, any awards or acknowledgements that I have gained, are all due to his hand in my life.

During my time in graduate school I gained a companion in life through marriage to Katia T.C. Ganesh. Her constant love and support have kept me afloat even through extremely trying times. Through marriage, I also became the proud owner of two peculiar cats who unknowingly have aided in keeping me sane. My parents, Dr. T.S. Ganesh, and Dr. Lalitha Ganesh provided support and bits of wisdom along with the perpetual, "When will you graduate?" and "Why are they keeping you so long?" They instilled me with the idea that I could accomplish great things. Additionally, I wish to thank my sisters, Shanti and Geeta Ganesh, who have gone on to become doctors of a different sort.

I am exceedingly grateful to Dr. Arthur Edison and Dr. Stephen Hagen for guiding me through my graduate career. Both of them opened their wealth of knowledge to me, providing me with space to grow and learn scientifically. Their perpetual pushing has made me into a better scientist, sharpening my abilities and firming the foundation of my scientific career. I feel that I have been under the tutelage of two top-notch, talented scientists who have given me priceless instruction.

I thank the members of my committee, Drs. Ben Dunn, Joanna Long, Linda Bloom, and Adrian Roitberg. Their suggestions and guidance and encouragement have greatly aided my graduate studies. I have been very fortunate to have a talented committee to guide me.

Additionally, I thank the entire Biochemistry and Molecular Biology department and the Interdisciplinary Program in Biomedical Sciences. I have had the pleasure of interacting with various members of the faculty and staff throughout the years. I wish to especially note Pat Jones, Regina Corns, Terry Rickey, Denise Mesa, Brad Moore, Valerie Cloud-Driver, and Susan Gardner who have provided me with a great deal of assistance.

The friendship of the many people who have come through the Edison, Hagen and Long laboratories made my time at UF not only enjoyable, but fun and exciting. Camaraderie between everyone really does help make tough days and long hours more worthwhile. There are almost too many people to list here, but some of the most important are: Cherian Zachariah, Ramazan Ajredini, Iman Al-Naggar, Aaron Dossey, Heather Cornnell, Seth McNeill, Vijay Antharam, Ranjani Narayanan, and Pablo Perez. Previous members of all the labs, such as Suzette Pabit, LinLin Qiu, Terry Green, Jim Smith, Julie Vanni, Jonathan Lane, and Mini Samuel-Lantiser also made my time here much more enjoyable.

There are other people who also deserve acknowledgement such Alfred Chung and the staff at the Protein Core, Jim Rocca, Al Guarino, many undergrads and high school students, and many others. Frankly, it is impossible for me to list all the people who have helped me, or all the ways in which people have assisted me over the years.

I also need to thank my church family. Through the years I have become synonymous with the perpetual student: always here and always working on a mysterious project that involves proteins. Many people ask me what I study, but my church family showed real interest and concern. To list everyone in my church family who has helped me would require another chapter in this dissertation; instead, I want to thank them all, and let each of them know that they have made a difference in my life.

Finally I need to point out that graduate school is not as bad as it seems. While my trials may have been exceedingly tough, it is worthwhile to recognize not all things put into fires turn to ash. Fires are also used to strengthen and purify, thereby raising the value and rarity of an object. So, I must thank the University of Florida for providing a place where I could be tossed into a fire, yet come out stronger and better.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1,5-IAEDANS	5((((2iodoacetyl)amino)ethyl)amino)naphthalene1sulfonic acid
CD	Circular dichroism
DSC	Differential scanning calorimetry
EPR	Electron paramagnetic resonance
ER	Endoplasmic reticulum
his_6 tag	C-terminal addition to IA_3 with the amino acid sequence LEHHHHHH
HSQC	Heteronuclear single quantum coherence
IA_3	Inhibitor of YPrA
IAEDANS	1,5-IAEDANS
IDP	Intrinsically disordered protein
ITC	Isothermal titration calorimetry
NMR	Nuclear magnetic resonance
PMSF	phenylmethanesulphonylfluoride or phenylmethylsulphonyl fluoride (a serine protease inhibitor)
RF	Radiofrequency
SVD	Singular value decomposition
TFE	2,2,2-trifluoroethanol
UF	University of Florida
YPrA	Yeast proteinase A, Yeast protease A

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

INTRINSIC DISORDER IN PROTEINS: FOLDING OF THE YEAST PROTEINASE INHIBITOR IA_3

By

Omjoy Kumar Ganesh

May 2008

Chair: Arthur S. Edison
Cochair: Stephen J. Hagen
Major: Medical Sciences -- Biochemisty and Molecular Biology

Intrinsic disorder is a common feature in proteins yet is not completely understood. Analysis of the folding of simple proteins may hold the key to understanding the benefits of intrinsic disorder. IA_3 is a 68 amino acid endogenous inhibitor of yeast aspartic proteinase A (YPrA). In solution IA_3 can be found as an unstructured protein; however in the presence of YPrA, the N-terminus of IA_3 can be found as a helical structure bound to YPrA. This transition can be induced by the addition of the alcohol cosolvent 2,2,2-trifluoroethanol (TFE).

Circular dichroism studies of IA_3 in the presence of TFE indicated that the refolding midpoint between the coil and helix state is 16.3% TFE. However, these studies cannot conclusively state which residues are involved in the folding. Comparison of the 2D ^{15}N-HSQC spectra of IA_3 in water and in 23% TFE indicated that the protein undergoes an unstructured to structured transition in the presence of TFE. The refolding midpoint found by NMR is 18.3%, in agreement with the CD studies. The NMR studies allowed a residue level look how individual amino acids move toward a helical structure, with a greater helical tendency at the N-terminus.

Inhibition assays of YPrA with several IA_3 constructs show that only the N-terminal half of the protein is necessary to inhibit YPrA with the same specificity as the wild type protein. Isothermal titration calorimetry was employed to understand the function of the C-terminal portion of the molecule within its biological context. These studies demonstrated a marked difference in interactions between the wild type protein and the N-terminal half of IA_3 with YPrA, suggesting that the C-terminus plays a role in the interaction between the two proteins apart from inhibition.

Assessment and comparison of the folding rates within a biological context and in TFE may shed light on the validity of the models used to understand such a simple system. Laser-induced temperature jump experiments were used to determine if the folding rate of IA_3 changes in the presence of YPrA.

This research demonstrates that the intrinsically disordered IA_3 displays a rich set of interactions with YPrA.

CHAPTER 1
INTRODUCTION

Introduction

A modern understanding of the protein sequence-structure relationship was first formally described by Christian Anfinsen in the 1960s (*1*). In his experiments, he noted that an unfolded protein chain can refold to its unique 3D structure spontaneously and reproducibly (*2, 3*). This has led to a dogma in molecular biology: sequence begets structure and structure begets function (*4-6*). Since then, vast strides have been made in our understanding of protein function and structure. However, this viewpoint has led to a greater focus on globular or well-structured proteins, progress in the study of unstructured or intrinsically disordered proteins (IDPs) has been much more gradual (*7*).

In this dissertation, I will focus on the intrinsically disordered protein IA_3. This protein is an endogenous inhibitor of the aspartic proteinase, YPrA, and binds to YPrA with nanomolar potency, with the N-terminus of IA_3 transitioning from disordered to a helical structure on binding (*8-11*). Before continuing with the IA_3 story, it is important to define what intrinsic disorder is. Additionally, I will describe two examples of intrinsic disorder, illustrating the importance of disorder in biological systems, and will briefly describe a few tools that were used to study IA_3. Finally I will present some background on the aspartic proteinases, YPrA, and IA_3.

Defining Intrinsic Disorder

In their landmark 1963 paper, Gopalasamudram Narayana Iyer Ramachandran, Chandrasekaran Ramakrishnan and Viswanathan Sasisekharan proposed the plotting of the backbone dihedral angles φ versus ψ for each residue in a protein as a way to assess protein structure (now commonly known as a Ramachandran plot, Ramachandran diagram or Ramachandran map) (*12*). In well-ordered proteins, φ and ψ angles (and backbone atoms) are

distinctly affected by local packing environment and consequently exhibit small amplitude motions relative to neighboring atoms/angles (*4, 12-14*). Secondary structures formed in well-ordered proteins tend to fall in several specific regions of the Ramachandran plot. Also, the ensemble of structures measured for a well-structured protein tends to cluster in certain regions of the Ramachandran plot (Figure 1-1). Intrinsically disordered proteins, however, are characterized by a distribution of dihedral angles that are unhindered by packing interactions and can vary significantly over time for each amino acid residue, giving rise to a *dynamic ensemble* of interconverting structures (*15-17*). A Ramachandran plot of a disordered protein would show a diverse distribution of φ and ψ angle pairs which would change over time (*18, 19*).

Intrinsic disorder has been referred to by several different terms in literature: "natively/intrinsically disordered" (*20, 21*), "natively/intrinsically unfolded" (*22*), "natively/intrinsically unstructured" (*5, 23*), "floppy structure" (*24*), "loopy structure" (*25*), and "random coil", "coil" or "coil-like" (*24*). All of these terms are in contrast to a stable three-dimensionally well-ordered protein, which Linding *et al.* defines as "typically contain[ing] regular secondary structures" (*26*). Comparison to well-structured proteins can lead to a simple definition of intrinsic disorder: no regular secondary structure and potentially a large amount of flexibility in the backbone φ and ψ angles (*7, 22, 26*).

Biological Importance of Disorder

Despite a lack of three-dimensional structure, the importance of disordered protein states has been gaining recognition in biochemistry. Several authors have noted that unstructured proteins play a large role in normal and abnormal biological functions, such as protein transcription (*27, 28*), translation (*29, 30*), translocation (*31*), degradation (*32*), assembly (*33, 34*) and aggregation (*32, 35*). Additionally, many authors have noted that disordered proteins can commonly be found in regulation, signaling and control pathways (*36-40*). To illustrate the

commonness of disorder normally and in disease, I will focus on two examples, High Mobility Group (HMG) proteins and cardiovascular disease (CVD).

Disorder Is a Common Feature in Biological Processes: HMGA Proteins

The High Mobility Group proteins (HMG proteins) are a group of multifunctional chromosomal proteins. These proteins have been known to aid in transcription, in addition to replication, recombination, and DNA repair (*41*). The high degree of flexibility implied in their name is due to the distinguishing structural characteristic that members of this group have very little detectable secondary structure (*42*). Binding to a substrate, however, causes HMG proteins to undergo an unstructured to structured transition, which is similar to the transition necessary for inhibitory action of IA_3. HMG proteins have been shown to interact with DNA and a variety of transcription factors (*43*). An important example member of this group of proteins is HMGA, with the C-terminal tail of HMGA presenting an interesting story. The C-terminus of HMGA is not necessary for binding to transcription factors and other proteins, however it has been found to enhance transcription factor activity (*44, 45*). This effect may be similar to the C-terminus of IA_3, which is not needed to inhibit YPrA, but may enhance the interactions between YPrA and the inhibitory N-terminus of IA3. This will be further touched upon in Chapter 4.

Disorder Found in Disease: CVD

Diseases of the heart and arteries are generally grouped under the term Cardiovascular Disease (CVD). Among the diseases known as CVD are heart attack, stroke, heart disease, high blood pressure, heart failure, and atherosclerosis (*46*). CVD is currently the leading cause of death for individuals in the United States, affecting more that 910,000 people each year (*46, 47*). It was estimated that in 2005 around $450 billion was spent on CVD treatment and therapy (*46*). Currently there are around 400 prescription drugs for treatment of CVD targeting various risk factors associated with the disease (*48*).

Several authors have shown that amino acid sequences for CVD related proteins that encode for disordered regions in a protein are considerably different from those that encode for ordered regions (*6, 20, 22, 49*). The differences lie in several factors, including local amino acid composition, flexibility, hydropathy, and charge (*6, 20, 22, 49-51*). Disordered regions in proteins are characterized by low complexity primary sequences which are biased towards disorder-promoting residues, and depleted in order-promoting residues. Cheng *et al.* analyzed 487 CVD-related proteins and was able to show that there exists a high degree of intrinsic disorder among CVD proteins (*50, 52*). Their analysis suggests that this set of proteins was enriched in disorder-promoting residues, and depleted in order-producing residues, suggesting that CVD proteins have an unstructured nature which is similar to IA_3. It has been postulated that this intrinsic disorder could be important for control and regulation of processes associated with CVD via molecular recognition (*53*).

Tools of the Trade: Investigating Disorder and Disorder to Ordered Transitions in Proteins

The intrinsic lack of fixed 3-D structure makes studying IDPs challenging. Many different experimental methods have been used to detect, investigate and characterize the disorder in IDPs. Several of these methods are discussed in this section.

Bulk Properties

Susceptibility to proteases

It has long been understood that disordered regions of proteins become targets for proteolysis more readily than structured regions (*54-58*). Studies of proteolytic activity have established that proteolytic sites are usually located within flexible loop or disordered regions, are typically solvent exposed, and are generally devoid of secondary structure (*54-62*). Disordered regions in proteins tend to jut outward from the protein surface into the solvent environment making them more open and accessible targets for proteases as compared to the

compactness of well-folded regions (*56*). Consensus in the literature indicates that a protein that is folded into a complex structure tends to be resistant to proteolysis (see for example (*54, 55, 58, 63*)). In this manner, susceptibility to proteolysis can crudely distinguish between ordered and disordered proteins (*54, 57, 60, 64, 65*).

Gel filtration

Size exclusion chromatography (also called gel filtration chromatography) is often said to separate proteins based on 'size'. For a globular protein, an increased 'size' can roughly correspond to an increased molecular mass (*66*). However, this correlation does not hold with intrinsically disordered proteins. In actuality, size exclusion chromatography separates proteins based on hydrodynamic volume (*67, 68*). Several authors have noted that disordered proteins possess hydrodynamic properties which are significantly different than globular proteins. Due to their extended conformations, the hydrodynamic radius of disordered proteins is generally larger than the hydrodynamic radius of globular proteins. This extended hydrodynamic volume leads to abnormal movement through a gel filtration column, and results in a disordered protein eluting at an apparent molecular mass which is larger than what would be expected. This is examined in Chapter 5.

Circular Dichroism spectroscopy

Circular dichroism (CD) spectroscopy can be used to monitor secondary structure through the differential absorption of right and left handed circularly polarized light. Chiral compounds will interact with circularly polarized light (*69*). In a protein, there are two optically active centers: side groups of aromatic amino residues, and peptide bonds. Far-UV CD (wavelengths of 180 nm to 250 nm) highlights the symmetry across the peptide bond and is indicative of secondary structure within the protein (*70*). Distinct CD spectra can be obtained for the alpha-helix, beta-sheet, and random coil secondary structures (*71*). Disordered proteins generally

exhibit a large negative ellipticity below 200 nm, however the spectra can also display features characteristic of other secondary structural elements (*70*).

CD spectroscopy of IA_3 was used extensively throughout the many experimental studies detailed in this dissertation. CD has been used as an indicator of the unstructured nature of IA_3 exhibiting a negative ellipticity below 200 nm, as in Chapter 4. The unstructured to structured transition of IA_3 by NMR is compared to CD results and discussed in Chapter 2. Finally, these structural changes were investigated using CD and correlated to temporal events as discussed in Chapter 3.

Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) measures the heat change upon binding of two (or more) reactants (*72*). While ITC cannot directly measure disorder within a protein, it can be used to report on the interactions between two proteins. The release or absorption of heat in a chemical reaction can be an indication that the reactants are interacting (*72*). Nonspecific interactions between disordered proteins or disordered regions of proteins and other proteins would result in a heat change in the system which can be monitored by ITC (*72, 73*). The N-terminus of the intrinsically disordered IA_3 does bind to YPrA, while the C-terminus appears unstructured and does not seem necessary for inhibition. The energetics of this interaction are discussed in Chapter 4. Additionally, a more detailed description of ITC studies of IA_3 can be found in Chapter 5.

Monitoring Changes Through Fluorescence

Intrinsic fluorescence

Fluorescence spectroscopy techniques have been widely used to monitor structural transitions within proteins. Aromatic amino acids are the main intrinsic fluorescent chromophores within a protein (*74*). Tryptophan is more often used as a fluorescent

chromophore (rather than other aromatic amino acids), since the quantum yield of phenylalanine is low in comparison and tyrosine fluorescence can be quenched due to many factors, including its ability to ionize or its proximity to carboxyl groups (*75*).

The emission characteristics of tryptophan are highly dependent on the polarity of the residue's local environment. In a nonpolar environment, tryptophan has an emission maximum around 320 nm; in a polar environment the emission maximum is close to 350 nm (*75-77*). While this wavelength change does not demonstrate that a protein is disordered, it can yield basic information about when the residue is in different local environment conditions (*75-79*).

IA_3 does gain structure upon the addition of the alcohol cosolvent trifluoroethanol (TFE), which can be monitored by intrinsic fluorescence. In Chapter 3, the intrinsic fluorescence of tryptophan as well as the added dansyl fluorophore are used to monitor these equilibrium changes.

Fluorescence resonance energy transfer

Fluorescence resonance energy transfer (FRET) has been a very useful tool in the study of protein behavior (*80*). FRET describes energy transfer between two chromophores through resonance (*79-81*). The basic idea is the interaction is similar to two oscillators that are separated by a small distance (*79*). One oscillator is in an excited state, which results in an electromagnetic field. This field can induce oscillation with the same frequency in the non-excited oscillator. This idea forms the basis of energy transfer between the donor and acceptor chromophores, however in FRET the chromophores are dipoles with matching resonance frequencies. The transfer of energy between the donor and acceptor chromophore in a dipole-dipole interaction is a resonance phenomenon.

FRET has been called a molecular ruler, since it is sensitive to distance as efficiency of energy transfer depends on the distance between the donor and acceptor chromophore (*79*). The

efficiency of energy transfer is proportional to the inverse sixth power of the distance between the donor and acceptor chromophore. FRET is routinely used to measure distances on the order of 10 to 100 Å.

FRET can be used to monitor the distance between residues in a disordered protein under different solvent environments, or through a temperature change. These changes can be followed as a function of time yielding rate information about local structure formation. This is useful in the comparison of the folding of IA_3 in TFE or in the presence of YPrA as discussed in Chapter 3.

Structure at the Molecular Level

X-Ray crystallography

X-ray crystallography can be used on protein crystals to determine the arrangement of atoms within the protein. The technique requires that a protein sample can be crystallized. However, disordered regions in proteins represent imperfections to the regular repeating structural units within a protein crystal. Processing crystallographic data and assigning atoms in space requires similarity/regularity in the crystallographic units within the crystal (*82*). Regions with high similarity between crystallographic units reveal an average high electron density aiding in molecular assignments. It is believed that disordered regions within a protein can adopt conformations which are different between crystallographic units, and leads to a diffuse electron density (*83*).

Diffraction methods to obtain high resolution structural information on disordered or partially disordered proteins often utilize antibodies or binding partners (*84*). Structures derived from these techniques are still suspect, however, since molecular interactions between the antibody or binding partners and the protein of interest are necessary to 'freeze out' a particular member of the conformational ensemble (*85*). The crystal structures of IA_3 are all in the

presence of YPrA. In these structures only the N-terminus of the protein is determined indicating that the C-terminus may still be unstructured.

Nuclear magnetic resonance

The nuclei of atoms with spin angular momentum can be affected by magnetic fields; if they are placed in a magnetic field, they will align to that field (*86*). In an NMR experiment, pulses of radio frequency (RF) energy are used to 'knock' the nuclei from the aligned equilibrium positions in the field (*86*). The nuclei respond by precessing around the axis of the magnetic field, and relaxing back toward that equilibrium (*87*). The frequency of this precession is characteristic of the physical and chemical environment for each particular atom. Using different NMR experiments, with different RF pulses, parameters such as distances between atoms, the local chemical environment of an atom and the number of bonds between two atoms can be assessed (*86, 87*). This information can be used to determine the placement of atoms within a protein structure or how a protein structure is changing with changes to the solution environment.

In their 2005 paper on the Database of Protein Disorder, Vucetic *et al.* stated: "NMR spectroscopy provides the most direct estimation of the internal motions that characterize disordered ensembles and can be used to identify each residue lacking fixed structure if peak assignments are made."(*88, 89*) Characterization of residue movements can still be made even without clear assignments of all residues. Both global analysis and individual residue characterization of IA_3 in the presence of TFE was carried out and is discussed in Chapter 2.

The advantage to using NMR over other techniques such as CD is that NMR can provide residue level information that other techniques cannot (*88-90*), and can also be used to measure dynamic rates. CD and related techniques rely on datasets that represent an average of the whole protein structure, and may miss small variations in structure, as they may represent only a small

fraction of the total dataset (*88*). In comparison to X-ray crystallography, NMR is more suitable for investigating the relationship between structure and dynamics within a protein.

Other Tools

There are many other techniques and tools which can be used to investigate disorder such as radius of gyration, optical rotary dispersion (ORD), Fourier transform infrared spectroscopy (FTIR), Raman spectroscopy, and others. Each of these techniques has certain positives and negatives, exploiting different properties to probe disordered proteins.

Understanding Disorder

In the 1970s, Peter Chou and Gerald Fasman examined the occurrence of particular amino acids in secondary structures solved by x-ray crystallography (*91-93*). This led to the empirical observation that certain amino acids have a propensity for certain secondary structures. This analysis can now be used to aid in secondary structure identification, but does little to identify disordered regions in proteins (*94, 95*). Several authors had noted that disordered regions in proteins tend to be enriched or depleted in certain amino acids as well (*96-98*). Disordered regions are known to be enriched in disorder-promoting amino acids (alanine, arginine, aspartic acid, glutamic acid, glutamine, glycine, lysine, phenylalanine, and serine) while they are depleted in order-promoting amino acids (asparagine, cysteine, isoleucine, leucine, phenylalanine, tryptophan, tyrosine, valine) (*96, 99, 100*).

Other properties have been used to identify disordered regions in proteins also (*101*). Uversky *et al.* have noted that a characteristic of disordered proteins is a low overall hydrophobicity and large net charge (*22, 102*). Additionally, it has been noted that disordered regions have a high degree of flexibility and are found only within looping or coil regions of a protein (however these regions do not necessarily have to be disordered) (*103*). Prediction of these regions can aid in identifying disorder (although it is not an absolute requirement).

Complicating a physical understanding of disordered proteins is the definition of disorder itself. One would expect that a disordered protein could explore all degrees of freedom, essentially spanning all conformations available. However, this is not the case as studies under very strong denaturing conditions (such as 8 M urea) have failed to observe this (*104-106*). Instead, the disordered ensemble for a particular IDP may contain a certain amount of residual structure, even under strongly denaturing conditions (*104, 106*).

Dyson *et al.* have postulated that due to its disorder, an IDP may have an advantage in making nonspecific, promiscuous contacts with other proteins which may aid in further interactions (*107*). Consider an unstructured protein as a fishing line that can cast out and explore its immediate environment (*108*). A structured protein would not be able to explore its environment as readily simply because it must diffuse through this space more slowly (*108, 109*). The center of mass of an IDP may diffuse slower due to a larger hydrodynamic radius, however this can be an advantage since an IDP can explore its environment over a shorter time than a well structured protein (*108*). This ability to explore its environment in combination with its ability to make nonspecific contacts with other proteins is known as the fly-casting model (*108-111*). This model, postulated by Shoemaker *et al.*, holds that an unstructured protein can act much like fishing line in that the unstructured nature enhances its ability to move through its environment to find a target protein (*108*). Initial nonspecific contacts can also be an enhancement over a rigid globular protein making specific binding contacts (*108*).

The IA_3·YPrA System

The focus of this dissertation is on the intrinsically unstructured protein IA_3. IA_3 displays an amazing nanomolar potency towards its cognate protease, in the process undergoing an unstructured to structured transition. Here I wish to touch on the general aspects of aspartic proteinases, and then focus on specific properties of YPrA, and finally the inhibitor IA_3.

General Overview of Aspartic Proteinases

Aspartic proteinases are important enzymes, characterized by an aspartic dyad that is necessary for proteolytic action (*112*). Structural studies of these enzymes show a bilobed molecule with an opening or cleft located at the interface of the two lobes (*112, 113*). Within this cleft, each lobe contains an invariant Asp-Thr/Ser-Gly sequence. The Asp residues of both lobes are positioned close to each other, and both are necessary for proteolytic cleavage of a substrate. The cleft is thought to stabilize and position a substrate within a particular distance of the aspartic dyad (*114*).

The steps involved in proteolytic cleavage by aspartic proteinases are known as the "push-pull proton transfer mechanism" (*113, 115-117*) (Figure 1-2). At an acidic pH, both aspartate residues form an extensive hydrogen bonding network with the solution and to each other, leaving one aspartate charged, and the other available to be deprotonated. In a single step, a proton from water can transfer to the carboxyl end of an aspartate residue causing a proton from the carboxyl group of the neutral aspartic acid to transfer to a carbonyl group in the backbone of the substrate. This protonation breaks the double bond of the carbonyl allowing a hydroxyl molecule to bond to carbon which previously had a double bond to oxygen, and now forming a tetrahedral intermediate (*118*). In a reverse step, the proton gained by the carboxyl group of the aspartate residue in the protease is transferred to the backbone amide group causing a partial charge imbalance. This forces a proton transfer from the hydroxyl of the carbon back to the aspartic acid of the protease, allowing a carbonyl to reform and breaking the scissile bond of the substrate.

Yeast Proteinase A

YPrA is synthesized from the ribosome as a 405 amino acid protein (*119, 120*). In this initial state it contains a propeptide segment that stabilizes the protease and inactivates the active

site, leaving the protease inactive; this inactive form is known as a zymogen (*121*). In addition to the propeptide, there is a signal sequence that targets the protein to the endoplasmic reticulum (ER) for extensive glycosylation (*119*). This full sequence is referred to as PreProYPrA, since it contains the signal sequence (the Pre portion), and the prosegment.

Within the ER, the 22 amino acid signal sequence is removed, and oligosaccharides are added to the protein through several post-translational modifications (*122*). These oligosaccharides are processed as the protein moves through the ER into the golgi complex and is shuttled into an acidic vacuole (*119, 122*). Two processes could occur within this acidic environment to activate the protease. Interaction with YPrB could cleave the 54 amino acid prosegment, or autoactivation of YPrA could result in a pseudo-YPrA which is able to cleave the prosegment from other YPrA molecules (*123, 124*) (Figure 1-3). Both processes will result in the same outcome, a mature and active 42 kDa protein (of 329 amino acids) which is localized to an acidic vesicle (*125*).

In the vacuole, YPrA could serve several different roles. There does exist some interplay between several yeast proteinases including YPrA, YPrB and YPrC (Figure 1-4). These proteinases can interact to activate one another by cleaving prosegments from inactive forms of the 'partner' protease. It has also been shown that YPrA can activate phosphofructokinase and aminopeptidase 1, as well as interact with other hydrolases (*126-129*). The activity of YPrA with these hydrolases does not stop at activation, however, since prolonged incubation of these proteins with YPrA will result in their degradation. Density centrifugation studies indicate that YPrA can be found bound to many different cellular proteins (*126*). Although the general action of YPrA does seem to be proteolytic degradation of protein substrates, it is conceivable that it

could interact with, activate, or inactivate other proteins through a scheme similar to its interaction with YPrB and YPrC.

Yeast Proteinase A Inhibitor: IA_3

IA_3 is a 7.7 kDa protein of *S. cerevisiae* (*10, 130*). The wild type protein is 68 amino acids, consisting of 47 polar residues (~70%), and six aromatic residues (~9%). The polar character of the molecule is due to thirteen lysine, seven aspartate, six glutamate and six serine residues. Within the sequence there are fifteen basic residues and thirteen acidic residues.

While YPrA is found within the acidic vacuole, IA_3 is found in the cytosol (*131*). This sequestration of the inhibitor from the protease suggests that molecules of IA_3 will not encounter YPrA unless the vacuole is ruptured. If the vacuole is disrupted then formation of an IA_3·YprA complex could prevent cell degradation by inhibiting the proteolytic action of YPrA.

Purified IA_3 is unstructured in solution (*10, 11*). Circular dichroism experiments indicate that the protein has a bulk unstructured nature (*9*). NMR experiments confirm that the general nature of the protein is unstructured with little cooperativity between residues to form secondary structure; however individual residues may have a tendency towards certain structure (*8*).

The IA_3 protein has the distinction of being the first specific inhibitor of an aspartic proteinase to be discovered (*130, 132*). IA_3 displays a remarkable specificity for YPrA, although YPrA is similiar in structure and function to many aspartic proteinases (*11, 133, 134*). While the inhibitor acts as a substrate for other similar aspartic proteases, it displays a sub-nanomolar potency towards YPrA (*11, 133, 134*). Both the wild type IA_3 and the truncated 2-32 N-terminus of IA_3 display nanomolar potency toward YPrA, strongly suggesting that only the N-terminal residues are needed for inhibition of the protease (*11*).

The crystal structure of the IA_3·YprA complex shows residues 2-32 of IA_3 bound as a helix to YPrA (*10*) (Figure 1-5). The IA_3 helix binds to the flap region of YPrA, with several specific

contacts forming between the protease and the inhibitor (*10*). The helix also displays an amphipathic character with hydrophobic residues facing into the active site cleft (*10*). By binding to the flap region of YPrA, IA_3 effectively blockades entry of substrates into the active site cleft, thereby inhibiting the protease (*10, 11*). Additionally, it has been suggested that binding to the flap region keeps IA_3 itself far enough away from the catalytic dyad of YPrA to keep the inhibitor from being a substrate (*10, 11*). It has also been noted that IA_3 could change the hydrogen bonding network around the catalytic dyad (*132*). Finally, while there is strong evidence for the role of the N-terminus in the binding of the inhibitor to the protease, there is possibly also evidence that the C-terminus of IA_3 could play a role in the binding affinity of IA_3 to YPrA (*135*).

Primary Goals of This Dissertation

The unstructured nature of IA_3, and the unstructured to structured transition that is necessary for its mode of inhibition provides a useful system to investigate intrinsically disordered proteins and their interactions with partner proteins. Through the use of several different biophysical techniques I will follow the structural and functional transition in the disordered IA_3 protein and its binding to YPrA. Chapter 2 of this dissertation will discuss the individual residue movements and the transition involved in the folding of IA_3 in the presence of the renaturant TFE using CD and NMR techniques. The kinetics and time resolution of the folding events of IA_3 in the presence of TFE or YPrA will be examined in Chapter 3 using CD, temperature jump, and fluorescence techniques. Finally, the energetics of the interaction of IA_3 with the protease will be explored in Chapter 4 using ITC.

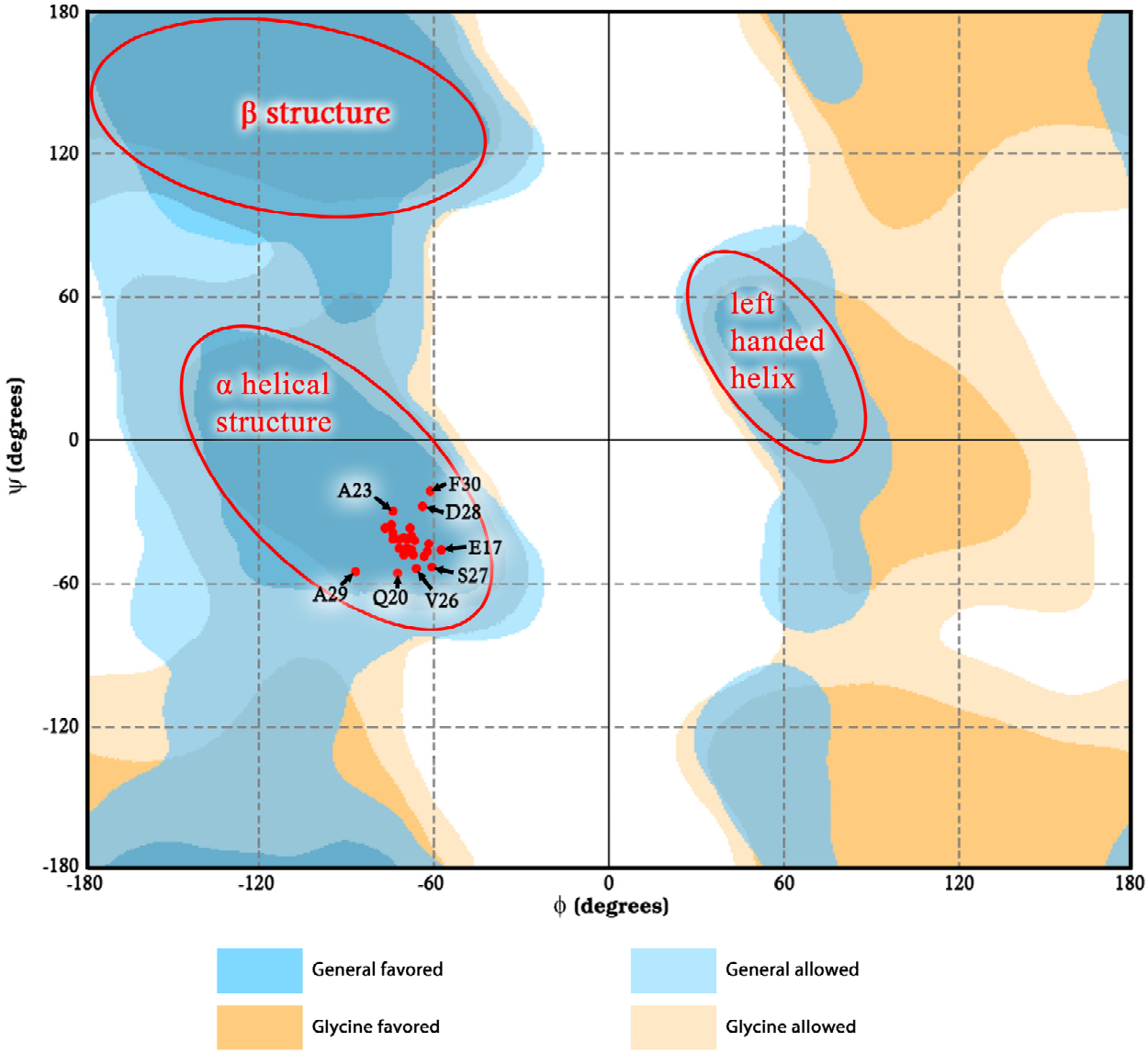


Figure 1-1: Ramachandran diagram of the N-terminal residues of IA_3 (residues 3-31) showing various secondary structures. The gradations on the plot were generated from data from Richardson *et al.* and represent the 98% (favored) and 99.98% (allowed) countours for all amino acids, and for glycine separately (*136*). The red circles indicate clustered regions of certain secondary structures. Red dots indicate the residues of IA_3 which are found in the 1DPJ crystal structure of the IA_3·YPrA complex and plotted onto the Ramachandran diagram. Several of the non-clustered residues are identified.

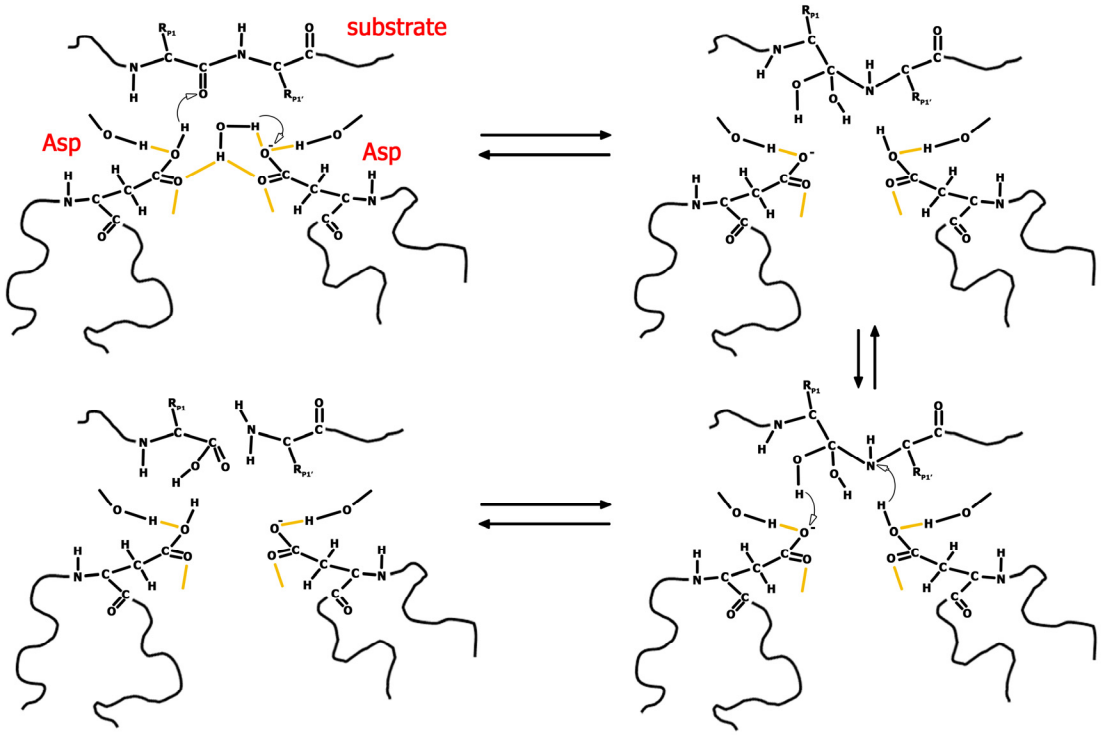


Figure 1-2: Proton transfer 'push-pull' mechanism used in aspartic proteinases to cleave peptide bonds. Illustrated here is the movement of protons within this mechanism. Hydrogen bonds are shown as yellow lines.

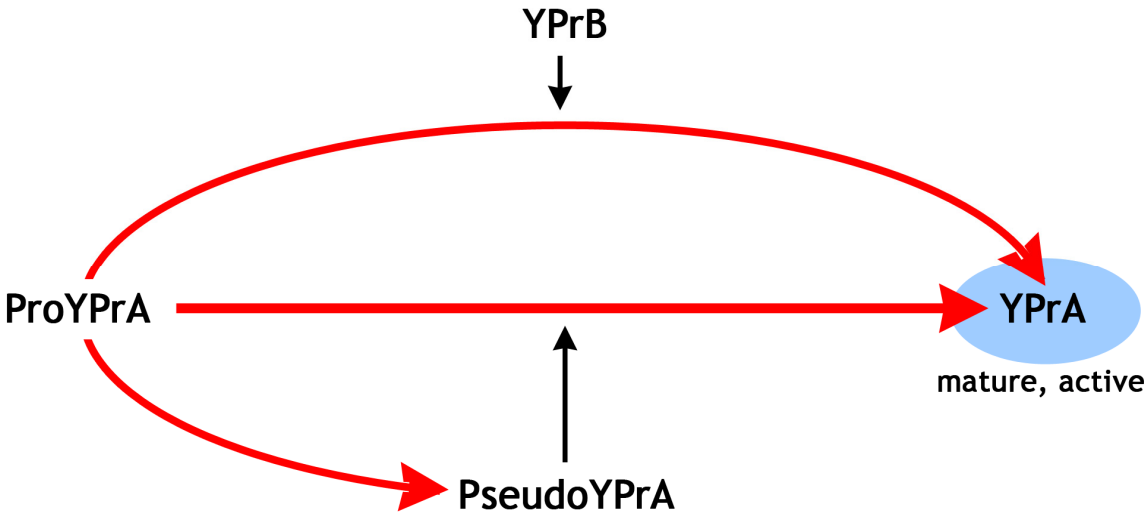


Figure 1-3: Activation of ProYPrA to mature active YPrA. Illustrated here are two pathways by which YPrA could become active: by cleavage of the prosegment by YPrB, or by the formation of PseudoYPrA which could then cleave the prosegment from ProYPrA resulting in mature, active YPrA.

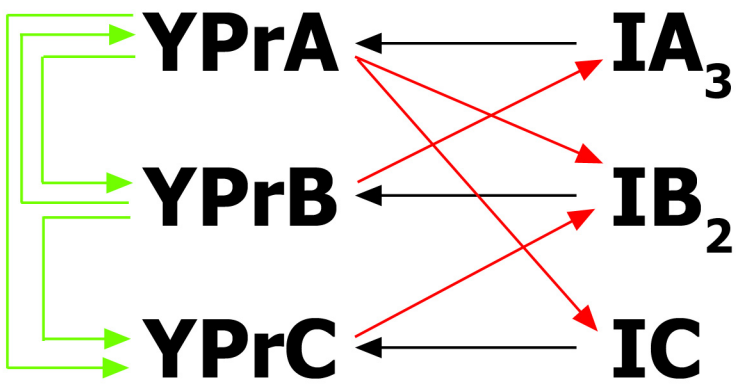


Figure 1-4: Interplay between YPrA, YPrB and YPrC. Inhibitors of these proteases are symbolized by black arrows. Red arrows indicate inactivation of the inhibitor as it can be a substrate for another protease. Green arrows indicate a role of the proteases in the processing and maturation of each other.

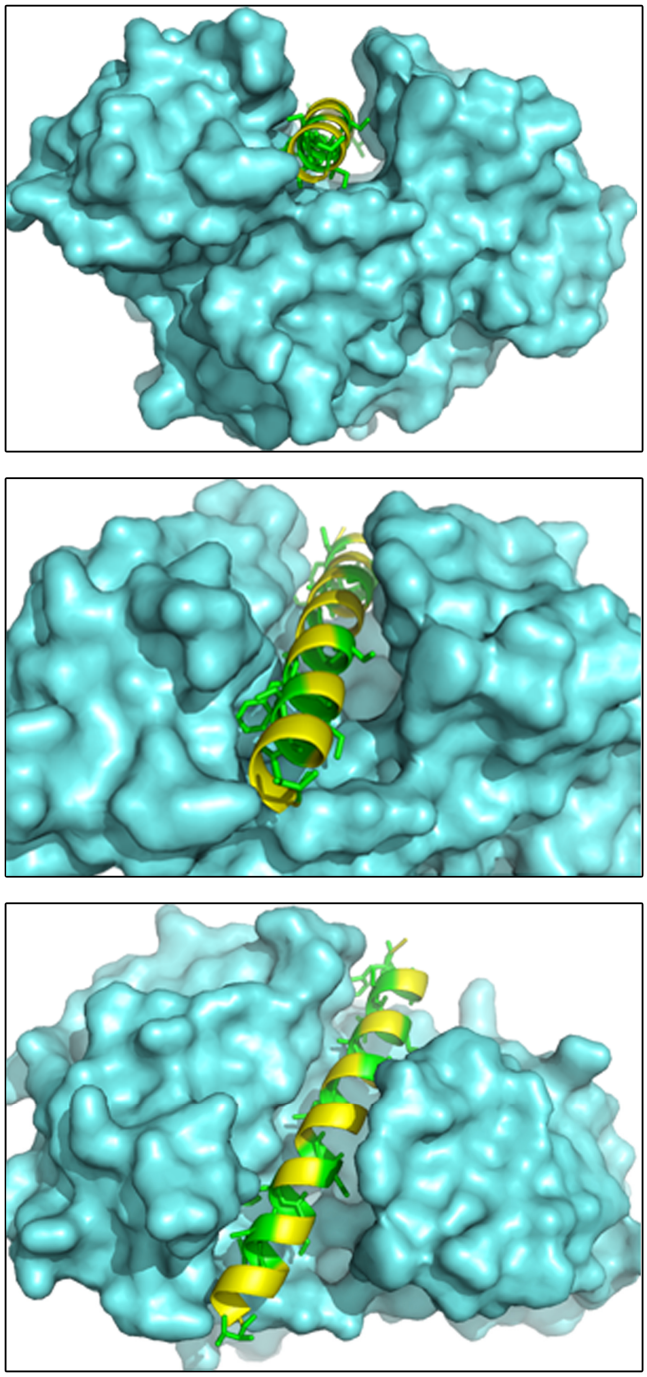


Figure 1-5: Views of IA_3 bound into YPrA from the 1DPJ crystal structure. YPrA in cyan is wrapped around the IA_3 N-terminal helix in yellow and green. Hydrophobic residues in the IA3 helix are indicated as green stick projections. These views of YPrA and IA_3 were generated using PyMOL.

CHAPTER 2
CHARACTERIZING THE RESIDUE LEVEL FOLDING OF THE INTRINSICALLY UNSTRUCTURED IA_3[1]

Introduction

IA_3 is a 7.7 kDa protein that is found in the cytoplasm of *Saccharomyces cerevisiae* and is an endogenous inhibitor of the yeast aspartic proteinase, YPrA. Previous biophysical studies have shown that in the absence of YPrA, IA_3 exists as an unstructured protein in solution (*9, 10*). In the X-ray crystal structure of the YPrA·IA_3 complex (PDB accession number 1DPJ), only the N-terminal residues Thr 3 through Lys 31 of the sixty-eight amino acid protein are seen; this N-terminus forms an α-helix in the active site cleft of YPrA. The remaining C-terminal residues are believed to be disordered since electron density from this region is not observed in the crystal structure (*10*). Only the thirty-four amino acid N-terminus of IA_3 is necessary to inhibit YPrA with the same nanomolar potency as the full length wild-type IA_3 (*11*).

Using far-UV circular dichroism (CD) we have previously shown that 2,2,2-trifluoroethanol (TFE) facilitates a two-state transition in IA_3 from unstructured protein to α-helix with a transition midpoint around 16.3% v/v TFE (*9*). Due to the global nature of CD, these studies could not establish which residues in the IA_3 sequence became helical, or whether the transition is uniformly two-state for all the residues.

NMR chemical shift data from ^{15}N heteronuclear single quantum correlation (^{15}N-HSQC) experiments on proteins contain important structural information (*137*). Folded proteins show greater distribution of chemical shift dispersion, while spectral cross peaks of unfolded proteins are less dispersed with possible overlap in NMR frequencies (*138, 139*). Here we use chemical

[1] This chapter is a published work (*Characterizing the residue level folding of the intrinsically unstructured IA_3.* Ganesh OK, Green TB, Edison AS, Hagen SJ.; Biochemistry. 2006 Nov 14;45(45):13585-96.), and is republished here with permission from the American Chemical Society.

shift data from ^{15}N-HSQC TFE titrations in conjunction with singular value decomposition (SVD) to monitor the transition of IA_3 in the presence of TFE (*140, 141*). Assignment of the protein through triple-resonance methods at intermediate TFE concentrations was used to aid in tracking the cross peaks as the TFE concentration varies. We also use parameters derived from a two-state fit to the TFE-driven transition to extrapolate to the high and low TFE states to extend our current understanding of how the protein folds.

The tracks taken by individual chemical shifts, in response to the addition of TFE contain a large amount of information, and it can be difficult to extract a clear picture of either the general trends or the individual, residue-level departures from the general trends. One approach to this problem could be to use chemical shift mapping (CSM) to identify residues that are undergoing large changes. CSM is also generally used to identify interactions between proteins and nucleic acids, other proteins, or membranes (*142-145*). In this method, the chemical shift of each residue is assessed and mapped onto an image of the primary structure or the folded structure of the protein (*142*). In the present case, we do not have a structure for the high-TFE state, and we cannot assume that this structure is identical to the crystal structure obtained for the biologically active complex. More importantly, CSM identifies the residues involved in an interaction but does not reveal the nature of the transition or indicate how closely an individual residue does or does not follow the chemical shift of other assigned residues.

As an alternative, we used singular value decomposition (SVD). SVD is a mathematical tool for identifying the most important patterns in a dataset (*146*). It breaks apart a two-dimensional matrix of data (in this case, chemical shifts recorded for an array of sequence positions and TFE concentrations) into a sum of simpler matrices. Each of those component matrices contains a single two-dimensional pattern, *i.e.* a particular pattern of chemical shifts

versus sequence position, which varies with a particular TFE dependence. The original data matrix, with all of its sequence-dependent and TFE-dependent variations, can be reconstituted exactly by adding together all of the component matrices. The primary benefit of applying SVD to a complex data set is that it essentially identifies all the major patterns that are present in the data, and ranks them in order of descending strength or significance. SVD has been an invaluable tool in other areas of biophysical spectroscopy, such as transient absorbance spectroscopy (*146*).

By applying SVD to the chemical shift data, we show that the major transition is two-state, and that it occurs in the same residues necessary to inhibit YPrA. Additionally, another component to the folding is shown to be sensitive to changes in the amide proton chemical shift, and may be indicative of changes to the hydrogen bonding of backbone amide protons. A biochemical interpretation of the residue interactions, combined with the global description provided by SVD, results in a relatively complete residue-level picture of the structural changes induced by TFE. Our results demonstrate how the application of SVD to a two-dimensional NMR dataset can efficiently extract both a detailed description of the global structural changes in a molecule and the rather small, residue-specific deviations from that global behavior.

Experimental Procedures

Sample Preparation

IA_3 was expressed in *E. coli* and purified as previously described by Phylip *et al.*(*11*). IA_3 samples were prepared at a concentration of 1 mM sample in 250 µL of 50 mM PO_4 buffer, pH 5.6 as described previously (*9*), with the inclusion of increasing volumes of TFE. A TFE titration from 0% to 23% v/v was made by the addition of constant volumes of 5 µL of TFE to the IA_3 sample. In addition, a fresh IA_3 sample was prepared at the endpoints and middle of the TFE titration for triple resonance experiments.

Nuclear Magnetic Resonance Data Acquisition

All NMR data were collected on Bruker Avance spectrometers in the Advanced Magnetic Resonance Imaging and Spectroscopy (AMRIS) facility at the University of Florida's McKnight Brain Institute. ^{1}H data were referenced to DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid) (0.0 ppm). ^{15}N and ^{13}C 0.0 Hz frequencies were obtained through indirect referencing to DSS using the chemical shift ratios provided by the BMRB database (*9, 147*). All NMR experiments were run at 20 °C.

Two-dimensional ^{15}N-HSQC (*148, 149*) spectra of the TFE titrations were collected at 750 MHz on a Bruker Avance (DRX)-750 console in a 17.6 Tesla magnet using either a 2.5 mm or 5 mm TXI probe. The ^{1}H carrier frequency was centered on water which was reduced using a 3-9-19 WATERGATE sequence (*148*). ^{15}N-HSQC spectra were collected with 2048 complex points in the ^{1}H dimension, and 256 complex points in the ^{15}N dimension. Quadrature detection in the indirect dimension was achieved using the States-TPPI method (*150*). All 2D spectra were collected with a 10.69 ppm (8012.82 Hz) spectral width in the proton dimension, and 30.0 ppm (2280.30 Hz) in the nitrogen dimension. The carrier frequencies for ^{1}H and ^{15}N were centered on 4.86 ppm and 116.68 ppm respectively.

Triple resonance data were collected at 600 MHz on a Bruker Avance (DRX)-600 console in a 14.1 Tesla magnet equipped with a 5 mm TXI probe. Spectra were collected with a 11.02 ppm (6613.76 Hz) spectral width in the proton dimension, and 21.7 ppm (1315.79 Hz) in the nitrogen dimension. For the HNCACB (*151*) and CBCA(CO)NH (*152*) the spectral widths were 50.9 ppm (7692.31 Hz) and carrier frequencies were 42.51 ppm for the carbon nuclei. The HNCO (*153, 154*) carbonyl spectral width was 9.5 ppm (1428.57 Hz) with a carrier frequency of 176.09 ppm. Quadrature detection in the ^{15}N dimension was achieved using the Echo-AntiEcho method (*149, 155, 156*), and in the ^{13}C dimension using the States-TPPI method. All three-

dimensional spectra were collected with 1024 complex points in the $^{1}H^{N}$ dimension. HNCO spectra were collected with 32 complex points in the ^{15}N dimension and 30 complex points in the ^{13}C dimension. HNCACB spectra were collected with 32 complex points in the ^{15}N dimension and 50 complex points in the ^{13}C dimension. CBCA(CO)NH spectra were collected with 18 complex points in the ^{15}N dimension and 50 complex points in the ^{13}C dimension.

Processing and Referencing of NMR Data

Multi-dimensional spectra were processed using NMRPipe (*157*) and the resulting spectra were analyzed with NMRView software (One Moon Scientific, Trenton, New Jersey) (*158*). We applied cosine-squared apodization functions in the direct and the indirect dimensions during processing of the 2D ^{15}N-HSQC spectra. To increase the digital resolution, spectra were zero-filled to 4096 points in the proton dimension, and to 1024 points in the indirect dimension. Polynomial baseline correction was applied in both dimensions.

For the CBCA(CO)NH, HNCACB, and HNCO triple resonance spectra, the direct dimension was zero-filled to 2048 points. Cosine-squared apodization and zero-filling were applied in all dimensions. Linear prediction was used to expand the data in the indirect carbon dimension only. We used model-free baseline correction to remove noise from the resulting 3D spectra (*157, 159*).

Assignments and Cross Peak Tracking

Resonance assignments were made manually using standard triple resonance methods. We tracked individual residue chemical shifts in the 2D ^{15}N-HSQC as a coordinate from the centroid of the cross peak. The cross peak $^{1}H^{N}$ and ^{15}N coordinates of individual residues were tracked as they varied with TFE concentration, and displayed as two-dimensional tracks. We tracked only cross peaks for which an unambiguous assignment was possible and where the cross peak could

be followed through the whole TFE titration. Of the 68 residues in IA_3, 40 cross peaks were assigned and tracked in this study.

SVD Analysis

Random-coil chemical shifts and sequence-dependent correction values were obtained from methods developed by Schwaringer *et al.* (*160, 161*), using a program written in PERL by the authors. The chemical shift data, after this correction, can be represented by two separate matrices containing the $^{1}H^{N}$ and ^{15}N values respectively. Each matrix has 40 rows corresponding to the 40 different residues tracked, and 16 columns corresponding to the 16 different TFE concentrations studied; the element in row *i* and column *j* then gives the value of the NMR chemical shift for residue *i* at TFE concentration *j*. We then decomposed each matrix by singular value decomposition (SVD). SVD is a method for producing a unique factorization of a given matrix as the product of three basis matrices (*137, 162-164*): it extracts the essential features of a data matrix, preserving the significant information in the matrix while sharply reducing the number of parameters needed to represent that information. As SVD itself is a purely mathematical and model-independent procedure, it lacks intrinsic bias toward any particular interpretation of the data, and as such reliably identifies interesting features in the data that rise above the noise background.

SVD decomposes a data matrix *A* into the matrix product,

$$A = U \cdot S \cdot V^{T} \tag{2-1}$$

where, for a matrix *A* of size 40 × 16, *U* is a matrix of size 40 × 40, *V* is a matrix of size 16 × 16, and *S* is a matrix of size 40 × 16. The 16 diagonal elements of S are the only nonzero elements and are known as the singular values, s_n (with n = 1 to 16). The columns of *U* comprise an

orthonormal set of 40-element vectors, u_n, (n = 1 to 40) and the columns of V comprise an orthonormal set of 16-element vectors v_n (n = 1 to 16).

From Equation 2-1, the data matrix A can be reconstructed exactly as a sum.

$$A = \sum_{n=1}^{16} u_n \cdot s_n \cdot v_n^T \tag{2-2}$$

In our case the vectors u_n may be interpreted as functions of sequence position (residue number), and the vectors v_n as functions of TFE concentration. The dataset can then be reconstructed from the series:

$$A(residue, TFE) = u_1(residue) \cdot s_1 \cdot v_1(TFE)^T + u_2(residue) \cdot s_2 \cdot v_2(TFE)^T + \ldots \tag{2-3}$$

Since the singular values are ordered from largest (s_1) to smallest (s_{16}), and the u and v vectors are normalized, the weight or importance of each successive term in this series is smaller than that of the preceding term. More precisely, the sum of the first n terms in the series provides the best (in the least-squares sense) n-component representation of the complete dataset A (*164*).

Fitting and Rotation Matrices

If the response to TFE is primarily the induction of a two-state folding transition, then this transition should be represented in the *V(TFE)* data and can be described by the fraction of molecules folded, F_f (*165*).

$$F_f = \left[1 + \exp\left(-\Delta G / RT\right)\right]^{-1} \tag{2-4}$$

where ΔG, the free energy of unfolding, is given by $\Delta G = \Delta G_0 + m[TFE]$ in the conventional case of a linear relationship between folding free energy and TFE concentration (*166*). This two-state model requires only two free parameters, m and ΔG_0, to fit the data. A two-state transition will be represented by two SVD components. Therefore we expect the v_1 and v_2 vectors obtained from both the $^1H^N$ and ^{15}N datasets to have the TFE dependence of Equation 2-4. We fit these

two vectors in both datasets simultaneously in order to find the parameters. The residue dependence should then follow $s \cdot v_n(TFE) = a_n + b_n F_f$ for the constants a_n, and b_n, $n = 1, 2$.

Extrapolation to TFE → + ∞ or TFE → - ∞ should then result in limiting spectra. TFE → + ∞ would correspond to a high TFE state with a large population of folded molecules (essentially $F_f \approx 1$); TFE → - ∞ would correspond to a low TFE state with a high population of unfolded molecules ($F_f \approx 0$). The limiting values of v_1 and v_2 can be calculated as $sv_n = a_n + b_n$ for the high TFE state and $sv_n = a_n$ for the low TFE state. Combining the limiting TFE dependence with the u_1 and u_2 vectors produces the high and low limiting spectra.

$$limiting\ spectra = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \cdot \begin{bmatrix} sv_1 \\ sv_2 \end{bmatrix} \tag{2-5}$$

The first two *V(TFE)* components were fit extremely well by a two-state function (Equation 2-4). The vectors of each of the basis sets *U* and *V* produced from the SVD are orthogonal, therefore the physical phenomena of interest in the data are most likely spread across several of the SVD components. Accordingly it was noted that rotation of the components of the $^1H^N$ and ^{15}N basis sets would minimize the deviations between the two-state fit and the first two SVD components, resulting in more accurate ΔG_0 and *m*-values. A rotation of the basis vectors can preserve the orthogonality of the set, while ensuring that separate physical phenomena are maximally represented in the minimal number of components (one or few). The rotation, however, is not a fit or approximation and does not change or reduce the information contained in the data.

Rotation of *V* requires us to find an optimal rotation matrix *R* (*164*). To find *R*, we rotated the dataset while simultaneously fitting to the two-state equation (Equation 2-4). In doing so, we were able to find a new set of basis vectors that spanned the same vector space, yet also provided

the best fit to the two-state equation for v_1 and v_2, while preserving the orthonormailty of v_3. The third component would then also contain deviations which did not conform to a two-state fit.

Results

Initial NMR Analysis

Using the ^{15}N-HSQC, individual residue chemical shifts were tracked as a $^{1}H^{N}$ or ^{15}N coordinate and as a function of TFE. We subtracted random coil values (*160*) from all chemical shift values and corrected the resulting values for sequence dependent effects (*161*). Several residues could not be tracked throughout the titration, owing to overlap of cross peaks at some concentrations, or due to the difficulty of completely assigning all peaks in the TFE titration. However, 40 of the 68 residues in IA_3 were fully tracked through the titration. Initial analysis of these residue tracks indicates that there is a general tendency towards an α-helical structure according to cut-off values given by Wishart and Sykes (*137*). As shown in Figure 2-1, the tracks of many of the residues throughout the protein exhibit a large curvature while moving in a generally α-helical direction with the addition of TFE. $^{13}C^{\alpha}$ chemical shifts at a few TFE concentrations verify the helical transition (data not shown). However, several residues do not follow this general trend. Additionally, residues such as Asp 22, Asp 28, Phe 30, Ala 34, Ala 45, Glu 49, Tyr 53, and Asn 58 move in a manner contrary to the general motion of the other tracked residues of the protein.

Evaluation of SVD Analysis

The ^{15}N and $^{1}H^{N}$ coordinates of the chemical shift tracks were collected as separate data matrices as described in the Experimental Procedures section. We used SVD to decompose the chemical shift data into three matrices, *U*, *S*, and *V*. The *V* matrix describes the TFE dependence of the chemical shift, with the columns of *V* representing the TFE dependence. The columns of the *U* matrix represent the residue dependence of the chemical shift. *S* is a diagonal matrix of

singular values, with diagonal elements representing the weights for the corresponding columns of the U and V matrices. Hence the n^{th} column of the V matrix, together with the n^{th} column of U and the n^{th} diagonal element of S, together comprise the n^{th} SVD component of the data (Equation 2-2).

The weights of the first three components from the SVD describe more than 96% of the information in the chemical shift data matrix (Figure 2-2). All remaining higher order SVD components are less than 1% of the total chemical shift data. Further, the TFE dependence and residue dependence of the higher components appeared randomly distributed around zero, indicating that these components essentially reflect noise in the data (Figure 2-3). Therefore, in the discussion that follows we focus only on the first three components of the SVD analysis.

TFE Dependence

The overall effect of TFE on the coil to helix transition of IA_3 can be evaluated from Figure 2-4. It shows the TFE dependence of the chemical shifts as a function of increasing TFE concentration for the first three components of the SVD analysis. Plots of the first three components of this TFE dependence revealed that both the $^{1}H^{N}$ and ^{15}N exhibited similar responses to TFE. The first two components of the TFE dependence of both the $^{1}H^{N}$ and ^{15}N chemical shifts exhibit sigmoidal trends with increasing amounts of TFE. These movements are strongly indicative of a two-state transition (*167*), where one state is progressively depopulated in favor of a second state as the TFE concentration rises. The third component in each figure exhibits a weak maximum near 15% v/v TFE. This third component corresponds to only a small portion of the signal, around 4.1% of the $^{1}H^{N}$ and 2.6% of the ^{15}N chemical shifts (Figure 2-2).

We fit the first two components of both the $^{1}H^{N}$ and ^{15}N simultaneously to a two-state folding-unfolding model. This model implicitly assumes that the free energy of unfolding is linear in the TFE concentration. The fit yielded a unfolding free energy in water, ΔG_0, of -21.2

kJ/mol with an *m*-value, which measures the helix stabilizing effect of TFE, of 1.2 $(kJ\ mol)^{-1}$ at 20°C. Simultaneous fitting of the first and second components of the TFE dependence of the chemical shift for both $^{1}H^{N}$ and ^{15}N cross peak movements yielded a TFE refolding midpoint of 18.3%, which was in agreement with the TFE refolding midpoint obtained from far-UV circular dichroism (CD) studies of IA_3 in increasing concentrations of TFE (*9*). Fitting of $^{1}H^{N}$ and ^{15}N datasets separately yielded TFE refolding midpoints of 18.4% and 18.1% respectively, indicating that the model provides a robust fit to the data (Table 1).

Residue Dependence of Chemical Shift

Plotting the residue dependence of the chemical shifts versus the amino acid sequence provides insight into the folding of IA_3 at the residue level (Figure 2-4). Bar plots of the first through third components of the residue dependence on the chemical shift proved to be very revealing. The first component of the $^{1}H^{N}$ residue dependence indicates that IA_3 exhibits a large tendency towards α-helical structure throughout the length of the sequence. This tendency is present, albeit not as clearly, in the first component of the ^{15}N residue dependence. The second component of the residue dependence for the $^{1}H^{N}$ and ^{15}N can be broken down into distinct regions. In $^{1}H^{N}$ the N-terminal half of the protein (residues 1-28) exhibits a negative trend. This trend appears to be minimized or reversed in the C-terminal portion of the protein. A central region of positive values, with flanking negative values, can also be seen around residues 28-31. The second component of the ^{15}N residue dependence also shows this trend with the N-terminus showing generally larger negative changes as compared to the C-terminus, with residues 29 and 30 being positive. The third component residue dependence indicates negative deviations in both ^{1}H and ^{15}N.

The limiting high and low TFE spectra obtained from the fit and the SVD components are shown in Figure 2-5. Figure 2-6 shows the 0% and 23% ^{15}N-HSQC TFE spectra, along with the

same spectra as reconstructed from just the first three SVD components (i.e. omitting all higher order components). The original and reconstructed spectra agree well, illustrating that discarding the higher order components of the SVD has a negligible effect on the initial NMR data, and that the rotation included in our SVD treatment preserves all the information in the data.

The Third SVD Component

Panels G and H of Figure 2-4 illustrate the deviations seen in the third component. All the deviations seen in this component are small in both the $^{1}H^{N}$ and ^{15}N dimension. Residues such as Thr 3, Gly 40, Glu 49, and Tyr 53 show bars which are very small or nearly zero. A very small third component, or the apparent lack of a third component, indicates that the chemical shift tracks of these residues can be described primarily by the first and second components, both of which were well fit by a two-state transition. Ser 9, Val 26, Gly 62 and other residues which are influenced by this third component are indicated by very large bars in panels G and H.

Reconstruction of the NMR chemical shift tracking spectra based only on the third SVD component and residuals allowed us to evaluate the third component (Figure 2-7). Residues which have no third component in $^{1}H^{N}$ and ^{15}N would be represented here as a track in which all the points lie on top of each other. Closer inspection of such a track would reveal that the points have small deviations from one another and are randomly distributed. Figure 2-7 provides several examples of this: Ala 23, Phe 30, Gly 40, Thr 43, Glu 49 and Tyr 53. Residues such as Ser 9, Val 26, Lys 32, Met 33, Ala 34, and Gly 62 in Figure 2-7 are examples of large deviations of the third component seen in Figure 2-4. From Figure 2-7 we can see that deviations in the third component primarily manifest themselves as chemical shift movements along the $^{1}H^{N}$ axis.

Discussion

The native IA_3 provides an ideal model system to study the conformational features of intrinsically unstructured proteins. It is known to be unfolded in solution, however the N-

terminus of the protein forms an α-helix upon binding to YPrA (*10*). We have previously shown by far-UV CD that TFE can induce an α-helical structure in IA_3 (*9*), and here we expand this work using NMR spectroscopy along with SVD analysis to understand changes in chemical shift behavior to gain an insight into the folding of IA_3.

TFE Driven Transition in the NMR Data

It is well understood that $^1H^N$ shifts are influenced by local electronic environment (*168*) and ^{15}N shifts are sensitive to amino acid type and the effect of neighboring amino acids. Both $^1H^N$ and ^{15}N shifts are influenced by hydrogen bonding and the φ/ψ_{i-1} backbone dihedral angles (*169-171*). In general, cross peaks which move to lower frequencies are indicative of helical structure, while cross peaks which move to higher frequencies indicate beta structure (*168, 170*).

Subtraction of random coil and sequence dependent effects from the chemical shift data allowed us to evaluate the change in secondary structure of a given residue due to the increase of TFE. According to the analysis of Wishart *et al.*, a residue track starting near the origin and ending in the upper-right of each graph in Figure 2-1 would clearly indicate a transition from coil to helix. Titration tracks generated by following residue chemical shifts as a function of TFE (Figure 2-1) indicate that the general chemical shift movements of most of the residues of IA_3 appear to be consistent with the formation of an α-helical structure as the concentration of TFE in solution is increased. Most of the paths of the individual residue $^1H^N$ and ^{15}N cross peak titration tracks contain this positive slope movement in a $\Delta\,\delta^1H^N$ and $\Delta\,\delta^{15}N$ plot. The residue tracks of Lys 18, Gly 21, Lys 24, Val 26, Ser 27, Lys 32, Ser 35, Lys 41, Gly 62 and Gly 64 exemplify this behavior.

However, this type of analysis does not cover the large variation in chemical shift tracks when we consider all of the tracked residues. Many of the residue chemical shift tracks maintain some degree of a positive slope as the TFE concentration is increased. This indicates that there

is a large degree of helical tendency (*170, 172*), yet very few of these tracks have a strictly positive linear slope. In Figure 2-1, the residue tracks of Asp 4, Glu 10, Ser 15, Lys 16, Ala 29, and Ala 63 have slopes which are nearly horizontal, showing that the chemical shift movement occurred in the $^{1}H^{N}$ dimension only. Val 8, Ser 9, Gln 13, and Met 33 have residue tracks which bend towards the vertical axis at later points in the titration. More extreme cases of this bending can be seen in the tracks of Ala 34, Asp 45, Ser 48, and Asn 58 which bend past the vertical and turn downfield on the $^{1}H^{N}$ axis. The track of Glu 17 is nearly vertical. Also the residue tracks of Thr 3, Ser 14, Ala 23, Asp 37, Asp 39, Thr 42, Thr 43, Asp 44, and Asp 46 show positive linear slopes, yet start from positions father away from the $^{1}H^{N}$ and ^{15}N origin thereby confusing the standard coil to helix argument.

A few residue tracks are not consistent with the conventional coil to helix transition. Residues Asp 22, Asp 28, Phe 30, Ala 34, Ala 45, Ser 48, Glu 49, Tyr 53, and Asn 58 move in a manner completely contrary to what is expected for an ideal coil to helix transition. Asp 22, Asp 28, Ala 34, and Asn 58 make drastic hook-like turns as the TFE concentration increases. Ser 48, Glu 49, Tyr 53 do not have long tracks, however they move downfield in $^{1}H^{N}$ and upfield in ^{15}N. Finally, the residue track of Phe 30 is linear but moves downfield as the TFE concentration increases, and is the only residue track to have this kind of movement.

Qualitatively, many of the chemical shifts move in a manner consistent with a coil to helix transition. Following the chemical shift tracks of each individual residue, however, it is unclear whether these shifts constitute a uniform or global, two-state transition. This led us to use SVD on the chemical shift tracking data.

Unraveling the Transition

In our previous study, SVD analysis of the far-UV circular dichroism data yielded basis states consistent with a high TFE folded state and a low TFE unfolded state. In the presence of

TFE, IA_3 transitions from an initial disordered state to a helical conformation through a two-state transition, with the midpoint of this transition observed at 16.3% TFE (*9*).

From the individual residue chemical shift titration curves in Figure 2-1 it is difficult to see the two-state transition. Using SVD allowed us to focus on the important cooperative transitions inherent in the chemical shift data without being misled by the differences between the chemical shift tracks of individual residues. This approach is fundamentally different from chemical shift mapping in that it allows us to identify and quantify the transition that IA_3 is undergoing in increasing amounts of TFE. Additionally, the use of SVD allowed us to compare and analyze the TFE and residue dependence inherent to the residue chemical shift movements separately.

SVD analysis of the 2D ^{15}N-HSQC data indicated that the first two components of the TFE dependence of both the $^{1}H^{N}$ and ^{15}N chemical shifts could be fit to a two-state thermodynamic transition, as in the earlier CD study *(9)*. We simultaneously fit these four datasets to yield a TFE folding midpoint of 18.3%. This value is in close agreement with the value obtained from earlier CD studies. In Green, *et al.* (*9*) the far-UV CD studies of IA_3 in increasing concentrations of TFE exhibited very broad transitions which have been known to influence ΔG_0 and *m*-values (*173*). This, coupled with differences in sample preparation, can explain the small difference in the TFE folding midpoint found between these two techniques.

From Figure 2-2, we can see that the bulk of the TFE dependence can be represented by the first two TFE components from the SVD analysis. This indicates that, to a large extent, most residues can be described as participating in a global, two-state transition.

Different Behavior in the Two Halves of IA_3

Our previous study of IA_3 had shown that at 0% TFE the bulk of the residues in IA_3 had a small tendency towards α-helical structure in ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$, and $^{1}H^{N}$ chemical shifts (*9*). From panels C and D in Figure 2-4, we see that the first component of the $^{1}H^{N}$ residue

dependence of the chemical shift from the SVD analysis does indeed agree with our previous assessment (*9*). The first component of the ^{15}N residue dependence on the chemical shift qualitatively agrees as well, however the SVD analysis indicates that several residues around the middle of the sequence may have a propensity towards a different structure.

The data presented by Green *et al.* indicated that at 0% TFE, there was a moderately populated α-helix in the C-terminus of the protein (*9*). The amplitudes of second components of both the $^{1}H^{N}$ and ^{15}N residue dependence over the whole TFE range separate the N-terminus from the C-terminus of the sequence (panels E and F in Figure 2-4). In both panels the N-terminus of the sequence shows larger negative bars than the C-terminus, indicating that the N-terminus shows a larger propensity towards a helical structure, with residues around the middle of the sequence exhibiting a positive trend. As the percentage of TFE is increased, the ^{15}N - HSQC chemical shift cross peaks of these residues move downfield along the $^{1}H^{N}$ and ^{15}N frequency axes.

However, the difference spectrum of the limiting high and low TFE states (Figure 2-5) shows larger negative bars in the N-terminus of the protein. The first implication of this is that the N-terminal residues are undergoing a more pronounced TFE induced transition than residues in the C-terminus. Secondly, since the bars are predominantly negative in the N-terminus, this suggests that the N-terminus shows a larger overall helical change than the C-terminus as the TFE concentration is increased. Comparisons of the frequency shifts of these residues with limiting values from Wishart and Sykes (*170*) indicate that this region is largely α-helical.

The crystal structure of the YPrA·IA_3 complex shows only residues 3-31. Our analysis shows helical propensity throughout the sequence with the region around the middle acting as a separator between the two halves of the protein. Some difference in structure appears to exist in

this region separating the N and C termini; this is exactly the boundary found in the crystal structure.

Phylip *et al.* (*11*) had previously shown that the inhibitory action of the protein could be induced with only the first 34 residues of the N-terminus, with a K_i of less than 0.1 nM at pH 4.5. From the crystal structure of YPrA·IA_3, it is understood that formation of an α-helix in the N-terminus of the protein is necessary for inhibition of YPrA. The folded state of IA_3 induced by TFE may not be identical with the structure found in the biologically active complex. However, this analysis shows that the TFE induced two-state transition from a coil to a helical structure occurs in the same residues seen in the YPrA·IA_3 crystal structure, and is consistent with the formation of an α-helix in the N-terminus of IA_3. The predisposition of the N-terminus of IA_3 to form an α-helix, which is indicated by our analysis, may be a direct consequence of its sequence and is necessary for its function (*174*). Results of this study indicate that TFE is driving the same kind of structural transition that occurs during the inhibition of the protease.

Additional Folding Phenomena?

While the main transition seen in the chemical shifts of the residues of IA_3 appears to be two-state, a third component did become apparent from the SVD analysis. This component represented less than 5% of the observable transition in the data, yet stood out from the analysis since the first three singular values were much larger than the subsequent singular values (Figure 2-2).

The third SVD component of the TFE dependence of the NMR chemical shift showed a weak positive peak, as if an additional state becomes populated up to 15% TFE, then depopulates as the concentration of TFE increases further (Figure 2-4). This component contains all deviations from perfect two-state folding since the rotation procedure (see Experimental Procedures) optimized the first and second components for their fit to a two-state transition. Any

trend in the TFE-dependence of the data that does not match up with two-state behavior will appear in this third component (or else in a higher order component). Several factors could contribute to this third component including differential line broadening (*175*), direct interactions with carboxylate moieties on acidic side chains (*176, 177*), and any other deviation from a linear two-state transition.

It is well-known that 2,2,2-trifluoroethanol (TFE) stabilizes secondary structures in peptides (*178-181*). While TFE has been shown to stabilize many different types of secondary structures, several authors have demonstrated that there exists a good correlation between helical propensity and primary sequence (*175, 182, 183*). Although the exact mechanism of action of TFE in the inducement of helical structure is not fully understood, several theories of how TFE stabilizes helical structures have been proposed and continue to be studied. TFE may induce secondary structure simply by offering a more favorable environment for secondary structure formation, possibly by reducing the hydrophobic effect or decreasing the affinity of an unstructured polypeptide to hydrogen bonding with the surrounding solvent (*173*). However, TFE is also known to be a better hydrogen bond donor and a poorer acceptor than water (*173, 184*). These effects may assist or detract from folding by stabilizing transient structures.

One possibility to explain the third TFE-dependent component of the SVD analysis could be regarded as a low populated intermediate prior to the two-state folding in which the carboxylate moieties shift their transient hydrogen bonds from the amide protons along the backbone to TFE. This frees the amide protons, allowing them to proceed with *i*+4 hydrogen bonding found in the α-helix. The weak peak at 15% in both the $^{1}H^{N}$ and ^{15}N could represent the limit at which the amide proton is most free of the transient hydrogen bonds, just before it

participates in cooperative hydrogen bonding of the α -helix. This could be an interesting area for future research on this system.

Conclusion

We have presented a method to identify folding behaviors using residue level information on IA_3. In this study, we have shown that the use of 2D ^{15}N-HSQC spectroscopy combined with an SVD analysis can be used to identify folding behaviors and is an excellent method for following individual residue movements as a function of perturbant.

The addition of TFE enhances the inherent propensity of IA_3 to form a helical structure. From the above analysis, we determined that the N-terminus of IA_3 has a greater transition towards an α-helical structure than the C-terminus of the protein as the TFE concentration is increased. Using SVD of the ^{15}N-HSQC NMR data, we have shown that a two-state coil to helix transition occurs in the same residues seen in the α-helical structure of IA_3 found in the crystal structure of the YPrA·IA_3 complex.

The folding of IA_3 derived from this study indicates several features that were determined by using SVD analysis of the NMR data. We also show that to a large degree the transition of each individual residue in the protein is two-state like. However, the presence of a weak third component in the analysis indicates that a two-state description of IA_3 folding is not entirely complete.

Table 2-1: Computed ΔG_0, *m*-values, and $TFE_{midpoint}$ from separate $^{1}H^{N}$, ^{15}N, datasets and the combined $^{1}H^{N}$ and ^{15}N dataset.

Dataset	ΔG_0 (kJ/mol)	*m*-value (kJ (mol · % TFE)$^{-1}$)	$TFE_{midpoint}$ (% TFE)
$^{1}H^{N}$	-21.2 ± 7.3	1.15 ± 0.21	18.4
^{15}N	-16.6 ± 6.6	0.92 ± 0.30	18.1
$^{1}H^{N}$ and ^{15}N combined	-20.4 ± 4.1	1.11 ± 0.12	18.3

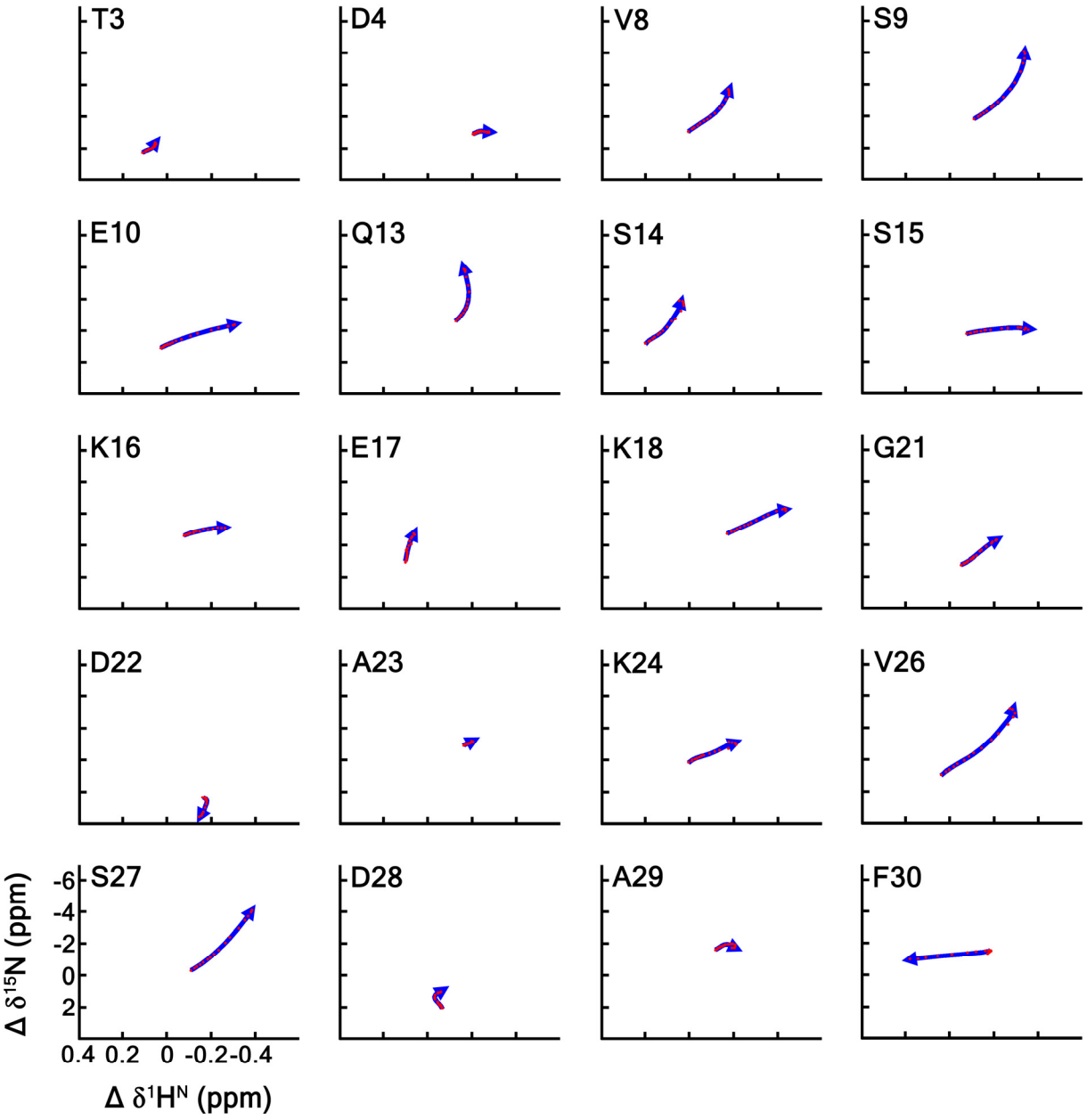


Figure 2-1: Changes in the chemical shifts ($^{1}H^{N}$, x-axis; ^{15}N, y-axis), in response to increasing TFE concentration, for the 40 residues of the IA_3 sequence followed in this study. Random coil values have been subtracted from the $^{1}H^{N}$ and ^{15}N chemical shifts (*160, 161*). Individual residue chemical shifts were tracked as a function of TFE concentration from 0% TFE (v/v) to 23% TFE, via analysis of sequential ^{15}N-HSQC spectra. Red dots indicate data points, arrowhead indicates direction of track as TFE concentration rises.

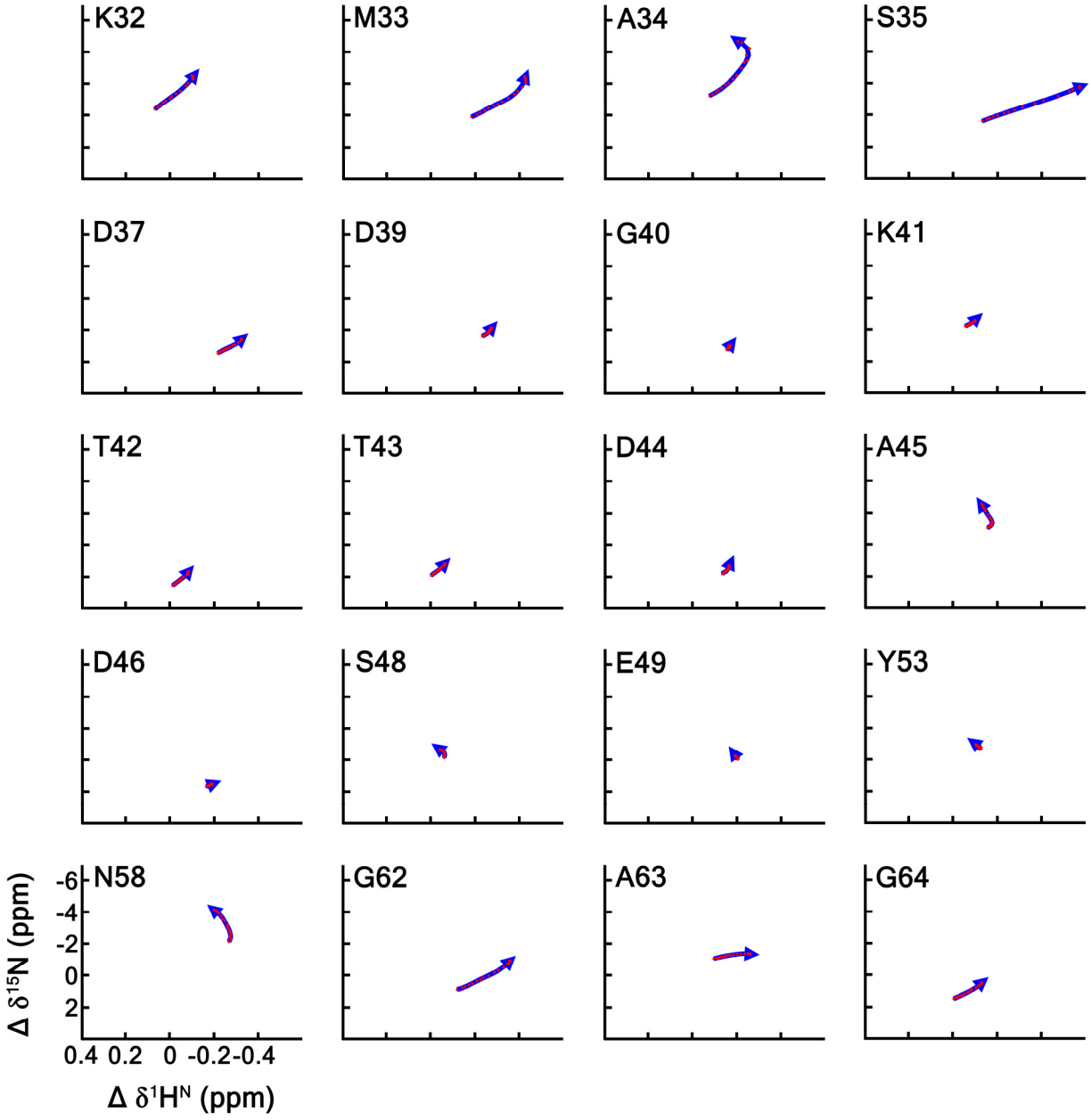


Figure 2-1 continued

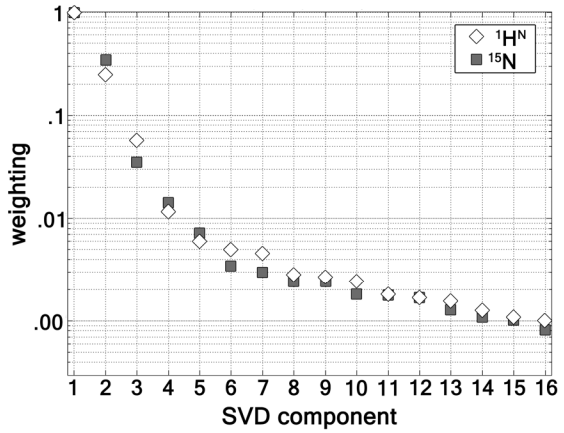


Figure 2-2: Relative weighting of the SVD components obtained from the $^{1}H^{N}$ (◇) and ^{15}N (■) chemical shifts. The relative weighting of component i is defined as *weighting* = s_i/s_1, where s_i is the i^{th} element along the diagonal of the weighting matrix S obtained by SVD of the data matrix A (see Equation 2-1). The first three components of the SVD analysis constitute 96.4% and 96.7% of the $^{1}H^{N}$ and ^{15}N matrices, respectively. Each of the remaining components of the $^{1}H^{N}$ and ^{15}N decompositions is under 1%. This analysis indicates that the $^{1}H^{N}$ and ^{15}N chemical shift matrices can be represented by the first three components of the SVD analysis.

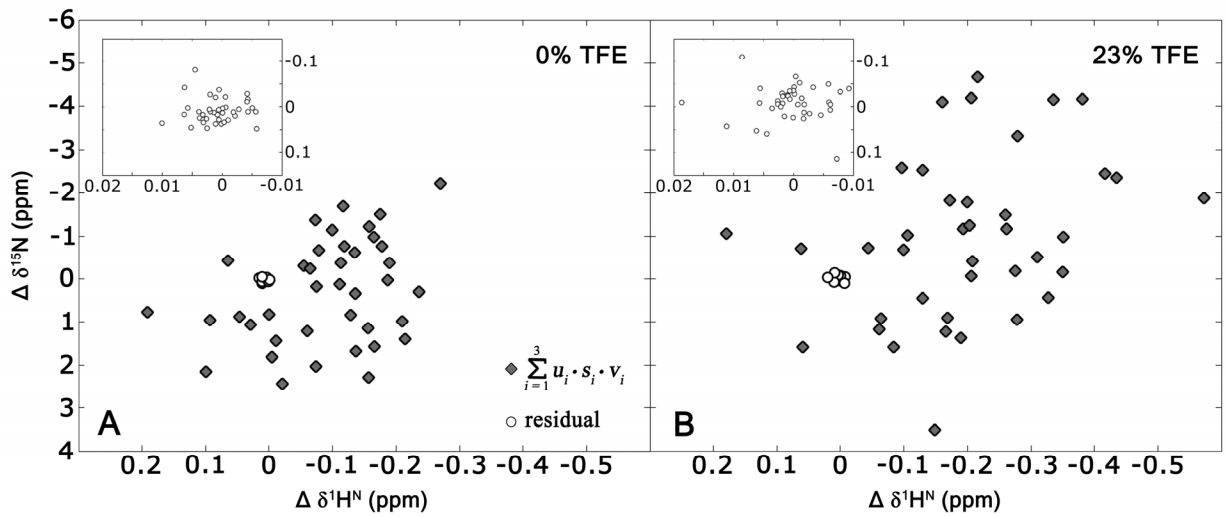


Figure 2-3: Comparison of the first three components of the SVD to the higher order components indicating that the fourth and higher order components could be noise and play a small role in the signal. Residual (○) is the difference between the complete dataset and the reconstruction of the first three SVD components (◆). Panel A is at 0% TFE, while panel B is at 22% TFE. The insets give an expanded view of the residual.

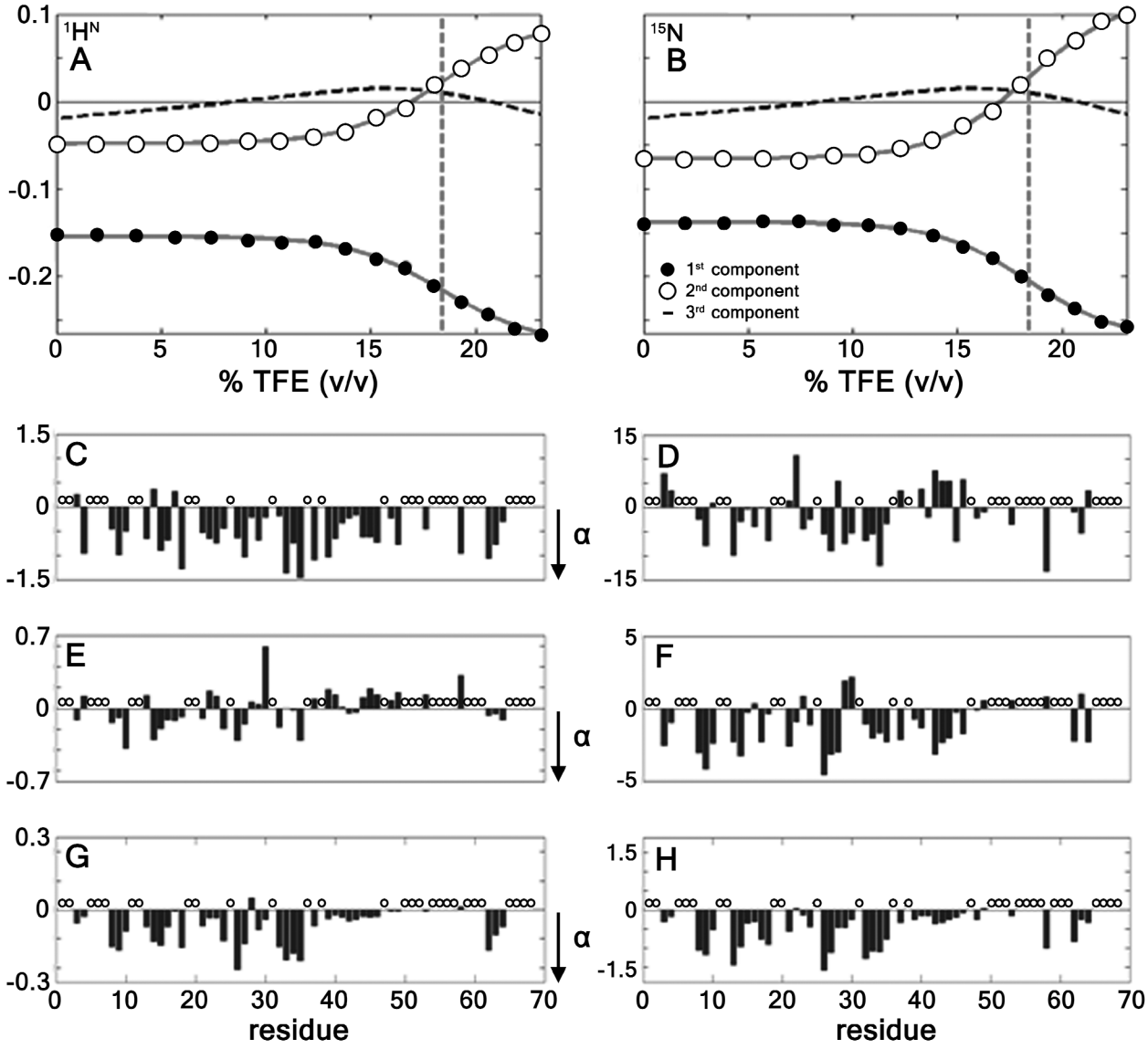


Figure 2-4: First three components of the TFE dependence (u_1, u_2, and u_3 from Equation 2-3) of the chemical shifts for $^{1}H^{N}$ and ^{15}N resonances are represented in panels A and B respectively. The first and second components of the TFE dependence are shown as open and filled circles respectively, while the third component is shown as a dashed curve. The solid lines show a two-state model, simultaneously fit to the first two TFE components of both the $^{1}H^{N}$ and ^{15}N together. The dashed vertical line indicates the TFE refolding midpoint which was found to be 18.3%, in agreement with the TFE refolding midpoint derived from CD (*9*). Panels C – H illustrate the residue dependence of the chemical shift (the *v* components from Equation 2-3). Panels C and D show the first component of the residue dependence shifts for $^{1}H^{N}$ and ^{15}N resonances associated with the first TFE dependent curve in panels A and B, panels E and F show the second component, and panels G and H show the third component. Open circles in panels C-H indicate residues not tracked. The arrow indicates helical tendency as given by Wishart and Sykes (*170*).

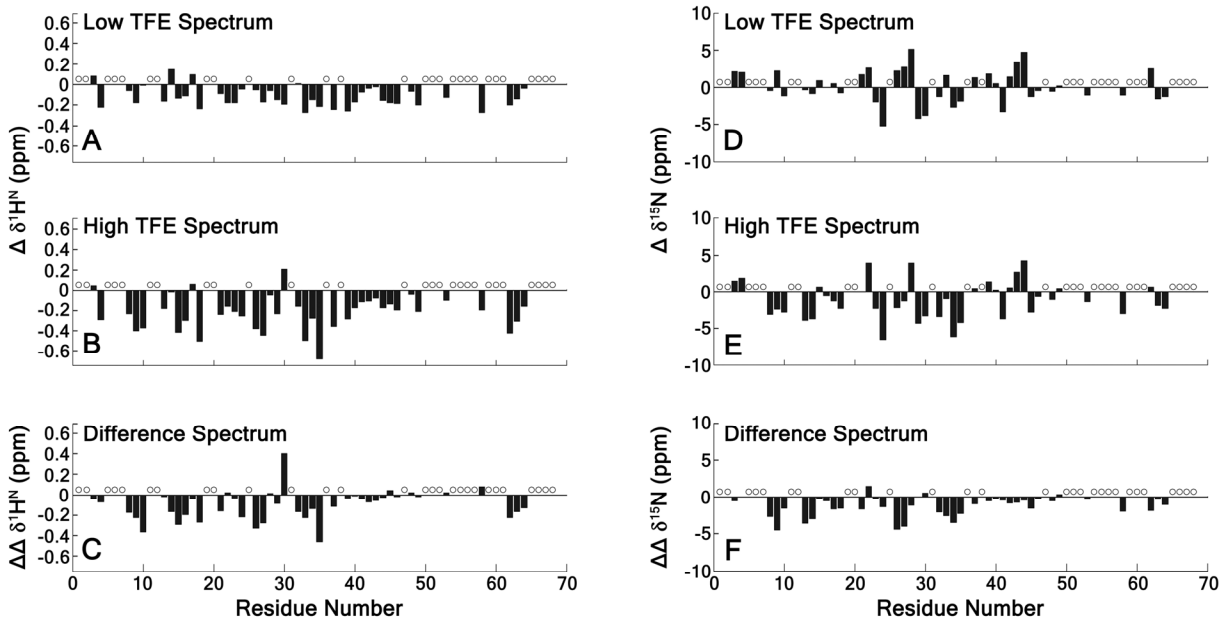


Figure 2-5: Computed high and low TFE spectra. Panels A and B indicate the computed residue dependence in the limit of low and high TFE concentrations for $^{1}H^{N}$; panel C shows the difference between the high and low TFE spectra. Panels D – F are the ^{15}N analogues to Panels A – C. Residues that were not tracked are indicated by open circles.

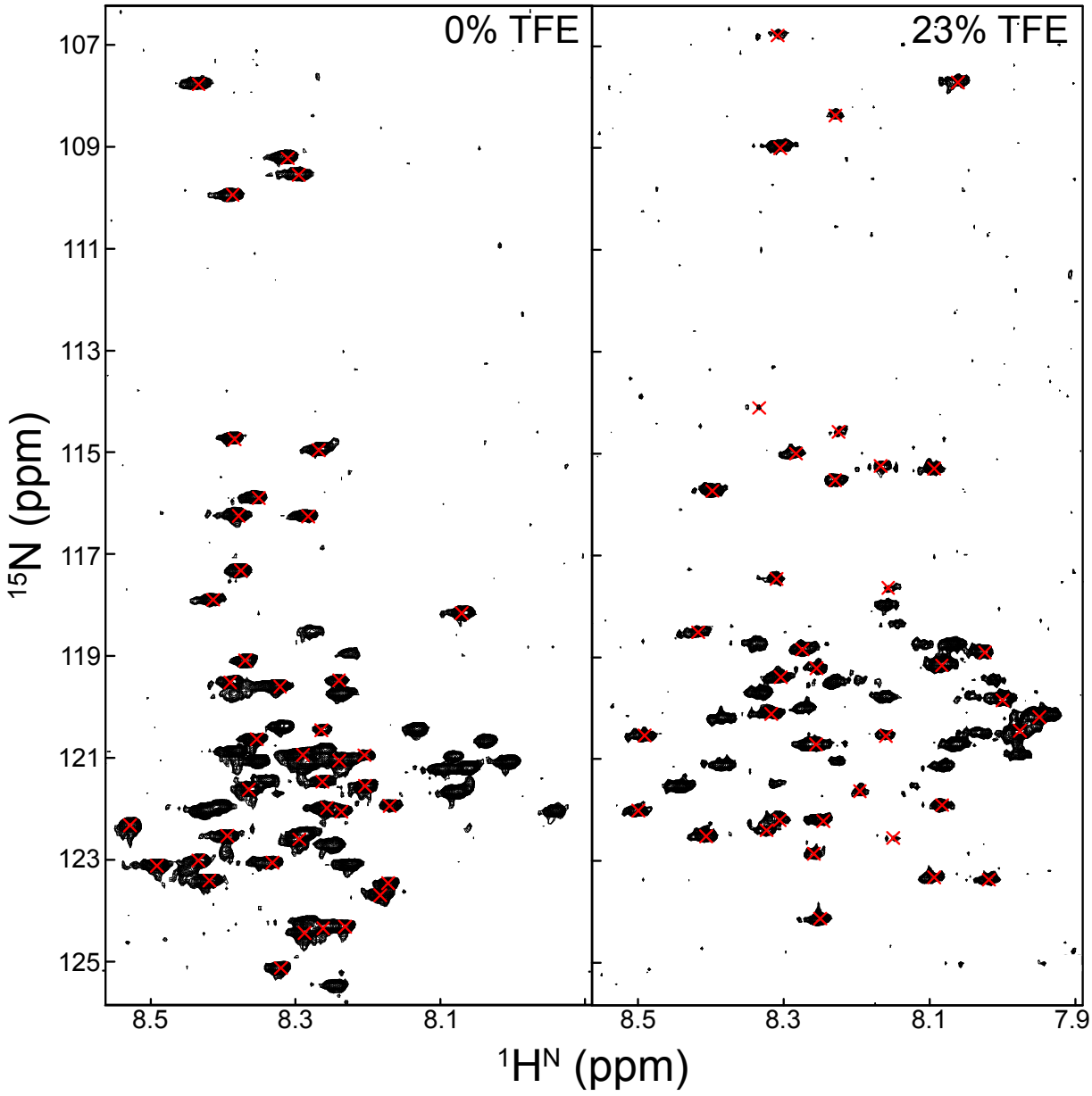


Figure 2-6: Experimental and reconstructed (red x) ^{15}N-HSQC spectra. Panel A shows the low TFE spectrum and Panel B shows the high TFE spectrum.

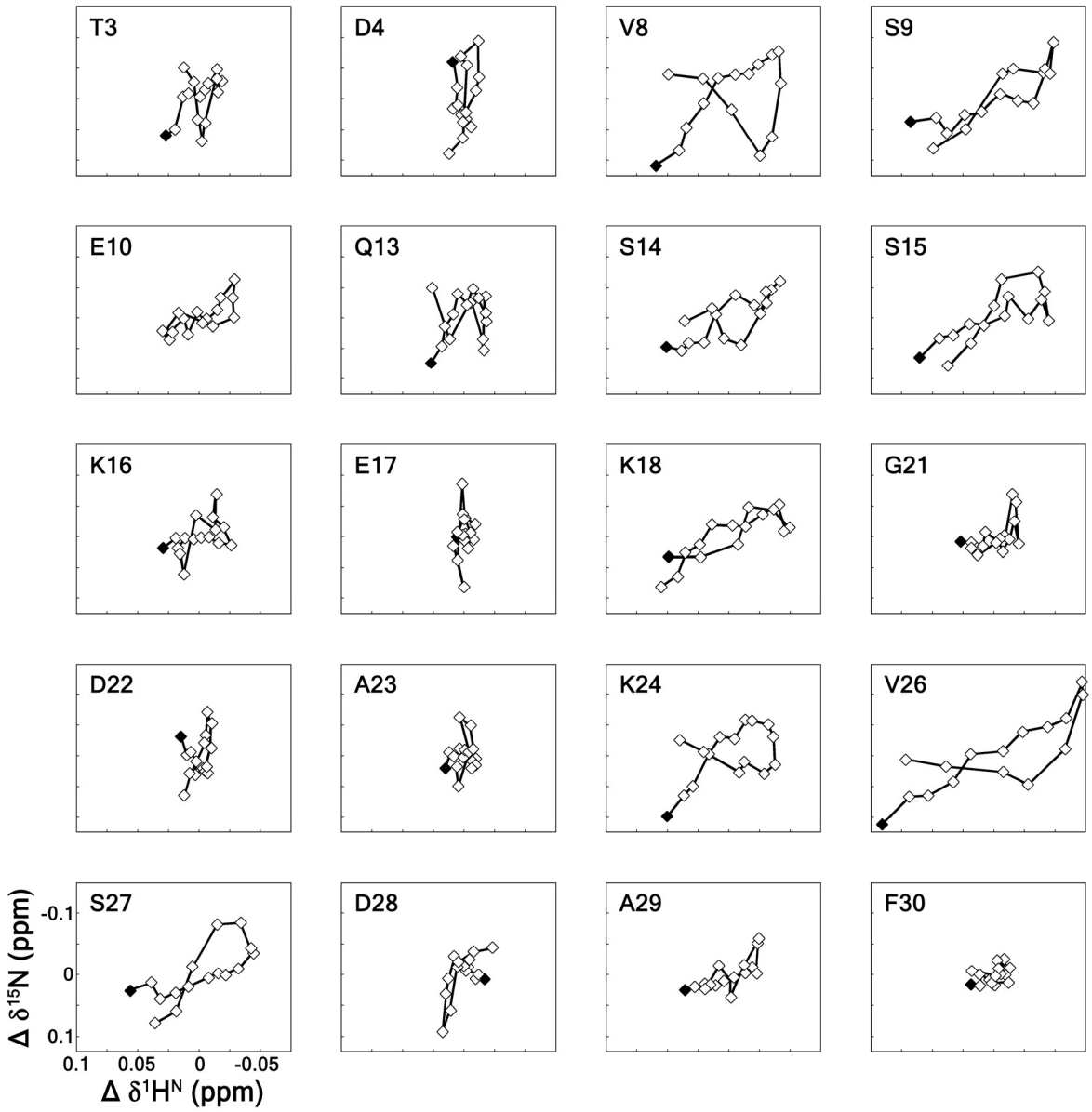


Figure 2-7: Analysis of the third SVD component. Each panel shows the reconstruction of the third component and residuals plotted for the residues tracked in this study. All panels are on the same scale. Each residue track illustrates the chemical shift movement of the third component from the SVD analysis for that residue as a function of TFE. Residues that do not have an appreciable third component appear to be randomly distributed around the origin with no discernable trend (e.g. Phe 30, Gly 40, Tyr 53). Residues with large third components show an outward looping track with beginning and ending points close to each other, and are very extended along the 1HN axis (e.g. Ser 9, Val 26, Met 33). The beginning of the TFE titration (0%) is indicated by a filled diamond.

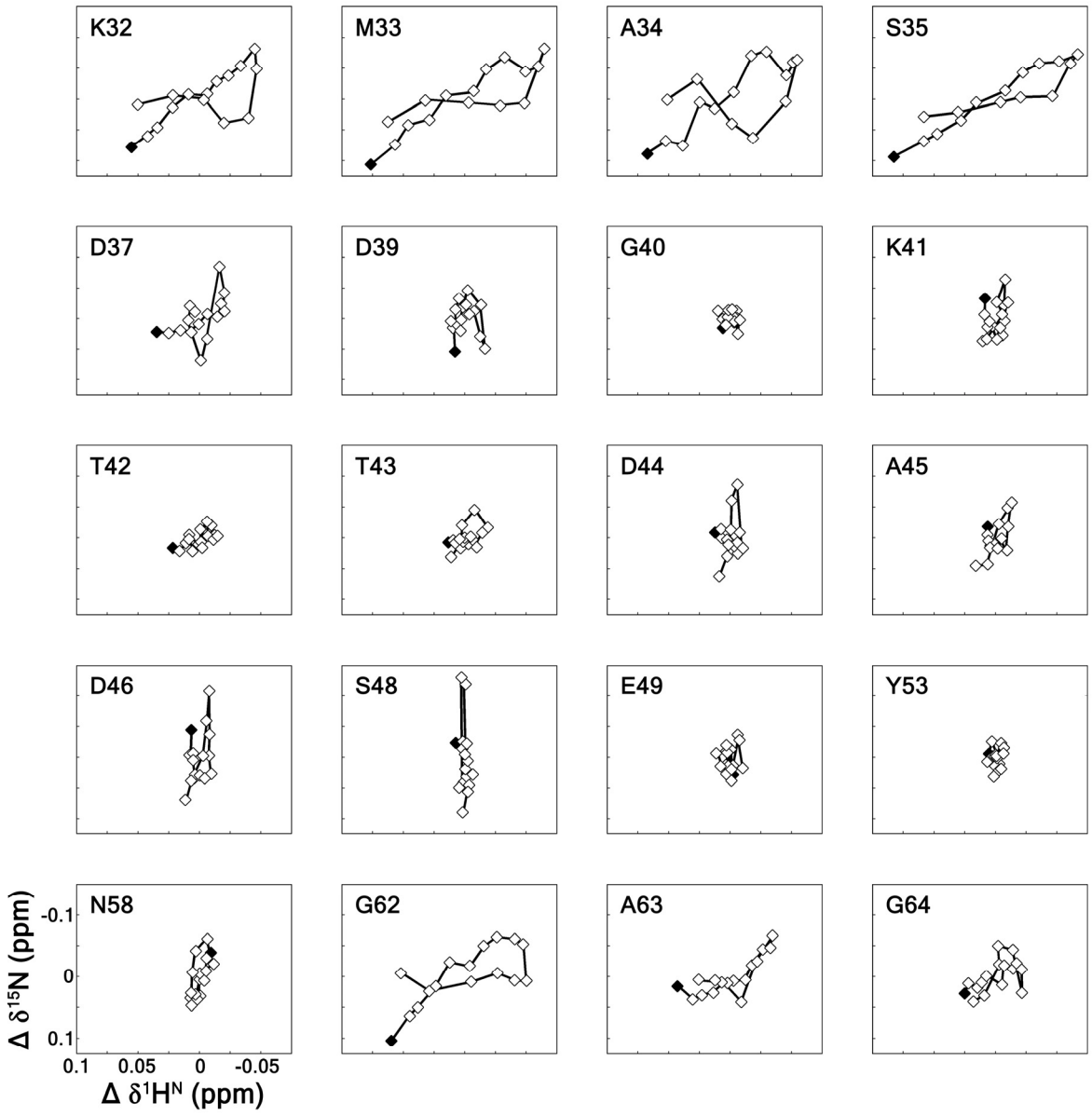


Figure 2-7 continued

CHAPTER 3
KINETICS OF IA_3 FOLDING

Introduction

A central notion of molecular biology is that a protein must have a defined structure to have a function (*185*). However, for many proteins this is not the case, as intrinsically disordered proteins (IDPs) defy this dogma (*186*). An intrinsically disordered protein can be found with very little well ordered structure in a biological context (*20, 103*). The action of an IDP occurs upon binding to target molecules (*40*). Crystal structures of IDPs bound to their partner molecules generally show structure of the IDP within the binding region of the complex (*7, 174*). There are several reasons why disorder may benefit a protein. For example, enhanced flexibility may allow an IDP to interact with more target proteins, or to assist in bringing an IDP and its target together (*7, 108, 110*). This raises some interesting questions: what is the method of folding and binding for an IDP, and are these processes coupled?

Chapters 1 and 2 explained that IA_3 is a small protein that is unstructured in solution. Crystal structures of the IA_3·YPrA complex indicate that biologically the N-terminus of IA_3 is found as a near perfect α-helix bound to YPrA (*10, 11*). We have demonstrated by CD and NMR that the addition of the alcohol cosolvent TFE folds IA_3 into a helical structure (*8, 9*). Using NMR we showed that while the whole protein transitions from unfolded to folded upon the addition of TFE, the N-terminus of IA_3 displays more of an α-helical character than the C-terminus of the protein. The controllable induced folding of IA_3 both in the presence of YPrA or TFE lends itself to investigation of the interactions between an intrinsically disordered protein and its target protease.

Kinetics of the folding and binding in IA_3 have not been studied prior to this work. One can imagine two simple models for the interactions between the inhibitor and the protease

(Figure 3-1). In one model the unstructured IA_3 molecule nonspecifically interacts with the surface of YPrA. This nonspecific interaction then aids in bringing the N-terminus of the inhibitor closer to active site, allowing a residue within the inhibitor an increased chance to find the target site within the protease and make a specific contact. In this scenario folding of IA_3 occurs upon binding to YPrA. A second model is that the N-terminus of IA_3 first folds into a helix, and this preformed helix then makes specific contacts and binds to YPrA. Folding of the helix in the second model occurs before binding of the inhibitor to the protease. The folding kinetics of IA_3 in these two models could be detectably different since in the first model folding is affected by contacts with YPrA. These contacts raise the local concentration of the N-terminus around the binding site and hold the position of the N-terminus near the binding site.

Measurement of the folding rate of IA_3 in water was obtained at increasing temperatures, and will be correlated with the folding rate of IA_3 in the presence of YPrA. To accomplish this, we used equilibrium far-UV CD measurements to verify structure and characterize the folding thermodynamics of several IA_3 mutants as a function of TFE concentration and temperature. A FRET probe was attached to IA_3 to allow us to detect the unfolding kinetically by fluorescence spectroscopy. Folding kinetics of IA_3 were then directly probed by temperature jump spectroscopy. This allowed us to measure the relaxation times for unfolding of structure within IA_3 as induced by TFE. Our ultimate goal is to constrain the folding and binding interaction models of IA_3 with YPrA by the comparison of IA_3 folding kinetics in the presence or absence of YPrA.

Methods

Choice of Donor and Acceptor Fluorophores

Fluorescence resonance energy transfer is a distance sensitive phenomenon (*187*). In FRET, a donor chromophore is excited, and transfers energy from this excited state to an

acceptor fluorophore through a nonradiative mechanism (*187, 188*). The mechanism is dipole-dipole coupling between the donor and acceptor chromophores (*187, 188*). In FRET, the efficiency of energy transfer between the two chromophores can be written in terms of distance between the donor and the acceptor, *r*.

$$E = \frac{1}{1 + \left(r / R_0 \right)^6} \tag{3-1}$$

In Equation 3-1, R_0 is called the Förster distance (or Förster radius) (*188, 189*). At $E = 1$ the efficiency of transfer is 100%, meaning that 100% of the donor energy is transferred to the acceptor and the donor's own emission is quenched. This occurs when $r = 0$, indicating that the donor and acceptor fluorophores are in near contact. As $r \rightarrow \infty$, the efficiency of transfer approaches 0% and the acceptor emission is diminished. The distance R_0 is characteristic for a donor acceptor pair. It is the distance at which there is a 50% probability of energy transfer for each donor excitation event (*188, 189*). The Förster dye pair used in this study is tryptophan (donor) and 5-((((2-iodoacetyl)amino)ethyl)amino)-naphthalene-1-sulphonic acid (1,5-IAEDANS or IAEDANS) (acceptor).

Tryptophan is a naturally occurring amino acid with a indole moiety. The absorbance of this moiety is very broad, extending from 265 nm to 290 nm (*75, 189*). The emission of tryptophan is equally broad, extending from 330 nm to 350 nm (*75, 189*). Since it is a naturally occurring amino acid, which absorbs low wavelength radiation, it is often used as a FRET donor.

1,5-IAEDANS is an organic fluorophore with an absorbance at 336nm (*189*). The emission wavelength is found above 500 nm(*189*). The free dye is soluble in DMF, however reactions between the iodo group at the tail of the molecule and thiols will result in attachment of the naphthalene rings to the sulfhydryl (through the IAEDANS scaffold) in a cysteine (Figure 5-

7). This reaction is covered in more detail in chapter 5. The absorption wavelength of 1,5-IAEDANS falls within the emission band of tryptophan (*189*). This overlap makes 1,5-IAEDANS useful as a FRET acceptor for a tryptophan donor.

The tryptophan (donor) / 1,5-IAEDANS (acceptor) Förster pair has a Förster radius of 18-21 Å (*189-191*). The placement of the donor and acceptor within the IA_3 molecule was chosen through an analysis of the 1DP5 crystal structure of the IA_3·YPrA complex (Figure 3-2). In this complex the N-terminus of a mutant IA_3 is found as a helix bound to the protease (*10*). The helix formed by this mutant IA_3 is nearly identical to the helix formed by WT IA_3 in the 1DPJ crystal structure, with the main difference (aside from the Lys to Met mutations at positions 31 and 32) being that the asparagine at residue 2 is visible in the 1DP5 structure (*10*).

Choice of Fluorescent Dye Position

Distance measurements between Asn2 and Lys16 of the IA_3 crystal structure yield distances of 20.3 Å and 19.4 Å for the α and β carbon separations respectively. The distance between these residues in the folded structure of IA_3 is essentially equal to the Förster radius for the tryptophan donor and IAEDANS acceptor. FRET labels at these residues could therefore be used as a probe of the unfolding of free IA_3. The protease, YPrA, contains 4 tryptophan residues. If a dansyl labeled K16C IA_3 is bound to the protease, three of the native Trp residues of YPrA are within the Förster radius for tryptophan and 1,5-IAEDANS: Trp39, Trp190, and Trp241. The fourth tryptophan residue at position 181 is over 28 Å away, which would lead to an efficiency of transfer of ~15%. Table 3-1 lists the tryptophan residues in YPrA and their distances from Lys16 in IA_3.

In the N2W-K16C·dansyl labeled IA_3 double mutant, distances between the tryptophan residue at position 2 in IA_3 and an IAEDANS labeled cysteine residue at position 16 of IA_3, and the distances from the IA_3 WT position 16 to the tryptophan in YPrA could be different from the

C_α distances due to the length of the side chains of these molecules. While this concern is worth noting here, the errors are relatively small and unlikely to affect the conclusions of this study.

Inhibition Constant (K_i) Determination

The attachment of the dansyl adds a large hydrophobic moiety to the inhibitor. It is possible that such an addition could alter the binding of the inhibitor to the protease. Binding of IA_3 to YPrA can be assayed by competitive inhibition assay. Samples of YPrA and IA3 in 0.1M NaAc at pH 4.5 were prepared as indicated in chapter 5. Inhibition constants were measured for all the IA_3 mutants as described previously (Table 3-2) (*192, 193*).

Florescence Resonance Energy Transfer Determination

We verified intramolecular FRET in labeled IA_3 by comparison of the equilibrium fluorescence spectrum of the N2W(his) single mutant and K16C·dansyl(his) labeled single mutant with the N2W-K16C·dansyl(his) labeled double mutant. Fluorescence spectra of each mutant at temperatures from 5 to 85°C in 5°C steps were taken. Samples in 17% TFE were excited with 266 nm wavelength radiation which would cause a broad but strong emission peak for tryptophan at 330-340 nm, and a weak emission peak for dansyl at 520 nm. If FRET occurred then the tryptophan peak would increase in intensity while the dansyl peak would be diminished in intensity during the unfolding process.

The integrated fluorescence intensity of each peak was taken by integrating the fluorescence spectra from 310 nm to 424 nm for tryptophan and from 425 nm to 550 nm for dansyl. These values were then plotted as a function of temperature (Figure 3-3).

Circular Dichroism Data Analysis

Far-UV CD spectra were collected for each IA_3 mutant in the presence of various concentrations of TFE, and at different temperatures as described in chapter 5. Assuming a two-state model of folding for IA_3 in TFE, we can say that the CD signal at each wavelength,

temperature, and TFE concentration is due to the fractions of molecules that are folded (f_f) or unfolded (f_u) (*8*) as in Equation 3-2.

$$CD_{\lambda,T,TFE} = a + b \cdot f_f + c \cdot f_u \tag{3-2}$$

In a two-state model, all the molecules are either in the folded or unfolded states.

$$f_f + f_u = 1 \tag{3-3}$$

This allows Equation 3-2 to be rewritten as an equation of one variable:

$$CD_{\lambda,T,TFE} = c_1 + c_2 \cdot f_u \tag{3-4}$$

for some constants c_1 and c_2.

The fraction of molecules that can be found in the unfolded state has a dependence on change in free energy of unfolding.

$$f_u = (1 + e^{\Delta G / RT})^{-1} \tag{3-5}$$

Since the concentration of TFE and temperature both affect the folded state of the protein, a simple model for ΔG can be written (*8*) as in Equation 3-6.

$$\Delta G(TFE) = \Delta G_0 + m \cdot [TFE] \tag{3-6}$$

That is, the free energy of unfolding varies linearly in TFE concentration and contains terms for the free energy in water (ΔG_0) and the helix stabilizing effect of TFE (the *m*-value). Recognizing that Equation 3-6 includes both a change in enthalpy and a change in entropy (by the Gibbs free energy equation), we can write the *m*-value as a combination of contributions from the enthalpy and entropy.

$$m = m_H - T \cdot m_S \tag{3-7}$$

Using Equation 3-7, Equation 3-6 can be rewritten as seen in Equation 3-8.

$$\Delta G(T, TFE) = \Delta H_0 + m_H \cdot [TFE] - T \cdot (\Delta S_0 + m_S \cdot [TFE]) \tag{3-8}$$

This suggests that we can fit the spectra derived from the large set of CD experiments at various temperature and TFE with only four parameters: ΔH_0, ΔS_0, m_H, and m_S. Uncertainties in these values can be estimated by bootstrapping the datasets.

Temperature Jump Spectroscopy

Kinetic data of the folding of the IA_3 protein were collected using laser induced temperature jump spectroscopy. In this experiment an IR laser is used to rapidly heat a sample, raising the temperature in 20-30 ns (*194*). The fluorescence of the sample is then probed by a UV laser as the protein relaxes to a new equilibrium at the raised temperature (*194, 195*). Our experimental setup allows us to capture not only the temporal relaxation of the system, but also the spectral distribution of the fluorescent emission over a 300 nm to 725 nm range.

Spectra are collected over a time range covering the ns to μs region, with ns resolution. The set of time resolved spectra are decomposed using singular value decomposition (SVD) to extract the dominant features from the data (*8*). A short description of SVD can be found in chapter 2. SVD of the dataset results in a set of spectral components, a set of time resolution components and a set of singular values which are used to weight the spectral and time resolution components.

Assuming that IA_3 is a two-state folding protein, the time resolution components from the SVD analysis can be fit to a single exponential function depending on a single parameter τ.

$$y = a + b \cdot e^{-t/\tau} \tag{3-9}$$

τ is characteristic of the relaxation of the system and is called the observed relaxation time. The inverse of this time constant is the observed rate k_{obs} which is the sum of the folding and unfolding rates.

$$\frac{1}{\tau} = k_{obs} = k_f + k_u \tag{3-10}$$

The ratio of the folding rate to the unfolding rate can be related to the change in unfolding free energy of the system by Equation 3-11.

$$\frac{k_f}{k_u} = e^{\Delta G/RT} \tag{3-11}$$

where ΔG is the change in free energy for a certain temperature and TFE concentration as deduced from equilibrium CD measurements. Combining Equations 3-10 and 3-11, we can arrive at a relationship between the folding rate, the observed relaxation rate and the change in free energy.

$$k_f = \frac{e^{\Delta G/RT} \cdot k_{obs}}{1 + e^{\Delta G/RT}} \tag{3-12}$$

Results

Inhibition of the IA_3 Mutants

We added a fluorescent IAEDANS acceptor on residue 16 and a tryptophan mutation at residue 2 of IA_3 in order to create a FRET probe of its folding/unfolding. This was implemented as an N2W and K16C double mutant with a dansyl label attached to the cysteine.

In the crystal structure of the IA_3·YPrA complex, the asparagine of IA_3 at position 2 fits into a pocket along the upper flap of YPrA (Figure 3-2) (*10, 11*). Mutating this residue to a larger tryptophan residue could alter the binding constant since the larger residue may not be accommodated as well into the pocket in the protease. The lysine of IA_3 at position 16 presents another concern since it is directly opposite the catalytic aspartates within the protease. It is possible that a lysine to cysteine mutation at this position could affect the inhibition constant by changing the positioning of the inhibitor directly above the active site of the protease. WT IA_3 exhibits a sub-nanomolar binding affinity towards YPrA. K_i measurements of the N2W(his) IA_3

single mutant were still within the nanomolar range (Table 3-2). Dansyl labeled K16C(his) IA_3 also had a subnanomolar K_i towards YPrA.

Circular Dichroism Analysis

To assess temperature dependent changes (unfolding) in IA_3, we collected CD spectra of each of the IA_3 mutants as a function of concentration of TFE and temperature in phosphate buffer. The CD spectra at 0% TFE for all mutants appeared unstructured at all temperatures (Figure 3-4). As the concentration of TFE increased, however, a second spectrum appeared which displayed an α-helical character. The structure formed in TFE could be melted with increasing temperature resulting in a spectrum that resembled the spectra at 0% TFE (Figure 3-4).

Globally fitting the dataset for each mutant (as in Figure 3-5) yields the representative folded and unfolded spectra. These spectra have characteristic lineshapes which resemble unfolded structures for the unfolded spectrum and α-helical structures for the folded spectrum (Figure 3-6).

The transition in TFE and temperature can also be mapped out from this global fit of the CD data (Figure 3-7). The transition midpoint is defined as the point at which ΔG = 0 kJ/mol, where 50% of the molecules are unfolded. Plotting the change in free energy as function of temperature and TFE concentration allows a view of how the ΔG is changing with each variable. In all cases the change in free energy has a very broad transition in both TFE concentration and temperature. The TFE midpoints of the transition at 25°C are listed in Table 3-3. These TFE midpoints and transitions are similar, and correspond to similar unfolding ΔG values.

Assessing Temperature Dependent Changes

Since we wanted to thermally perturb folded IA_3 and investigate the kinetics of the fluorescence relaxation, it was imperative to verify that there was an observable FRET

fluorescence change upon thermal unfolding. Measurements of IA_3 mutants containing either the tryptophan acceptor or IAEDANS donor fluorophore were compared to the double mutant containing both fluorophores in 17% TFE (Figure 3-3). This TFE concentration is at or near the midpoint of the folding transition at 25°C for all the IA_3 mutants. Tryptophan exhibits an intrinsic decrease in fluorescence as the temperature increases (*194*). This effect is quantifiable and amounts to a five fold decrease over the temperature range of the study. However, in the presence of the dansyl acceptor, as in the N2W-K16C·dansyl(his) mutant, the fluorescence of tryptophan decreases only two-fold (Figure 3-3). This diminishment in tryptophan intensity in the presence of the dansyl fluorophore suggests that there is an energy transfer between the donor and acceptor fluorophores.

The reciprocal experiment comparing dansyl fluorescence emission as a function of temperature indicates that fluorescence decreases as the temperature is increased. Direct investigation of the fluorescence of dansyl is confounded due to several factors, however this does suggest that there is a fluorescence intensity change in the tryptophan and dansyl emission spectra as a function of temperature.

Folding Rates

We calculated the folding rates of IA_3 at several TFE concentrations. The observed folding and unfolding rates were calculated from the observed relaxation rate and ΔG at a certain temperature using Equation 3-12. The folding and unfolding rates at 0% TFE were extrapolated from fitting the folding and unfolding rates at the various TFE concentrations (seen for 25°C in Figure 3-8) and are listed in Table 3-4.

Discussion

Intrinsically disordered proteins are found in a variety of biological contexts including signaling and transcriptional regulation (*7, 110*). These proteins can be found bound to their

target proteins with certain secondary structures, however when free in solution they do not appear to have any discernable secondary structure (*7, 40, 174*). Why would a protein, which can be found structured in a complex, exist with little structure by itself in solution? Is there a discernable advantage to the unstructured nature of an intrinsically unstructured protein?

Studies of the intrinsically unstructured IA_3 were used to shed light on these questions. The N-terminus of IA_3 can be induced to fold into a helical structure in the presence of the alcohol cosolvent TFE, or in the presence of its target aspartic protease, YPrA. If folding and binding of the inhibitor to the protease are coupled processes, then the unstructured nature of the inhibitor may assist in raising the rate at which folding occurs as compared to independent folding and binding processes.

Investigation of the Equilibrium Structure

CD studies of WT(his), N2W(his), K16C·dansyl(his), and N2W-K16C·dansyl(his) IA_3 indicate that these IA_3 mutants largely display an unstructured character at low TFE concentrations (Figure 3-4). CD spectra of all the mutants transition towards an alpha-helical spectrum with the addition of TFE. In all mutants with TFE-induced structure, the spectrum at higher temperatures resembles the unstructured spectra collected at 0% TFE. These data can be fit to a global two-state model over all temperature and TFE concentrations, strongly indicating that to a good degree the folding and unfolding of IA_3 by TFE and temperature can be described as a two-state transition: an unfolded state converting to a folded state with strong α-helical character (Figures 3-5 and 3-6).

Equilibrium CD of the IA_3 mutants at increasing temperatures and TFE concentrations allowed us to construct a linear model for the unfolding free energy, ΔG, as a function of these two variables. The results of this model indicate that the folding/unfolding transition midpoint in

both TFE and temperature show some variation among the IA_3 mutants (Figure 3-7). While this transition is similar in all of the mutants, the midpoint is slightly shifted towards higher TFE concentrations for the N2W(his) mutant, and even more so for the N2W-K16C·dansyl(his) mutant. Since this midpoint shift did not occur in the K16C·dansyl(his), it would appear that the N2W mutation subtly affects the TFE dependence of the transition. We note that the crystal structures containing the IA_3 helix start at residue 2 and do not contain the initial methionine; it is possible that helix initiation may begin at residue 2. If this is so, the tryptophan mutation may slightly supress helix initiation leading to a larger concentration of TFE being needed to initiate helix formation.

Observing Fluorescence

It is not possible to undertake kinetic experiments on IA_3 folding without a probe that is sensitive to changes in peptide secondary structure. Molecules displaying fluorescence and FRET can be used as probes since the fluorescence emission provides a readily identifiable signal, and reliance on distance can be exploited by observing changes in the intensity as a function of time needed for a protein to relax toward an equilibrium (*188, 189*).

Observations of the N2W-K16C·dansyl(his) IA_3 mutant indicate that at low temperatures the dansyl fluorescence is high (Figure 3-3). As the temperature is increased, the dansyl fluorescence emission declines. The tryptophan fluorescence for the labeled double mutant displays a decrease with rising temperature. If fluorescence tryptophan emission decreases with rising temperature, how do we know if there is a FRET signal?

A solution of free tryptophan has a characteristic decrease of 2% per degree Celsius in fluorescence emission as the temperature is increased (*194*). This decrease in emission would amount to a four fold decrease over the 5°C-85°C range. At low temperatures the tryptophan fluorescence of the labeled double mutant decreases by half as much as expected for an indole on

a polypeptide. This increase could be indicative of FRET. At higher temperatures this effect decreases, which is consistent with thermal denaturation of the peptide.

By CD we have seen that at lower temperatures IA_3 appears to have a helical structure, and could be found as a folded molecule (*8, 9*). At these lower temperatures the dansyl fluorophore displays a very intense fluorescence spectrum. At higher temperatures this spectrum is vastly diminished. This represents a quantifiable change in the fluorescence of the N2W-K16C·dansyl(his) IA_3 mutant, and is consistent with resonance energy transfer occurring within the folded inhibitor.

Folding Kinetics of IA_3

Using temperature jump spectroscopy, we measured the observed folding relaxation of the N-terminus of IA_3. By combining this observed folding rate with the unfolding free energy deduced from the CD study, we can calculate the unfolding rate constants (or folding rate constants) as a function of TFE at each temperature studied. By extrapolating these values to 0% TFE, we can obtain the folding rate in water for the inhibitor at a particular temperature (Figure 3-8).

The extrapolated folding and unfolding rates in water are listed in Table 3-4. Examination of these rates shows that as the temperature of the system is increased, the unfolding rate dramatically increases. The unfolding rates increase over two orders of magnitude, from $0.41\pm1.71\ 10^6$/s at 5°C to $25\pm15\ 10^6$/s at 50°C. This is well in line with our understanding of the system being folded at low temperatures and unfolding at higher temperatures.

Comparisons of the unfolding/folding rates in the presence of YPrA are needed to constrain the folding/binding model of the inhibitor to the protease. No change in the folding rate of IA_3 in the presence of the YPrA would indicate that the formation of a helix is not enhanced by binding, and suggest that helix formation occurs before binding. An increase or

decrease in the folding rate would suggest an enhancement in the presence of the protease, indicating that folding and binding are coupled processes. Early results of IA_3 in the presence of YPrA show a folding rate of ~10^7/sec (or ~100ns) indicating that IA_3 does interact with YPrA, and this interaction leads to a distinct folding rate enhancement.

Conclusions

The folding rate of IA_3 is found to be more sensitive to temperature than to TFE concentration. The TFE dependence in the observation relaxation rates is exceptionally weak, thus we can deduce with confidence the folding and unfolding rates of free IA_3 in water at various temperatures. These data are valuable as a step towards a comparison of folding rates in the presence of the protease since they allow us to interpret the folding rate of IA_3 in the presence of YPrA. Comparisons of the folding rates of IA_3 in water and with the protease will allow us to constrain the folding and binding interaction models of IA_3. Investigations of the kinetics of IA_3 folding in the presence of the protease are currently under way.

Table 3-1: Distances from tryptophan residues in YPrA to Lys16 in IA_3. These distances were found using the 1DPJ crystal structure.

YPrA residue	Distance to K16 in IA_3 (Å)
Trp39	19.0
Trp181	28.2
Trp190	18.4
Trp241	17.9

Table 3-2: Binding constant values of IA_3 to YPrA, for the mutants used in this study. Inhibition values for all the mutants were at or beyond the limits of instrument detection at 25°C, pH 4.5, indicating that they all lie in the subnanomolar range.

IA3 mutant	Inhibition constant (nM)
WT(his)	<0.1
N2W(his)	<0.1
K16C·dansyl(his)	<0.1

Table 3-3: The TFE folding midpoint values at 25°C for the IA_3 mutants in this study. These values are extracted from globally fitting the CD data for each mutant and extracting the TFE value corresponding to $\Delta G = 0$.

IA_3 mutant	TFE Folding Midpoint ($\Delta G = 0$) at 25°C
WT(his)	17.8±0.3%
N2W(his)	21.4±0.2%
K16C·dansyl(his)	17.0±0.3%
N2W-K16C·dansyl(his)	23.8±0.3%

Table 3-4: Folding and unfolding rates for the IA_3 helix, as measured in TFE and extrapolated to water (0% TFE). As the temperature increases there is a large increase in the unfolding rate of the protein.

T (°C)	K_f (10^6/s)	K_u (10^6/s)
5	0.02±0.10	0.41±1.71
10	0.04±0.09	0.68±1.59
15	0.07±0.10	1.14±1.71
20	0.11±0.07	1.86±1.15
25	0.18±0.04	2.99±0.64
30	0.27±0.03	4.56±0.51
35	0.41±0.09	6.71±1.45
40	0.63±0.21	10.33±3.49
45	0.99±0.40	16.22±6.58
50	1.55±0.89	25.40±14.57

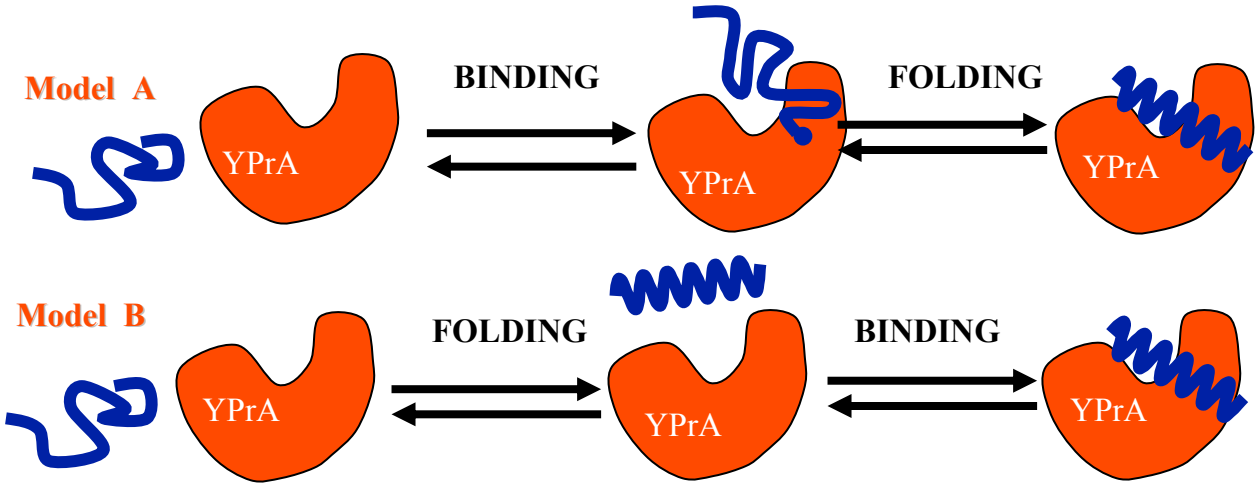


Figure 3-1: Simple folding and binding models for the interaction of IA_3 (in blue) with YPrA (in orange). In model A folding takes place after binding. In model B IA_3 folds into a helix, then this preformed helix binds to the protease. Figure reproduced with permission from Ranjani Narayanan.

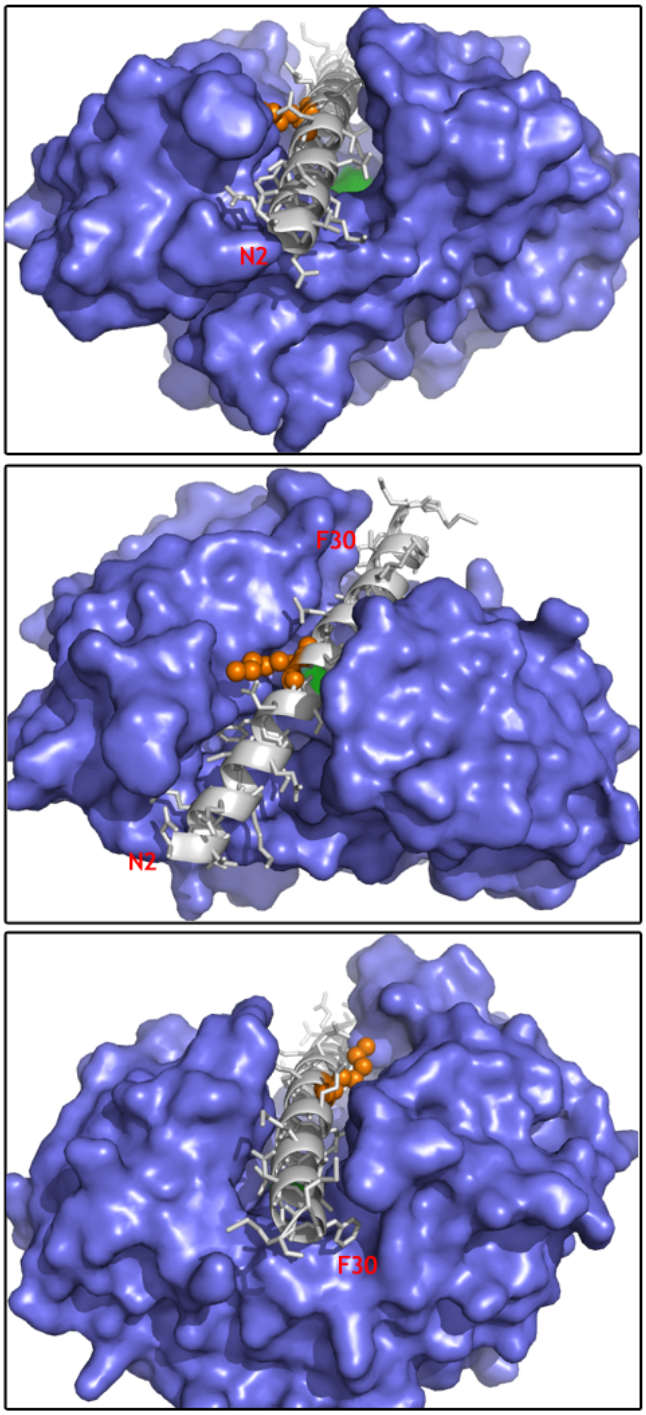


Figure 3-2: Several views of the IA_3·YPrA complex illustrating the placement of IA_3 residues within 'pockets' in YPrA. YPrA in blue is surrounding the helical N-terminus of IA3 in grey. The catalytic aspartates of YPrA are shown in green. Lys16 in IA3 is shown in orange space filling form. Residues Asn2 and Phe30 in the IA_3 helix are listed. These views of IA_3 and YPrA were generated using PyMOL.

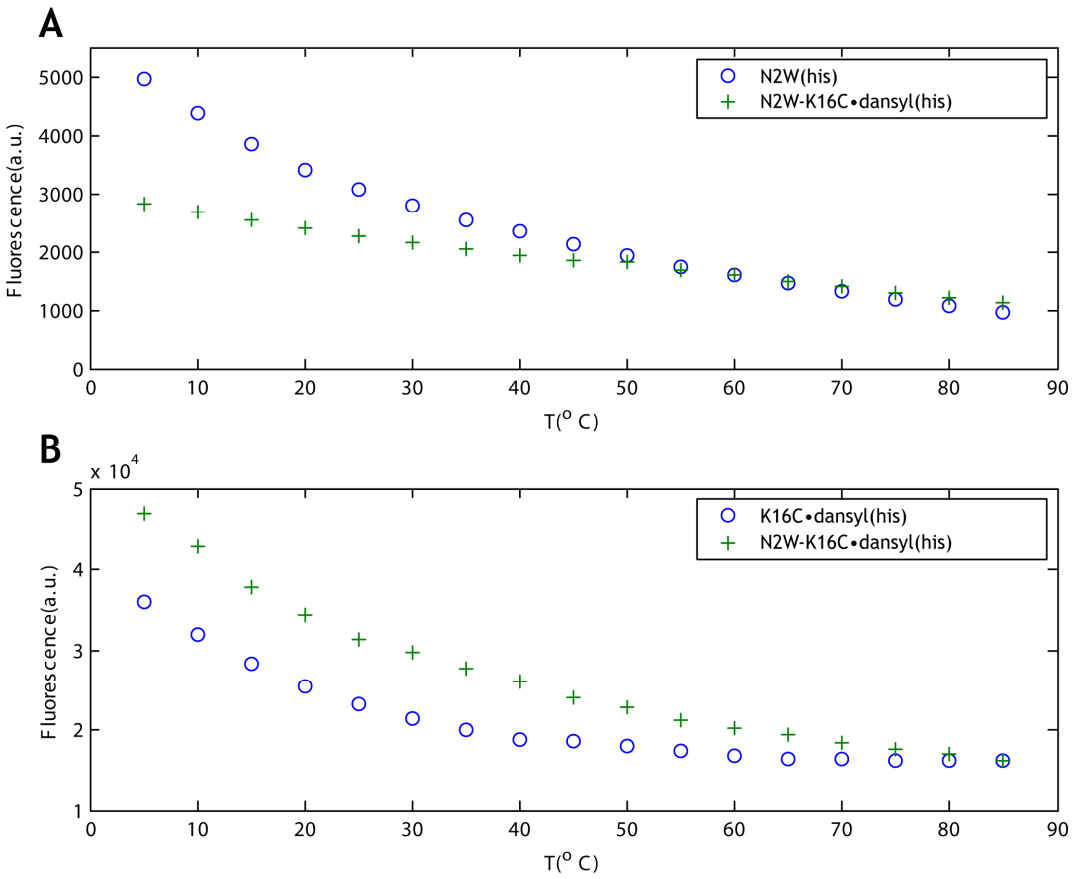


Figure 3-3: Fluorescence in the presence or absence of the donor or acceptor excited at 266 nm. A) Tryptophan donor fluorescence in the presence or absence of dansyl in N2W-K16C·dansyl(his). At lower temperatures the tryptophan fluorescence is higher alone than in the presence of dansyl. B) Dansyl acceptor fluorescence. In the presence of tryptophan the dansyl fluorescence is increased at lower temperatures.

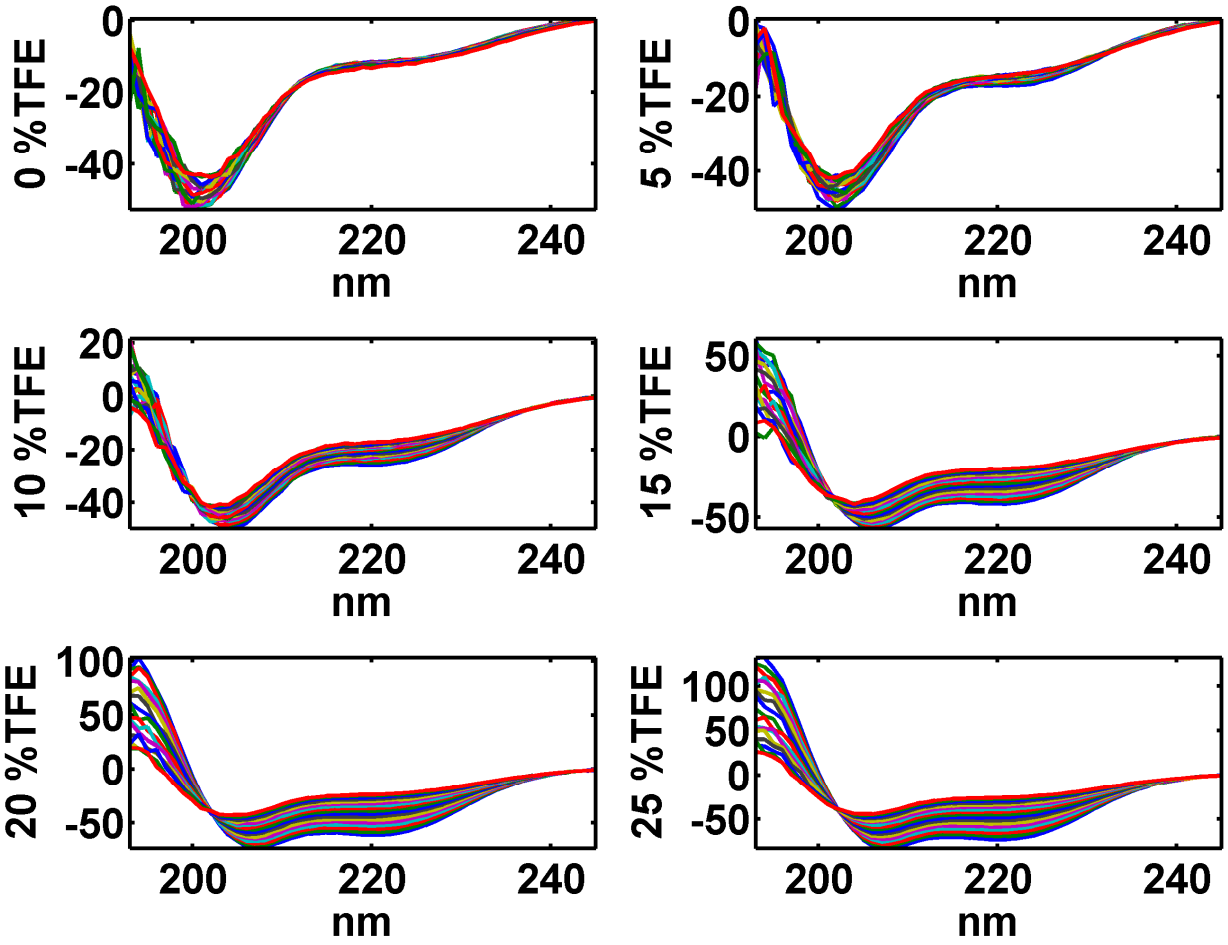


Figure 3-4: CD spectra gathered as a function of TFE and temperature. Shown here are the CD spectra for WT(his) IA_3, which is typical of the data gathered for all the mutants in this study. Note that at 0% TFE the line shapes of CD spectra at all temperatures are similar. As TFE is added the CD spectra change indicating α-helical structure. At low temperatures this α-helical structure is visible in the CD spectra. With increased temperatures these structures melt, resulting in an unfolded CD spectrum.

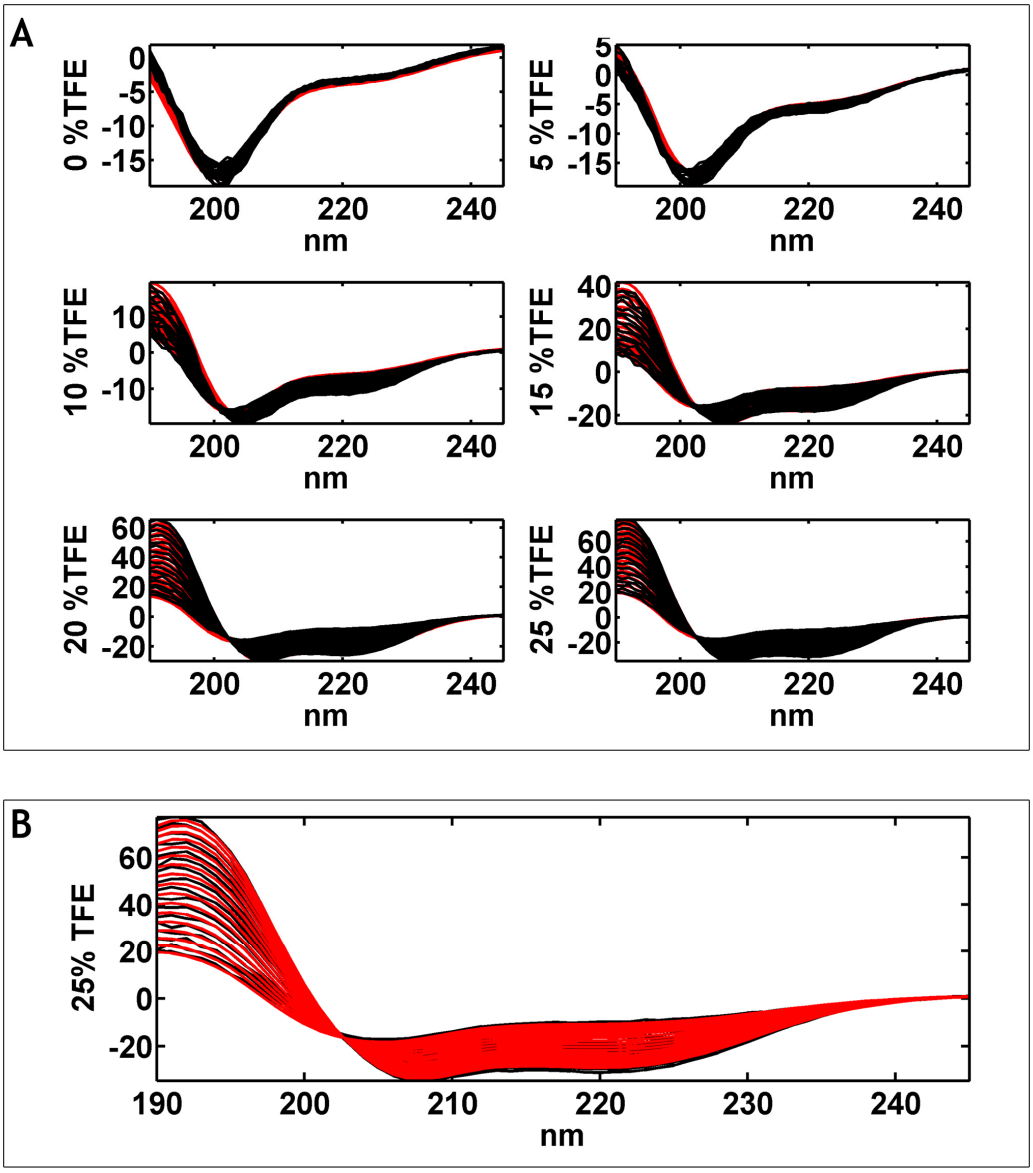


Figure 3-5: Typical fit to the far-UV CD spectra using the global two-state fit. A) The fit (in red) performed simultaneously on all CD spectra (in black) at all temperatures and TFE concentrations for K16C·dansyl(his) IA_3. B) Enlargement of the CD spectra and fit at 25% TFE.

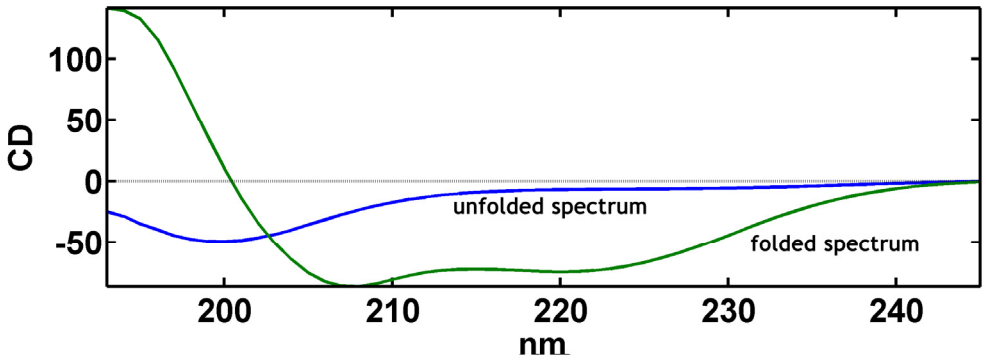


Figure 3-6: Folded and unfolded CD spectra of IA_3 as obtained from a global two-state fit of the spectra in Figure 3-4. The unfolded and folded spectra are characteristic of unfolded and α-helical spectra, respectively.

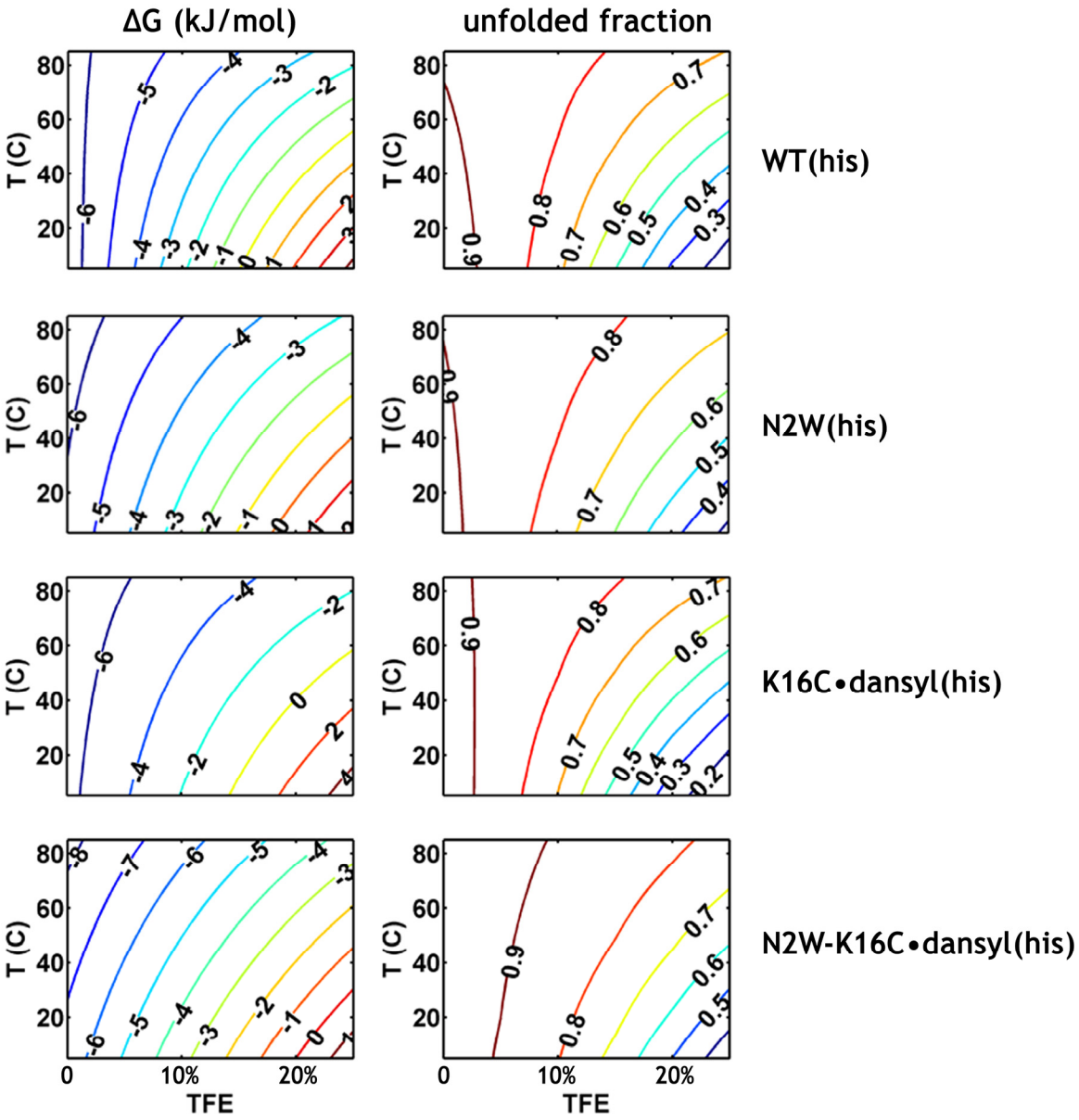


Figure 3-7: Contour plots of the free energy of unfolding (left column) and fraction of molecules that are unfolded (right column) as a function of temperature and percentage of TFE for all IA_3 proteins in this study, as obtained from the two-state global fit for each protein. All the contour plots display the same behavior: a broad transition in temperature and TFE concentration.

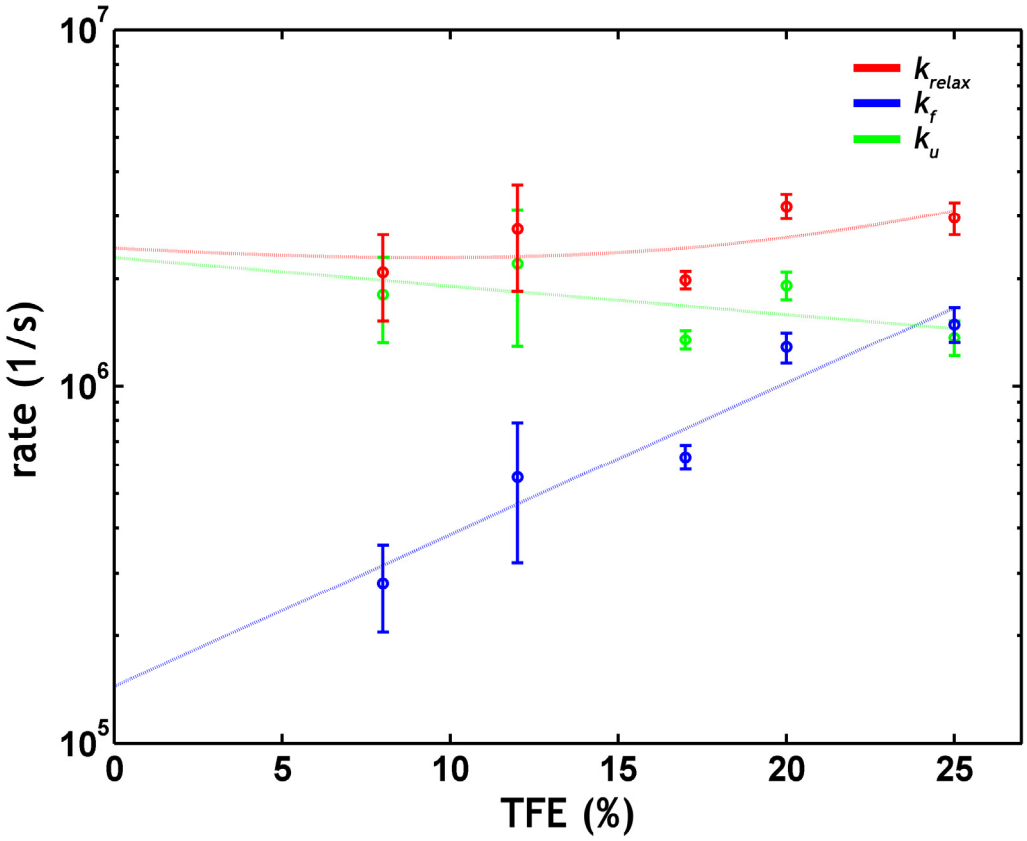


Figure 3-8: Plot of the observed relaxation rate, folding rate and unfolding rate as a function of TFE for N2W-K16C·dansyl(his) IA_3 at 25°C. Circles are data points, lines are fits to the data points allowing extrapolation to 0% TFE. This yields the relaxation rate, folding rate and unfolding of the mutant in water at 25°C.

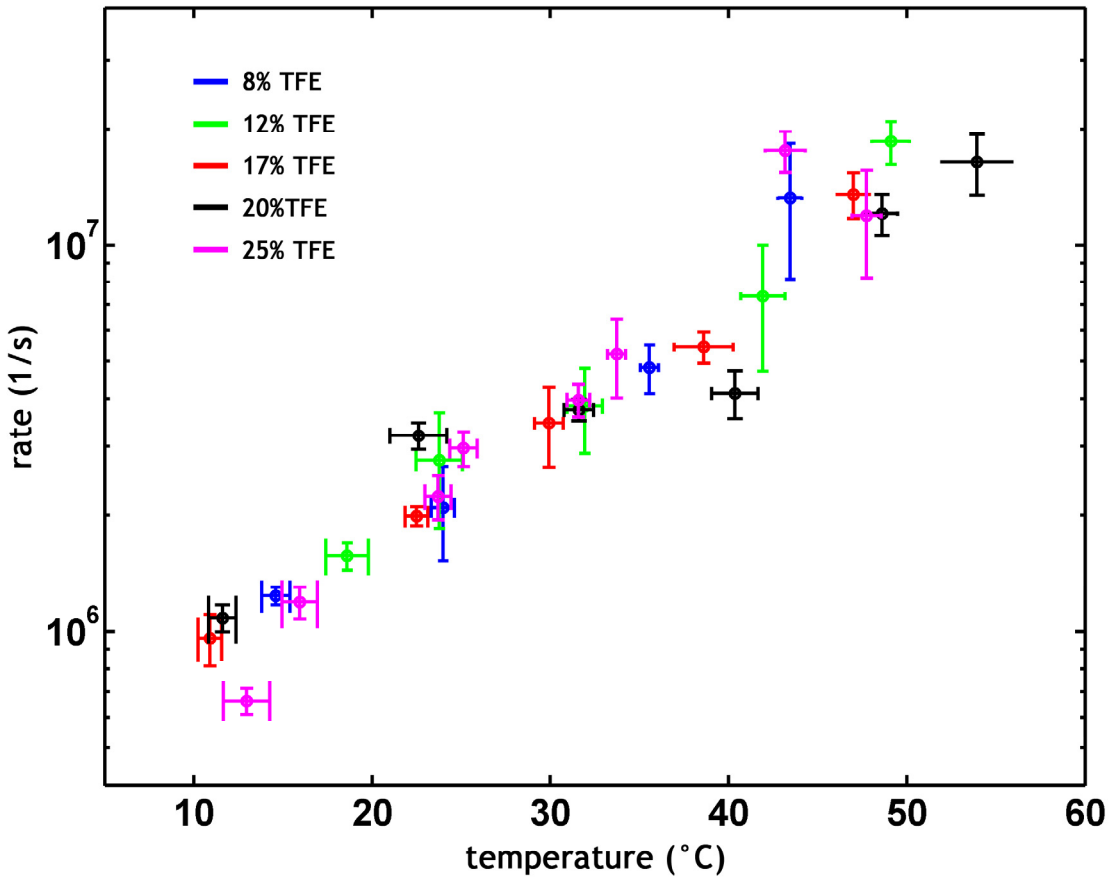


Figure 3-9: Plot of the relaxation rates at various TFE concentrations as a function of temperature. The folding rate of IA_3 appears to be more sensitive to temperature than to changes in TFE concentration.

CHAPTER 4
ISOTHERMAL TITRATION CALORIMETRY OF IA_3 PROTEINS IN THE PRESENCE OF YPRA

Isothermal Titration Calorimetry of the YPrA·IA_3 System

Introduction

All biological processes at the chemical level (and indeed chemical reactions generally) involve molecular recognition between the elements involved. Within living systems these elements can be very diverse, ranging from proteins to DNA to other small and large molecules. Additionally, proteins can be in different conformational states, which could allow for differences in molecular recognition. To understand any chemical or biological process, it is important to understand the nature of the recognition and interactions between elements at the molecular level, and elucidate the forces that stabilize the interactions (*196-199*).

The intrinsically unstructured protein, IA_3, can inhibit yeast proteinase A, YPrA. To inhibit YPrA, IA_3 undergoes an unstructured to structured transition; however, this transition does not manifest throughout the whole protein, but rather crystal structures of the YPrA·IA_3 complex show that the N-terminus of the protein adopts a helical structure (*10, 11*). The IA_3 structure found bound in the crystal structure of the inhibited YPrA· IA_3 complex are the N-terminal residues 2-32 found in a helix. These residues are less than half of the full length protein. The x-ray structure implies that the C-terminus of IA_3 is unstructured, and hence has no known function. Given that half of the IA_3 protein is not needed for inhibition, one wonders what benefit the C-terminal end may offer the whole protein. A possible effect of the C-terminus may be to increase the concentration of IA_3 near the active site of YPrA through weak interactions with the surface of YPrA. This explanation is similar to the study of IA_3 using an N-terminal his tag by Winterburn *et al.* (*135*). Here I present evidence for the interaction of the C-terminus with YPrA, and a possible explanation for this interaction.

Choice of IA_3 Proteins

To investigate the C-terminal interaction of IA_3 towards YPrA it was necessary to obtain three IA_3 variants:

- Wild type IA_3 – WT
- Wild type with a his_6 tag addition – WT(his)
- The N-terminal portion of IA_3, which is found as a helix in the crystal structure of the YPrA·IA_3 complex (PDB accession number 1DPJ) – Nterm

These variations were chosen based on several attributes (Figure 4-1 and Table 4-1). The crystal structure of IA_3 shows that residues 2-31 interact with the upper flaps of YPrA by binding to several residues as an α-helical structure (*11*). Although the full length inhibitor is 68 amino acids long, the C-terminal residues are not seen in the crystal structure which could indicate that this region is very mobile (*10, 11, 108*). Circular dichroism and NMR studies of the full length protein indicate that it is unstructured in the absence of YPrA (*9*). Several authors have shown that only the N-terminal half of the protein is needed for inhibition, and has potency similar to that of the wild type inhibitor (*10, 135*). Since the C-terminal portion of the protein is not seen in the crystal structure, it is possible that it is not necessary for inhibition of the protease.

In 2000, Shoemaker *et al.* proposed the fly-casting mechanism which postulated that natively unfolded proteins may have an enhanced ability to find a binding partner as compared to structured proteins (*108*). The authors hypothesized that this is due chiefly to two main effects:

- A well structured protein is restricted conformationally, and cannot explore the immediate solution environment for a binding partner as quickly as an unstructured protein.
- An unstructured protein may not be able to interact as strongly with its binding partner as a structured protein, however through weak nonspecific interactions the two proteins can draw together – enhancing existing interactions which may prompt the unstructured protein to fold.

This suggests that the C-terminus of IA_3 may serve to enhance the ability of the protein to find, bind, and fold onto YPrA. I used isothermal titration calorimetry to investigate the interactions of YPrA with wild type protein and the truncated N-terminal IA_3.

The additional WT IA_3 with C-terminal his_6 tag addition variant (WT(his))was tested as well. The WT(his) variant was made to ease in purification of IA_3 by making use of the affinity of the his_6 tag for nickel (*10*). However, it also served to check the relationship between charge and IA_3 binding to YPrA. Within the cell YPrA is located in an acidic vacuole (*132*). In such an environment, YPrA has a negative surface charge (calculated to be −7.6 at pH 4.5) (Figure 4-2) (*132, 200, 201*). Wild type and the truncated N-terminal IA_3 have surface charges of +5.4 and +2.3, respectively at pH 4.5. The addition of a his_6 tag to the wild type IA_3 significantly enhances the positive charge through the titratable histidine groups, with a resulting charge of +10.6 at pH 4.5. This his_6 tag addition to the wild type IA_3 then serves to verify if charge is a driving force in the C-terminus interaction of IA_3 with YPrA.

Methods

Protein Preparation

YPrA and IA_3 were prepared as described in chapter 6. Concentrations of stock solutions were determined using amino acid analysis at the University of Florida Interdisciplinary Center for Biotechnology Research Protein Chemistry Core.

Isothermal Titration Calorimetry Experiments

ITC experiments were performed on a VP-ITC titration microcalorimeter from Microcal, LLC. (Northhampton, MA). A typical experiment consisted of injections of IA_3 into YPrA. Samples of YPrA in 0.1M sodium acetate, pH 4.5, at 1μM were put in the sample cell. An IA_3 sample (in 0.1M sodium acetate, pH 4.5) was loaded into the titration syringe at a concentration of 13 μM.

The experiment consisted of 60 injections of IA_3. Each 10 μL injection lasted 20 seconds followed by an equilibration time of 240 seconds before the next injection. The equilibration time was sufficient for the system to return to baseline before the next titration. An initial injection of 1 μL, was deleted from the analysis as previously described (*9*). The sample cell was stirred at 460 rpm to mix the solutions during the experiment.

Control experiments, to determine the heat of dilution, were carried out by titrating IA_3 into the sample cell containing only buffer and by averaging the heat obtained per injection after saturation of YPrA by IA_3. The heats of dilution for control injections were averaged and subtracted from each injection of the binding experiment. Further explanation of the ITC experiment and methods can be found in chapter 5.

Isothermal Titration Calorimetry Analysis

The data were analyzed using Origin (OriginLab, Northampton, MA). The binding isotherm was deconvoluted by well-established methods (*72, 73, 202*), and a Levenberg-Marquardt nonlinear least-squares fitting method was then used to generate a best fit of the data and determine ΔH (enthalpy change of the binding reaction), K_a (association or binding constant), and n (stoichiometric ratio of inhibitor to protease binding sites). The binding affinity of IA_3 for YPrA is very strong, which causes the titration curve to be very steep and similar to a step function. K_a values determined from these ITC data are not accurate and were not used in our analysis (*9, 203*). A more detailed explanation of ITC analysis can be found in chapter 5.

Heat Capacity (ΔC_p) Determination

The relationship between the change in heat capacity upon binding (ΔC_p) (at constant pressure) and the change in binding enthalpy can be written as in Equation 4-1 (*73*).

$$\Delta C_p = \frac{d(\Delta H)}{dT} \tag{4-1}$$

Plotting the change in heat capacity of binding versus the temperature can provide insight into the equilibrium binding of the YPrA· IA_3 complex. The change in heat capacity was determined by fitting the ΔH values at each temperature to Equation 4-1. The fit was optimized using the error in ΔH provided by the ITC analysis. The optimization consisted of using the error in a particular ΔH value as a weighting of that value. Smaller errors in ΔH resulted in a higher weighting for that ΔH value.

Circular Dichroism Experiments

CD experiments of the WT, WT(his) and N-terminal IA_3 proteins were done as described in chapters 3 and 5.

Results

Optimizing the ITC Curves

The titration binding curve of WT and WT(his) IA_3 binding to YPrA appears sharp, and very similar to a step function; additionally the temperature difference as seen by the microcalorimeter after each injection of titrant into titrand is not large (Figures 4-3 and 4-4). Raising the IA_3 concentration would increase the temperature difference in each injection; however, doing so would sharpen the binding curve closer to that of a step function. Lowering the IA_3 concentration would broaden out the curve making fitting of the data more robust, however this increases the error in each data point since the signal to noise ratio would be much lower.

Finding optimal concentrations for the experiments was crucial to the success of the ITC measurements. Through several titrations at differing YPrA and IA_3 concentrations it was determined that a YPrA concentration of 1 μM and an IA_3 concentration of 13 μM would be sufficient to determine the values of ΔH and ΔC_p.

Figure 4-3 shows the raw data from the interactions of WT IA_3 and YPrA, measured by the ITC microcalorimeter. The binding reactions seen in these data are exothermic (*72, 73*). Integration of the heats of reaction of the experiment results in a binding curve (Figure 4-3). The distance between the beginning baseline of the curve and ending saturation is the enthalpy of binding, while the slope of the curve is proportional to stoichiometry multiplied by the binding constant. The titration data are fit using a least-squares fitting method to obtain the stoichiometry of binding for the ligand to the protease, change in enthalpy, and equilibrium binding constant. A similar analysis was done for WT(his) (Figure 4-4) and N-terminal (Figure 4-5) proteins at temperatures from 15°C to 45°C in 5C° increments.

A few of the titrations had a very low signal to noise ratio even at the optimal concentrations of the IA_3 and YPrA. The higher temperatures in all datasets appeared to be noisier. This was attributed to air bubbles forming in the sample cell during the experiment (*73*) despite degassing all samples before each experiment. Additionally, titrations of the N-terminal protein appeared to yield a lower signal than the other proteins.

Investigating the Heat Capacity

The ΔH values derived from fitting the ITC data were fit using Equation 4-1, as previously described, and are plotted in Figure 4-6. Linear fits of the ΔH values provided ΔC_p values which are listed in Table 4-2. All IA_3 proteins had differing ΔC_p values, however the plots of ΔH versus T for WT and WT(his) data seemed similar to each other when compared to the N-terminal data. The ΔC_p value for the N-terminal protein was found to be 387±32.3 cal/molK (1.6±.13 kJ/molK), while the WT and WT(his) values were –697±39.4 (–2.9±.16 kJ/molK) and –1090±82.3 cal/molK (–4.5±..34 kJ/molK) respectively.

Examining the Stoichiometry

A plot of the fractional stoichiometry of the interactions of the three proteins can be seen in Figure 4-7. The WT and WT(his) proteins had *n*-values which were close to each other up to 35°C. This value had a mean of ~.43. After this, both the WT and WT(his) *n*-value started to decrease. The *n*-value for the N-terminal protein had a mean of 1.3±.3.

Discussion

IA_3 Structure Before Binding to the Protease

Several studies have proven that the IA_3 has little intrinsic structure in the absence of YPrA (*8-11*). Circular dichroism measurements of WT, WT(his) and N-terminal IA_3 proteins indicate that without YPrA, these mutants have little secondary structure (Figure 4-8 for WT(his)). It is possible that these mutants are temperature sensitive, and that secondary structures could form at low temperatures. To test this, CD studies of the mutants at various temperatures were done. Examination of the CD spectra shows that there is very little variation in secondary structural content as temperatures are lowered.

IA_3 Structure After Binding to YPrA

Crystal structures of IA_3 (and IA_3 mutants) bound into YPrA have been made and studied by several groups (*9, 11*). All of these crystal structures show IA_3 as an amphipathic helix bound into the inner side of the binding cleft of YPrA (Figures 1-5 and 3-2). The hydrophobic side of the helix is surrounded by YPrA, while the hydrophilic side is mainly oriented out of the binding cleft towards the solvent (*10*).

CD spectra of IA_3 in the presence of YPrA do show a mix of secondary structure (*9*). Comparison of the CD spectra of YPrA alone with spectra of YPrA and IA_3 also show a slightly higher helical content, which could suggest that IA_3 is folded in the presence of YPrA (*9*).

Green *et al.* showed NMR studies of the YPrA·IA_3 complex (*9*). In these spectra it is evident that several of the resonances associated with the N-terminal residues of IA_3 shift or disappear upon binding to YPrA. These changes were interpreted as the N-terminus of IA_3 transitioning through a conformational change to a structure similar to the crystal structure. Additionally, it was noted that residue peaks identified in the C-terminus of IA_3 significantly broaden in the presence of YPrA, indicating intermediate exchange and suggesting that C-terminus is interacting with YPrA (*9*).

Potency of WT, Nterm, WT(his)

The interaction of IA_3 and YPrA is extremely strong. K_i values for the mutants in this study are all in the subnanomolar range (*9, 133, 135*).

Heat Capacity (ΔC_p) Value Comparison

The heat capacity of a protein can change upon binding (*204-206*). Binding can be seen as leading to a more condensed polypeptide environment within a sample. The physical effects of heat capacity are many and may include effects on solvent exposure, electrostatics, hydrogen bonding, among others (*207, 208*). Considering solvent exposure, heat capacity changes mirror these condensing interactions, by exhibiting a lowered heat capacity as the hydrophobic residues are buried and surfaces of both proteins are less solvent exposed (*204, 209, 210*). This has been extended to protein folding as well, as a lowered heat capacity can be used to qualitatively describe a protein which undergoes a transition from open, solvent exposed coil to condensed secondary structure (*211-213*). The IA_3·YPrA system involves folding of the IA_3 molecule and binding of IA_3 to YPrA; this complicates analysis of the enthalpy value since it involves terms from both folding and binding.

Empirical evidence states that changes in polar and non-polar surface area are correlated with changes in heat capacity (*204, 207*). However, several authors have suggested that heat

capacity changes may not simply reflect surface area changes and hydrophobic interactions, but are also influenced by electrostatic interactions, hydrogen bonding, and conformational entropy (*204, 214*). Of these effects, only solvent exposure of polar and apolar residues and their effect on heat capacity is understood (*203, 215-218*), leaving understanding of the other contributions uncertain.

The heat capacity values for the IA_3 proteins in this study appear to agree with these ideas as well. The ΔC_p values for the three proteins interacting with YPrA (all in the same solution environments) are markedly different (Table 4-2). The 68 amino acid WT IA_3 has a temperature dependent ΔC_p = –697±39.4 cal/molK (–2.9±.16 kJ/molK), while the 76 amino acid WT(his) IA_3 has a ΔC_p of –1090±82.3 cal/molK (–4.5±.34 kJ/molK). The N-terminal mutant IA_3 is only 31 amino acids, and has a ΔC_p of 387±32.3 cal/molK (1.6±.13 kJ/molK), which is more than half that of the WT IA_3. The ΔC_p values of all three proteins are negative indicating that the interactions between the IA_3 proteins and YPrA lead to less solvent exposed and more condensed systems. This is consistent with the idea that when IA_3 binds to YPrA, it buries hydrophobic residues into the proteinases, as well as covers over hydrophobic residues in the proteinases. In the process, YPrA and the unstructured IA3 form a more condensed system.

The ΔC_p values of the WT and WT(his) proteins are much more negative than the N-terminal protein. This could be indicative of other interactions of the C-terminus of the longer proteins with YPrA. At the pH of the experiments, the his_6 tag on the WT(his) IA_3 leads to a much more positive molecule. This positive tip at the C-terminal end of the WT(his) protein could interact with the negatively charged surface of YPrA more readily than that of the WT protein (Figure 4-9). While the C-terminal end of WT IA_3 could potentially act as a tether

between the N-terminus and YPrA, the C-terminus of WT(his) would act as 'stickier' tether between YPrA and the inhibiting N-terminal portion of IA_3.

The N-terminal protein, however, does not have the benefit of this tether. While only the N-terminus is necessary to inhibit YPrA, the smaller ΔC_p value seems to suggest that the nature of the interactions between the N-terminal protein and YPrA are not as strong as compared to the interactions between YPrA and the WT or WT(his) proteins.

***N*-value Comparisons**

The WT IA_3 consists of 68 amino acids, of which only half are necessary to inhibit YPrA. If the N-terminus of one IA_3 molecule interacted with one YPrA, then the ligand to protein ratio, the *n*-value, should most certainly be one. This is not completely the case, however, since the average *n*-value for the N-terminal protein is 1.3±.3. Noninteger *n*-values have been seen in studies of other protein systems, and have been attributed to additional binding sites with low cooperativity, additional folding events or misfolding (*219*), aggregation (*220*), or additional protonation states (*221*). The numerical value itself is confounding, however it may indicate that more than one N-terminal protein may be found interacting with each YPrA molecule. Inhibition of YPrA takes place when one IA_3 molecule binds into the flaps of the active site cleft. The ITC binding data could suggest that more than one N-terminal protein binds to YPrA, which further suggests that these proteins may be interacting somewhere other than the YPrA active site. Rather, once YPrA is inhibited by one N-terminal IA_3 molecule, other positively charged N-terminal molecules may be simply interacting nonspecifically with the negatively charged surface of YPrA (Figure 4-9).

The *n*-values of the WT and WT(his) proteins also hold an interesting story. The addition of the C-terminal residues to the N-terminal IA_3 causes a dramatic shift in *n*-value. The *n*-value of WT IA_3 averages to 0.41±.09, while the average value for WT(his) is 0.43±.05. The values,

however seem to be similar up to 35°C, after which they both become less positive. The average *n*-value from 10°C to 35°C, for both WT and WT(his) is ~0.42. This value indicates that each molecule is binding more than one YPrA. At temperatures higher than 35°C the *n*-values of both WT and WT(his) proteins decrease, which could be indicative of more solvent exposed binding regions due to a loosening of the YPrA·IA_3 complex, or an increase in the accessible surface area of more mobile regions of the complex. Aggregation of the samples at these temperatures was not seen.

N-terminus Versus C-terminus

Why would a cell spend resources and energy to create a longer protein when a smaller version would do the job sufficiently? From the above description and analysis it would seem that the N-terminal end of IA_3 is all that is necessary to inhibit YPrA. The C-terminus of the protein appears to play some role in the interaction of YPrA with the protein. The longer IA_3 proteins have more negative ΔC_p values indicating that they are more involved in producing a less solvent exposed or more condensed system, namely the interaction between the IA_3 protein and YPrA. This suggests that the C-terminus of the protein acts to bring YPrA and IA_3 together, while the action of the N-terminus of the protein is to inhibit YPrA (Figure 4-9). The N-terminal IA_3 is unstructured in the absence of YPrA so it does not serve as a strong test of the fly-casting model; however the addition of the unstructured C-terminus does change the heat capacity of the interactions between YPrA and IA_3. The C-terminus is not seen in the crystal structure which may point to not only the unstructured nature of the C-terminus, but that the interactions between the C-terminus and YPrA are weak and fleeting. In investigations of the interactions between IA3 and YPrA using NMR, Green *et al*. noted that the C-terminal residues of IA3 appeared to be interacting with YPrA. These weak interactions do not appear to enhance the inhibitory binding

of the N-terminus, but could instead promote the meeting of the N-terminus with binding partners in the active site of YPrA (Figure 4-9).

The WT(his) protein serves as an interesting variant of this theory. WT(his) IA_3 has a positive charge at pH 4.5 which is nearly double that of the WT protein. The ΔC_p values of the two proteins are different, however the ΔH values of both are reasonably close to each other at all temperatures as compared to the N-terminal protein. The source of this difference in ΔC_p values is unknown, but could perhaps be explained by the weak charge interactions between the surface of YPrA and WT IA_3 being fewer/less than the charge interactions between YPrA and WT(his).

The WT(his) explanation above is similar to that of Winterburn *et al.* (*135*). In Winterburn's work the authors modified the 68 amino acid WT IA_3 with an N-terminal his_7 tag. This addition did not affect the affinity of IA_3 for YPrA, however it broadened the specificity of the molecule to allow IA_3 to act as an inhibitor for a related protease, vacuolar aspartic proteinase from *Pichia pastoris*. Winterburn's explanation of this broadening of the selectivity was based on increasing the local concentration of IA_3 around the active site by taking advantage of the charge interaction of the his_7 tag and the surface of YPrA (*135*). In a similar fashion the charged his_6 tag on the C-terminus of the WT(his) IA_3 and the C-terminus of WT IA_3 generally may be responsible for raising the local concentration of N-terminus around the active site cleft in YPrA (Figure 4-9).

Conclusion

IA_3 is a unique protein molecule, both in secondary structure and the selective, potent, and distinctive way that it inhibits yeast proteinase A. The N-terminus of the inhibitor tightly binds YPrA, however the C-terminus has been mysterious and largely ignored due to its unstructured nature. Here we present evidence that the C-terminus may serve to enhance the meeting of the

N-terminus of IA_3 with the active site of YPrA. The C-terminus may do this by exploiting weak charge interactions between its positive end and the negative surface of YPrA, thus raising the local concentration of the inhibiting portion of IA_3 for the protease.

Table 4-1: Properties of the proteins in this study. The molecular weights, pI, and charge at pH 4.5 were calculated using EMBO tools. The molecular weights were verified using MALDI mass spectrometry.

Protein	Number of amino acids	Molecular weight	Calculated pI	Calculated charge at ph 4.5
WT	68	7707.4	7.63	+5.4
WT(his)	76	8772.5	6.82	+10.6
N-term	31	3483.9	9.20	+2.3
YPrA	329	35749.2	4.16	–7.6

Table 4-2: Heat capacity of binding (ΔC_p) and *n*-values for WT, WT(his), and N-term IA_3 derived from ITC data.

Protein	ΔC_p (cal/mol·K)	ΔC_p (kJ/mol·K)	*n*-value
WT	–697±39.4	–2.9±0.16	0.41±0.9
WT(his)	–1090±82.3	–4.5±0.34	0.43±0.05
N-term	387±32.3	1.6±0.13	1.3±0.3

Figure 4-1: Schematic representations of the IA_3 proteins. The sequence used is listed at the top. In red are representations of the three proteins, the 68 amino acid wild type (WT), the 76 amino acid wild type protein with a his_6 tag addition (WT(his)), and the N-terminal IA_3 protein (N-term). The region which is found in a helix in the IA3·YPrA crystal structure is indicated by blue H's.

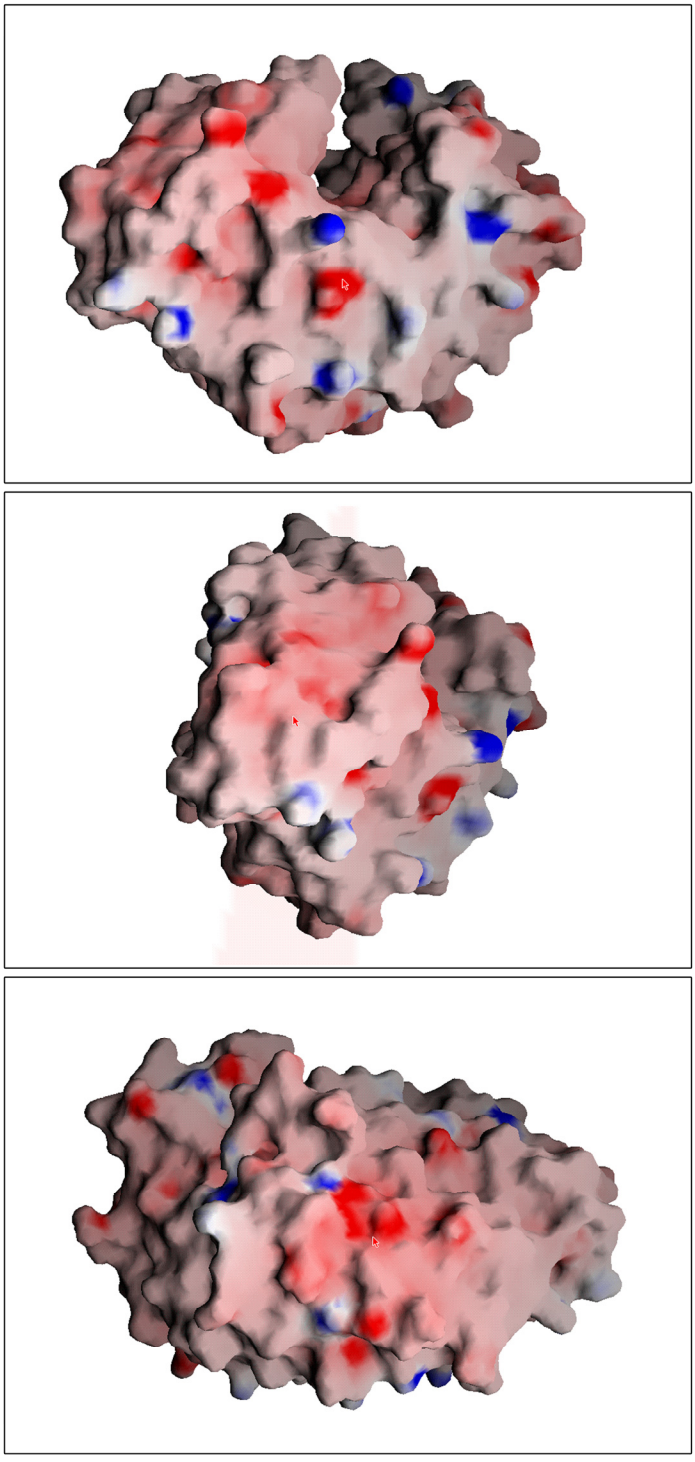


Figure 4-2: Several views of YPrA illustrating the surface charge on the protease. Negatively charged surface is in red, while positive elements are in blue.

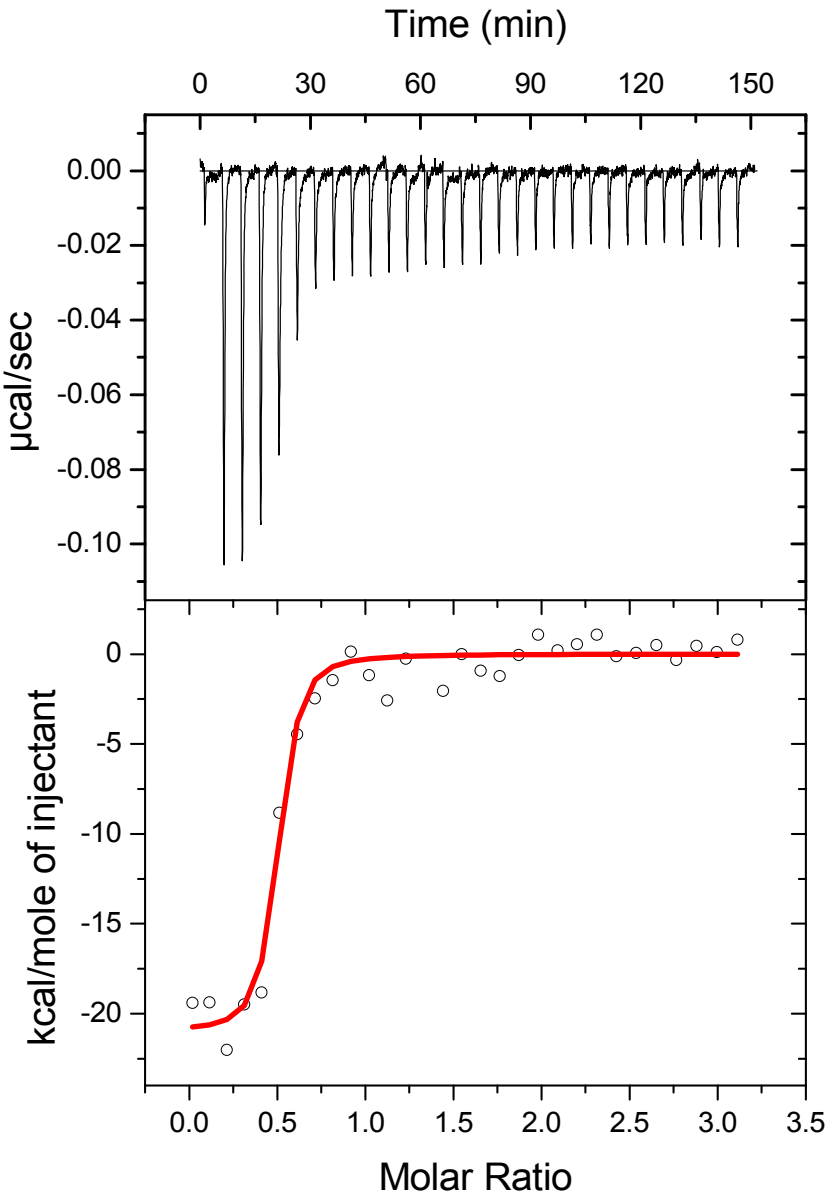


Figure 4-3: Representative ITC result for 13µM WT IA_3 titration into 1µM YPrA at pH 4.5. Panel A shows the raw power data from the microcalorimeter as a function of time. Panel B shows the integration of the baseline subtracted data at each injection as circles. The red curve is a fit to the data as explained in chapter 5.

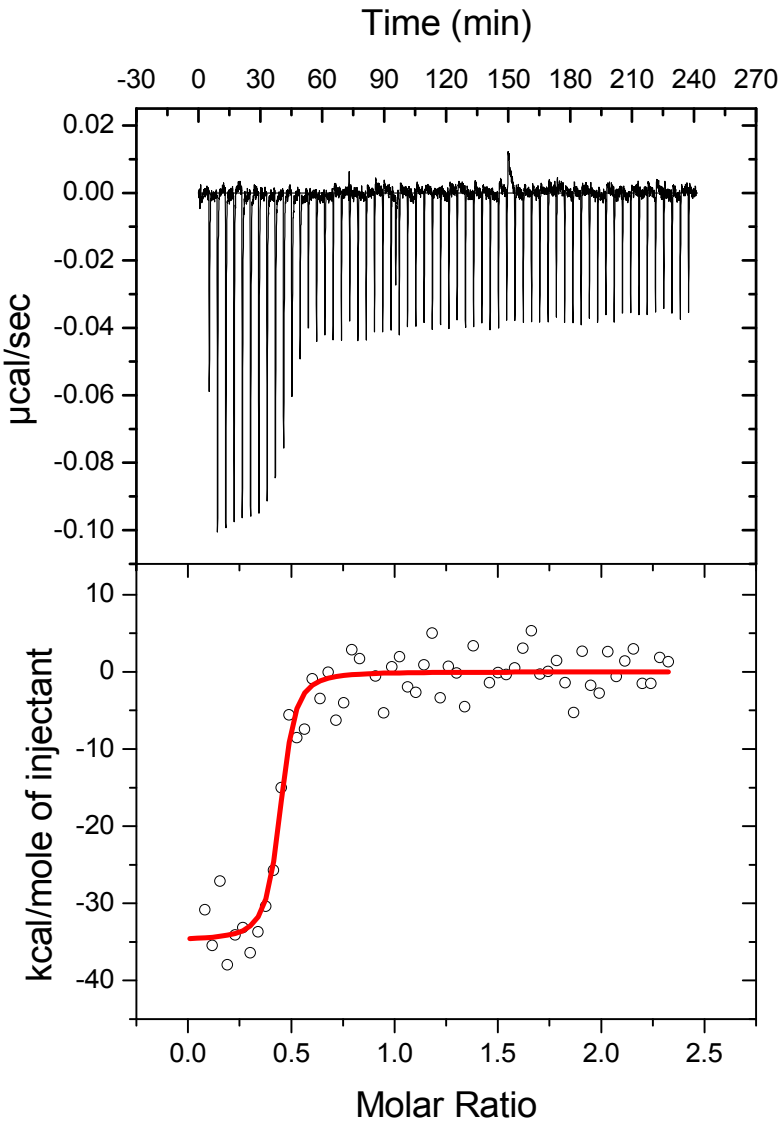


Figure 4-4: Typical ITC result of 13µM WT(his) IA_3 titration into 1µM YPrA at pH 4.5. Panel A shows the raw power data from the microcalorimeter as a function of time. Panel B shows the integration of the baseline subtracted data at each injection as circles. The red curve is a fit to the data as explained in chapter 5.

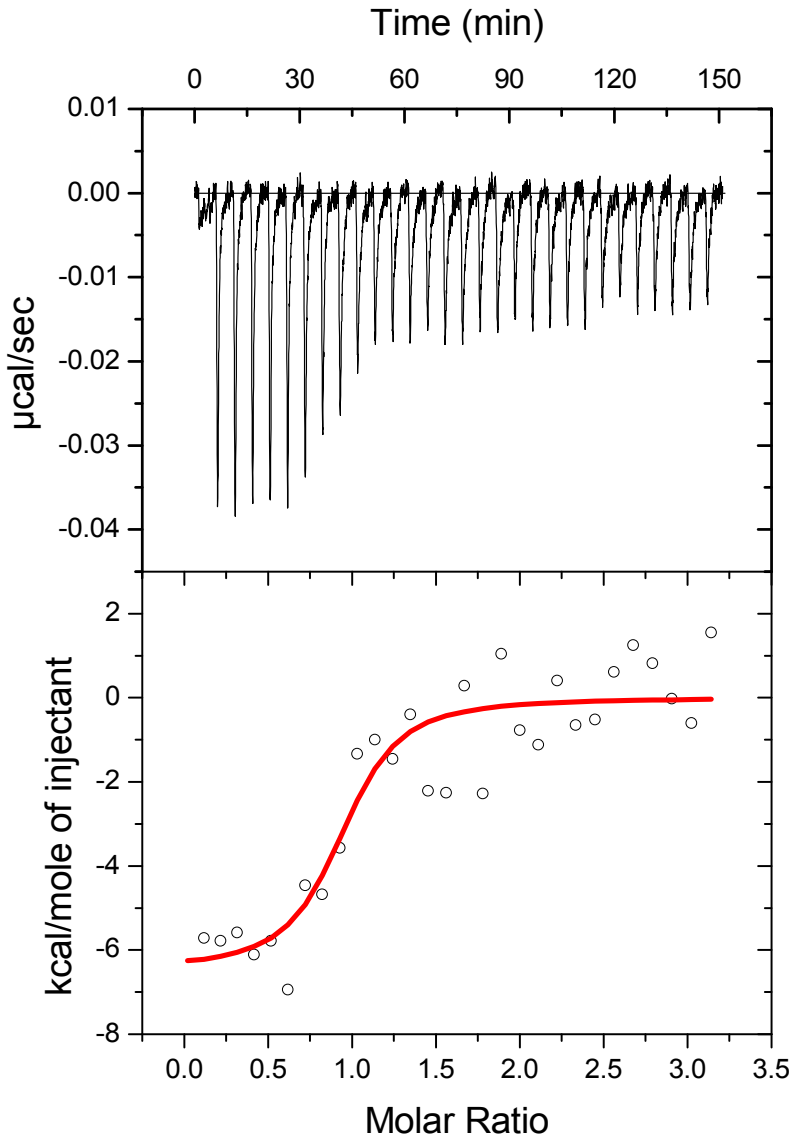


Figure 4-5: Typical ITC result for the 13µM N-terminal IA_3 titration into 1µM YPrA at pH 4.5. Panel A shows the raw power data from the microcalorimeter as a function of time. Panel B shows the integration of the baseline subtracted data at each injection as circles. The red curve is a fit to the data as explained in chapter 5.

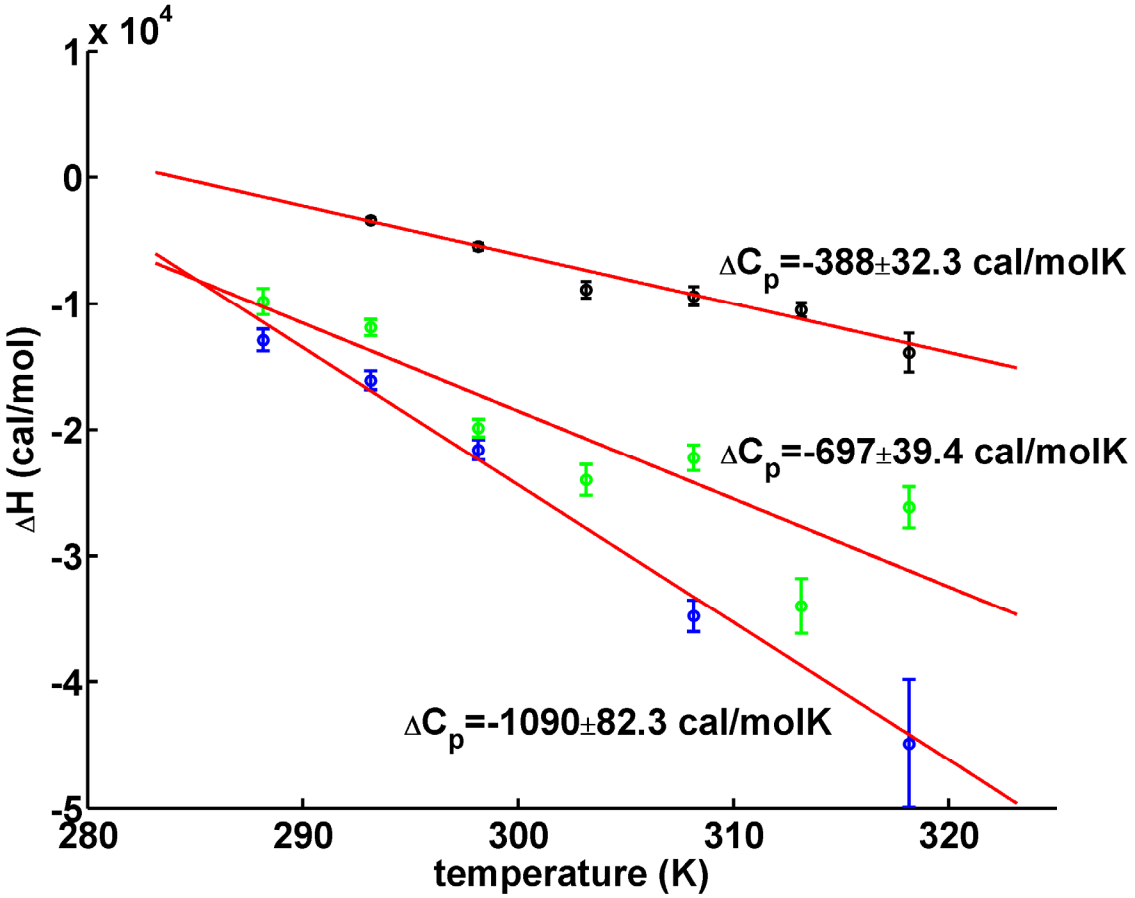


Figure 4-6: Change in enthalpy of binding (ΔH) values for the three proteins. The ΔH values for WT, WT(his) and Nterm are represented here in green, blue and black respectively. Fits of these values as a function of temperature according to Equation 4-1 are plotted as red lines, with the ΔC_p value listed accordingly.

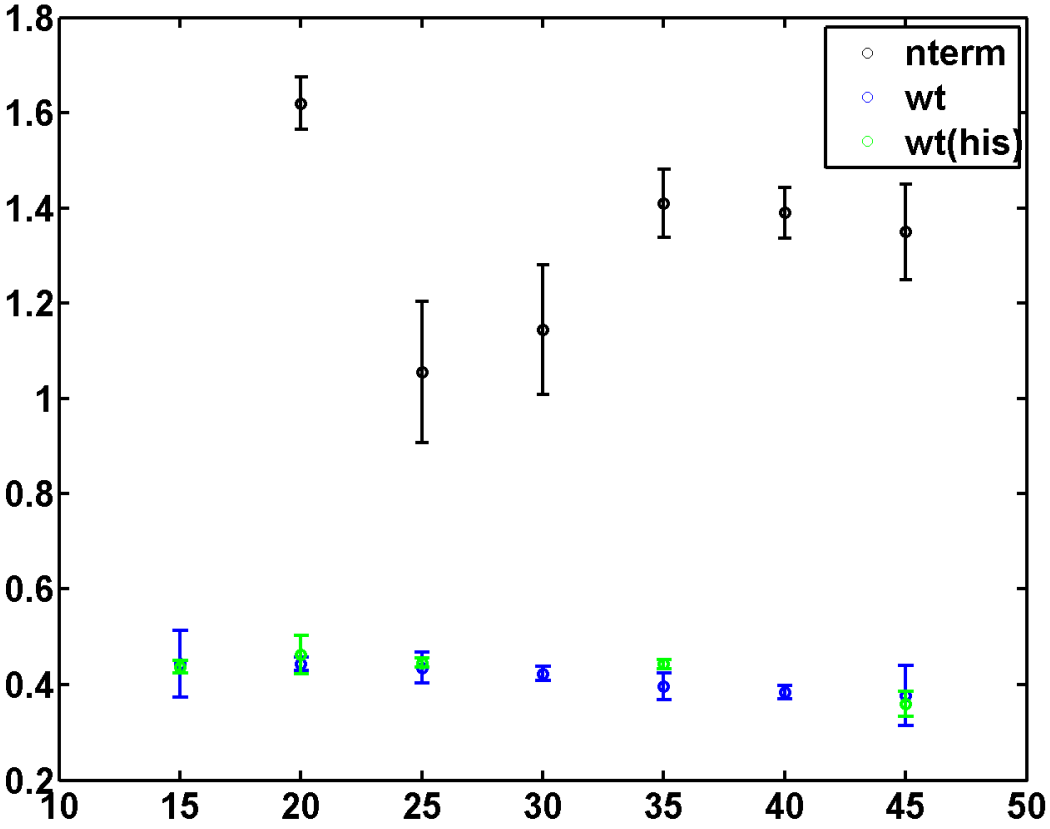


Figure 4-7: Plots of the stoichiometry value (*n*-value) for IA_3 protein to YPrA. The *n*-values for WT, WT(his) and Nterm are represented here in blue, green and black respectively.

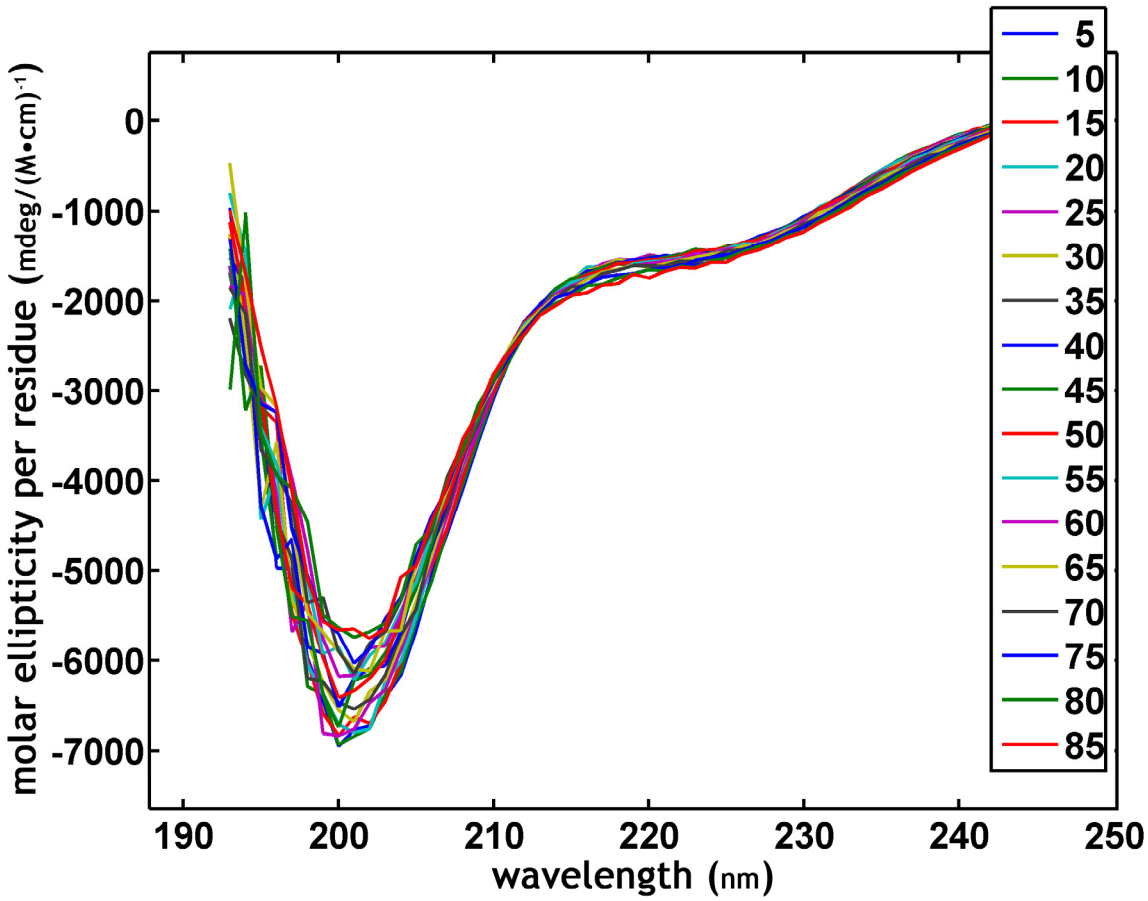


Figure 4-8: Circular dichroism spectra for WT(his) IA_3. WT(his) at 100uM is plotted for several temperatures ranging from 5°C to 85°C. All of these spectra have been overlaid to illustrate the lack of intrinsic structure in the IA_3 protein at all temperatures.

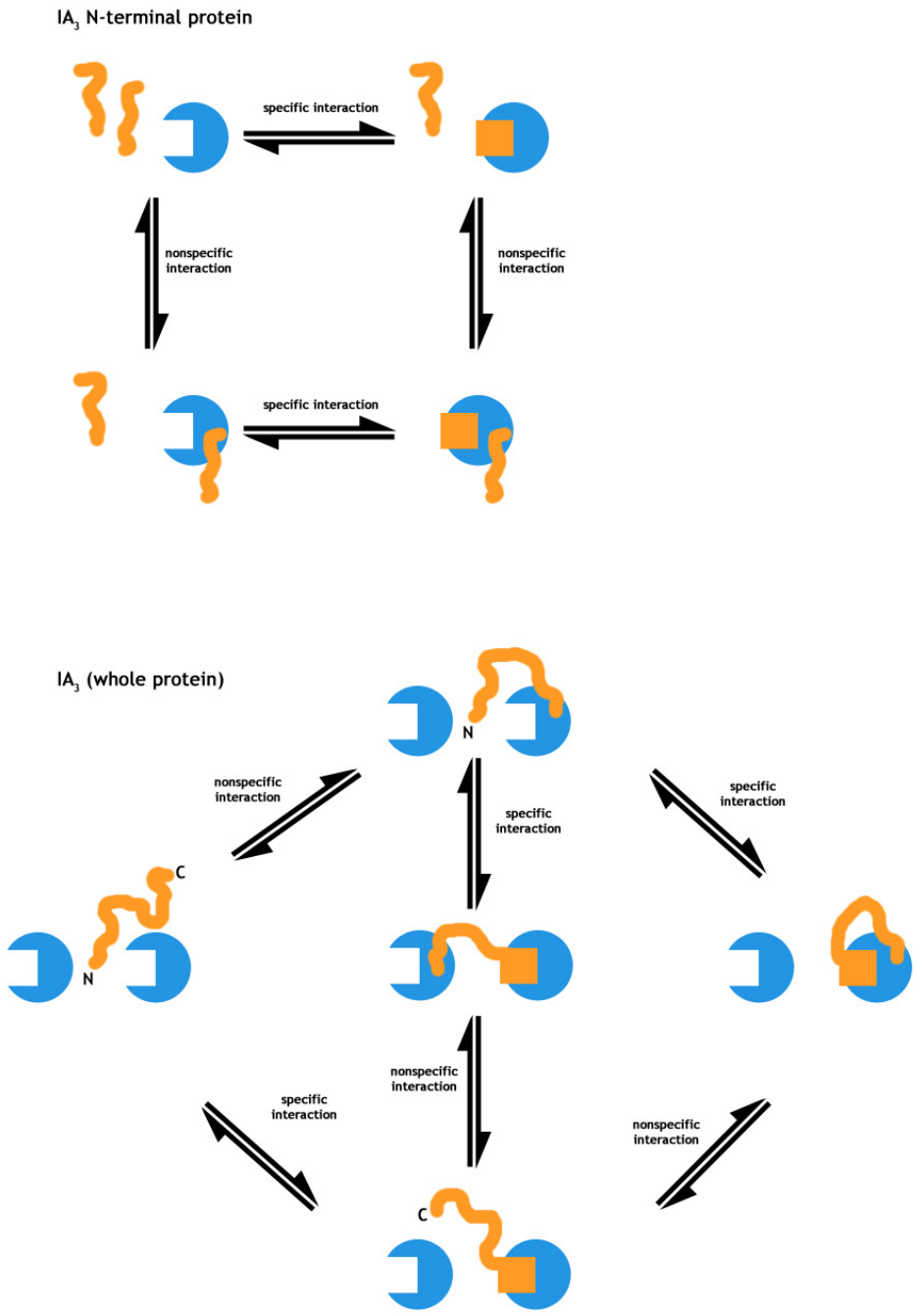


Figure 4-9: Possible models for IA_3 binding. IA_3 is in orange, and is represented as a squiggle when unfolded. Helical, bound IA_3 N-terminus is represented as an orange box, bound into YPrA in blue. Panel A shows binding of the N-terminal IA_3; Panel B shows possible interactions of the whole IA_3 protein with YPrA with the N and C termini of IA_3 labeled as indicated.

CHAPTER 5
METHODS AND PROTOCOLS

Due to the many different biophysical experiments that were done on IA_3 and YPrA, there are many methods associated with them. Many of these protocols deal with purification of proteins, particularly several different mutants of IA_3, WT IA_3, and YPrA. Additionally, some time will be spent illustrating some features of these protocols such as considerations to be noted concerning IA_3 cysteine mutants and the temperature stability of YPrA. A labeling protocol for attaching dansyl to cysteine mutants of IA_3 is also given below. These biochemistry protocols are followed by general methods used in several experiments.

IA_3 Purification

Purifying IA_3 Histidine Tagged Proteins

Working with IA_3

When working with IA_3, a main concern is contaminating proteases (*11, 130, 222, 223*), since unstructured proteins are particularly susceptible to proteolytic degradation (*224*). Within the cell proteolytic degradation is considered to be a method of quality control, attempting to ensure that misfolded or nonfolded proteins are removed (*224, 225*). The specificity of IA_3 towards YPrA and its inability to inhibit proteases which are even similarly to YPrA (both in sequence and structure) make it a candidate substrate for other proteases (*11, 130, 222, 223*).

An additional concern with IA_3 is the possibility of interaction with other molecules. A proposed advantage of a protein being intrinsically disordered or having a disordered region is so that it may interact with many partner molecules (*226*). Examples exist of unstructured proteins, or proteins with unstructured regions, interacting with multiple partners (*226, 227*). As an intrinsically disordered protein, IA_3 has the potential to interact with other molecules in a way

that could alter its interaction with the protease, or yield results which would confuse our understanding of the interaction of IA_3 to YPrA.

These issues are reasons why IA_3 must be separated from other contaminating molecules. Removal of other proteins allows us to investigate and characterize IA_3 and its interaction with YPrA without dealing with other effects such as degradation or other interactions.

Purifying IA_3

Histidine tagged IA_3 has several properties that are unique in a solution of *E. coli* cell lysates. Foremost is the his_6 tag which was added to the protein to aid in purification. This tag has a micromolar affinity towards nickel (and cobalt) (*228*), and is used simply by flowing the cell lysate over immobilized nickel and 'washing' away other molecules with various buffer solutions (*228, 229*).

An additional exploitable property of IA_3 is the unstructured nature of the protein. Boiling or freezing a structured protein will often disrupt secondary and tertiary interactions present within the protein (*230*). Often these disrupted structures lead to aggregation (*230, 231*). Unstructured proteins have higher tolerances for boiling and freezing as compared to most well structured proteins (*6, 7, 174*). The disordered nature of unstructured proteins can allow these proteins to evade this aggregation.

Purification protocol

The purification scheme that is laid out here is similar to schemes previously followed (*9, 11*). However, some significant differences allow this scheme to be more efficient and robust in producing IA_3. Additionally, this purification scheme serves as a template to compare differences with other purification schemes in this chapter.

- **Preparation**
  prepare the following buffers as indicated:
    - pH 8 buffer: 50 mM phosphate buffer + 300 mM NaCl at pH 8

- pH 6.3 buffer: 50 mM phosphate buffer + 300 mM NaCl + 10% glycerol at pH 6.3
- pH 4 buffer: 50 mM phosphate buffer + 300 mM NaCl at pH 4

- **Protocol**
  - Step 1: Boil cell lysates for at least 15 minutes, then rapidly cool to room temperature.
  - Step 2: Centrifuge the solution at 27,000 g for 15 minutes.
  - Step 3: Filter the solution through a .2 micron nylon filter to remove aggregates.
  - Step 4: Flow the filtered solution over a nickel column pre-equilibrated in pH 8 buffer
  - Step 5: Wash the column with 2 column volumes of pH 8 buffer, followed by 2 column volumes of pH 6.3 buffer
  - Step 6: Separate the bound histidine tagged IA_3 from the nickel column by washing the column with 2 column volumes of pH 4 buffer.

Further purification or characterization of the histidine tagged IA_3 can be done using size exclusion chromatography. The presence of DNA was tested for and was not found in purified proteins following this purification scheme. The protein can then be dialyzed to a new buffer or to water and lyophilized.

All histidine tagged IA_3 proteins have molecular weights around 8 to 9 kDa based on the amino acid sequences. SDS-PAGE gels of IA_3 show that these proteins appear at their expected molecular weights (Figure 5-1). All IA_3 variants and mutants run as expected under reducing and denaturing conditions, and display a band at the expected molecular weight.

Purifying WT IA_3

WT histidine tag free IA_3

Purification of the WT IA_3 presents a different challenge as compared to the other IA_3 variants. The eight amino acids present in the histidine tagged IA_3 mutants provides an affinity for nickel, which aids in purification. The lack of this tag in the WT IA_3 essentially eliminates

this technique as a method for purification. Different properties of the WT IA_3 need to be exploited, since the histidine tagged IA_3 variants rely on nickel affinity columns for purification.

The unstructured nature of IA_3 can still be used to separate it from a solution simply by boiling or freezing. Another important property is the mobility of the WT protein through a size affinity column, as the molecule is relatively small at 68 residues (average protein length is between 300 and 400 residues) (*232-234*). For an appropriate size gel filtration column WT IA_3 may move slower than other components in a solution. However, size exclusion chromatography relies on globular proteins to make its size estimates. Because IA_3 is unstructured, its mobility in a size exclusion column is different than a globular protein of the same molecular weight. An additional property is the charge of the WT molecule at a certain pH. Separation based on charge can be achieved using an anion exchange column (Figure 5-2).

The protocol

- **Preparation**
  prepare the following buffers as indicated:
  - Anion Exchange buffer: 50mM phosphate buffer, at pH 8

- **Protocol**
  - Step 1: Boil the sample as indicated in the histidine tagged IA_3 protein protocol
  - Step 2: After boiling, rapidly freeze the samples on liquid nitrogen (or in a -80°C fridge)
  - Step 3: Thaw the samples and centrifuge at 48,000 g for 20 minutes to pellet aggregates
  - Step 4: Filter the samples through a .2 micron filter
  - Step 5: load the sample onto a Superdex 75 (16/60) column which has been equilibrated with anion exchange buffer
  - Step 6: Run the protein solution through the column and collect the peak of interest
  - Step 7: Run the collected fractions through a HiTrap QFF column which has been equilibrated with anion exchange buffer and collect the flow through.

The protein can then be dialyzed to a new buffer, or to water and lyophilized.

Cysteine mutants and dimer formation

The cysteine mutants of IA_3 pose a unique concern as compared to the other IA_3 variants. The addition of a sulfhydryl group in these mutants adds a very reactive and strongly reducing residue to the protein. Cysteine residues in two IA_3 molecules can easily be oxidized, forming a disulfide bridge. Dimer formation could result in a radically different IA_3 molecule.

A dimer would affect results which characterize IA_3. The most obvious differences would be in molecular weight and the size of the protein. Dimerization would result in a protein which is (roughly) double in molecular weight, while the apparent 'size' of the molecule could potentially double as well. The formation of disulfide bonds affects higher order structures in proteins by linking protein molecules together.

Purification of all IA_3 mutants can be followed by SDS-PAGE gel analysis, and shows IA_3 as a band around its expected weight (Figure 5-1). All IA_3 mutants display similar characteristics in gel filtration with the exception of mutants which involve a cysteine mutation (Figures 5-3 and 5-4). These mutants display anomalous mobility during gel filtration chromatography as compared with WT IA_3, WT(his) IA_3, or other IA_3 mutations (Figure 5-3). Denaturing and reducing conditions during gel analysis indicate that cysteine mutants elute in later fractions from gel filtration chromatography (Figure 5-4). It is possible that dimer formation between two IA_3 molecules through disulfide bonding of the free cysteine residues is the cause of retarded mobility within the gel filtration column.

Native gels were used to evaluate possible dimer formation of the cysteine mutants of IA_3 (Figure 5-5). Native gels are similar to SDS-PAGE gels, however they differ in the removal of SDS and BME, and the samples are not boiled (*235, 236*). Without the denaturing effects of BME and SDS, a protein may still retain some three-dimensional structure, as opposed to

adopting a more open structure in the presence of BME and SDS (*235, 236*). With SDS removed, the charge of the protein can influence its mobility with a gel, since the negative charges of SDS no longer coat the protein (*235, 236*). Essentially the native charge of the protein is no longer masked out by a net negative charge due to SDS binding.

Due to these effects, mobility of a protein within a native gel is significantly different as compared to an SDS-PAGE gel (*236*). Proteins migrate through the gel according to their electrophoretic mobility; however, the sieving properties of the gel resist their movement within the gel (*237-242*). Lower molecular weight proteins are generally smaller and encounter less of this retarding force; consequently they migrate more quickly through a gel than larger proteins of higher molecular weight (*237-242*).

Native gels of the IA_3 cysteine mutants could indicate if dimer formation is present. Samples of IA_3 cysteine mutants were run on native gels and SDS-PAGE gels of similar acrylamide composition (Figures 5-6). In Figure 5-6 the K24C IA_3 single mutant appears similar to WT IA_3, suggesting that this mutant has not dimerized. The other cystine mutations, however, exhibit very different responses in the native gel. All of the cystine double mutants, and the K16C single mutant show significant smearing (Figures 5-5 and 5-6). The N2W-K16C IA_3 double mutant displays a band which is in line with WT IA_3, however it displays prominent bands above and below as well. All cystine mutants which contain the K16C mutation show a band with increased mobility below WT IA_3, however the large smear below this band is confusing. The SDS-PAGE gel of the same samples shows all the samples in line with WT IA_3. The results of this analysis indicate that IA_3 mutants containing the K16C mutation are forming structures which are not the same as WT IA_3.

Dansyl Labeling IA_3

Specifics of the Reaction

The thiol reactive probe 1,5-IAEDANS is used to add a dansyl moiety to IA_3 cysteine mutants. The reaction mechanism is nucleophilic bimolecular substitution (*243*) (Figure 5-7). In this mechanism cysteine acts as a nucleophile towards the IAEDANS molecule (*243*). The iodo group in the IAEDANS molecule draws electrons towards itself, causing a partial positive character on the adjacent carbon. The negative charge on the reactive sulfhydryl group is attracted to this carbon immediately attached to the iodo group (*243*). This not only enhances the attraction of the sulfhydryl group towards the carbon, but makes the iodo group an excellent candidate as a leaving group in the reaction (*243*). The sulfhydryl attacks the carbon as a backside attack, essentially popping off the iodo group (*243*).

The Protocol

The protocol provided by Molecular Probes (now owned by Invitrogen) has changed several times over the last 6 years. Over this time there has been varying degrees of success with these supplied protocols. Understanding the reaction allowed creation of a simple protocol which works well at room temperature. The reaction is sufficiently efficient using this protocol to achieve 50-80% labeling.

- **Preparation**
  prepare the following buffer solutions as indicated:
  - Labeling buffer: 50 mM phosphate buffer at pH 7.3
  - Dye solution: prepare the dye at a 20x mole concentration of dye to protein. Dissolve the IAEDANS dye in DMF using as little DMF as possible. A preferred dye concentration is 10 mM. Wrap foil around the dye solution to protect the solution from light.

- **Protocol**
  - Step 1: prepare the protein at a concentration of 10-100 μM in labeling buffer
  - Step 2: add TCEP to a final concentration of 10 times the protein concentration

- Step 3: allow the solution to stir for 10 minutes at room temperature
- Step 4: cover the flask with foil
- Step 5: add the dye solution dropwise to the reaction flask
- Step 6: blow nitrogen over the reaction then tightly cap
- Step 7: allow the reaction to stir at room temperature for 2 hours

Purification of the labeled IA_3 from the reaction mixture can be done using a Vydac C18 reverse phase column and watching for a peak at both 278 nm (for tryptophan) and 337 nm (and dansyl) (Figure 5-8). Unreacted dansyl, salts and DMF leave the column early, followed by the labeled IA_3, then unlabeled material. The labeled fractions can also be seen by shining a UV light on the collected fractions.

Working With YPrA

Purifying YPrA

Why is purification necessary?

YPrA that is purchased from Sigma may not have a purity that is satisfactory for our purposes. Past experience has shown that different lots of YPrA could contain a contaminating protease (possibly YPrB) (*private communication, T. Green*). This poses two problems. IA_3 is a specific inhibitor towards YPrA, and may not inhibit the other protease. Additionally, the contaminating protease may proteolyze both YPrA and IA_3. For these reasons further purification of the YPrA is necessary before use.

Purification concerns

Purification of YPrA is also problematic since the protease has been known to degrade itself. Two methods are used to combat the possibility of self degradation. The protease is active at acidic pH 4.5. All of the separation is done at pH of 6.5 in a buffer in which the

protease is stable but has a minimal activity. Additionally, the whole purification is done at 4°C to lower the amount of thermal energy present for proteolysis to occur.

The protocol

- **Preparation**
  prepare the following buffers as indicated:
    - Buffer A: 20 mM phosphate buffer with 50 mM NaCl, at pH 6.5
    - Buffer B: 0.1 M NaAc, pH 4.5
    - PMSF solution: 100 mM phenylmethylsulphonyl fluoride (PMSF) in isopropanol
    - Equilibrate a Sephacryl (S-100) High Resolution column with Buffer A, place at 4°C

- **Protocol**
- Step 1: Dissolve YPrA into 500 uL of Buffer A, and add to it 10 uL of PMSF solution
- Step 2: Load the sample onto the S-100 column
- Step 3: Run the column at 4°C, collecting 2 mL fractions, this will take approximately 3 hours. Collect the second peak.
- Step 4: Concentrate the pooled fractions using an Amicon concentration cell with a 3,000 NWML membrane. Do this at 4°C.
- Step 5: Switch the buffers from Buffer A to Buffer B using the Amicon cell. YPrA will be active in this buffer.

  Aliquot the sample into 1 mL vials, and store at -80 until needed.

The Stability of YPrA

Stability of YPrA is crucial for several of the experiments in this dissertation. The role of YPrA is to proteolytically degrade other protein substrates found within the same vesicle as itself (*223*). This action has a preference for other substrates, and disfavors degradation of itself and its inhibitor IA_3 (*222*). Nevertheless, at high concentrations or given enough time, YPrA could degrade itself or its inhibitor.

Temperature stability of YPrA

In an SDS-PAGE gel, mature YPrA appears as a band at ~42 kDa (*132*). Stability studies of YPrA under these conditions can be enlightening. As a protease, YPrA can cleave other

proteins including itself. A concern during any experiment is whether YPrA will cleave itself as data are being gathered. Twenty micromolar YPrA was allowed to equilibrate in 0.1M NaAc at pH 4.5. SDS-PAGE gels of a time course of YPrA allowed to equilibrate at various temperatures would show if YPrA is cleaving itself (Figure 5-9). YPrA which is not degraded would show up as a clear band in the gel, while the band for degraded YPrA would have a lower intensity, and possibly appear 'smeared' depending on the size and concentration of the degradation products.

While there may be slight degradation at the highest temperature and longest time period, it appears that 20 uM YPrA in 0.1M NaAc, pH 4.5 is stable up to 3 hours at 55°C (Figures 5-9 and 5-10).

Stability of the YPrA and IA_3 complex

Once it has been realized that YPrA appears stable over the course of a 3 hour experiment, the next concern is for the YPrA·IA_3 complex. It was important for us to verify that IA_3 would not be degraded by YPrA over the course of an experiment. One to one concentrations of YPrA and IA_3 were allowed to equilibrate in 0.1M NaAc at pH 4.5 for various amounts of time at 25°C (Figure 5-11). These results indicate that YPrA does not degrade IA_3 at 25°C for up to 3 hours.

Circular Dichroism Experiments

What Is CD Used For?

The differential absorption of circularly polarized light can be used to identify and characterize proteins and secondary structures (*244*). The measurement of the absorption of right and left circularly polarized light from a sample is known as circular dichroism spectroscopy (*244*). Secondary structures present in proteins cause distinct bands of differential absorption which can be viewed using CD (Figure 5-12).

While CD can illustrate the overall secondary structural character of a protein, it is probably best used to monitor secondary structural changes as a function of temperature or an

additional chemical. Wavelength scans can be used to assess secondary structural changes as a denaturant or renaturant is titrated into the sample (Figure 5-13). Temperature scans involve monitoring a certain wavelength while changing the temperature of a sample (Figure 5-14). These two modes are not independent of one another, and can be combined to assess how chemical addition and temperature simultaneously affect a sample.

Experimental Setup

CD wavelength scan experiments were typically done in the far-UV, from 190 nm to 260 nm. At these wavelengths differential absorption by secondary structures in proteins can be observed. Scans were taken with a 1 nm step size with a settling time of 3 seconds and a data collection and averaging time of 1-5 seconds. CD temperature scan experiments could typically be taken from 5°C to 85°C at 2-10C° increments. The temperature equilibration time depended on the particular experiment, however usual equilibration times were 1-5 minutes. Data were collected for 1-5 seconds at each temperature. The instruments used for these experiments were all Aviv CD spectrometers, models 202, 215 and 400 using CD software from Aviv Biomedical, Inc. (Lakewood, NJ).

ITC Experiments

What Is ITC?

Molecular level thermodynamics can point towards molecular interactions between molecules, and the energetics associated with their binding (*211*). Association events can be thermodynamically characterized by several parameters (*201*):

- Association constant (K_a) defined as the ratio of complex to reactants,
- Stoichiometric ratio of interacting molecules (n),
- Enthalpy of the binding interaction (ΔH),
- Entropy of the binding interaction (ΔS),
- Free energy of the binding interaction (ΔG),
- Change in heat capacity associated with binding (ΔC_p).

Many methods exist to characterize molecular interactions (such as equilibrium dialysis and radio ligand binding assays), however most of them are indirect methods that rely on an additional phenomenon or molecule to report on the recognition events, but are external to these events (*245, 246*). One method which can be used to investigate binding events directly is isothermal titration calorimetry (ITC) (*73, 247*). All binding events either release energy (exothermic) or absorb energy (endothermic) (*72, 73, 197, 202*). This evolved heat can be directly measured by ITC, and can be used to determine the thermodynamic parameters mentioned above (*248, 249*).

The ITC Instrument

An ITC instrument can be generally characterized as a sensitive calorimeter. The instrument is conceptually simple, consisting of two identical cells made of a highly conductive material and surrounded by an adiabatic jacket (*73*) (Figure 5-15). Thermocouples detect the difference in temperature between the cells and the jacket and between the cells themselves (*72, 73*). Heaters attached to the jacket and cells maintain these elements at identical temperatures (*72, 73*).

An ITC Experiment

To understand an ITC experiment, consider the simple reaction of two molecules, A and B, binding together to form the complex AB (*250, 251*).

$$A+B \rightleftharpoons AB \tag{5-1}$$

The reaction is simple, and can exemplify many different biological processes such as RNA-protein interactions (*252*), DNA-protein (*253*), lipid-peptide (*254*), antibody-antigen (*203, 255*), hormone-receptor (*72*), enzyme-substrate (*245, 256*), or various other protein-protein interactions

(*196-199, 257-259*). For simplicity we will consider B to be the substrate or ligand to A, the macromolecule. In an ITC experiment, a solution of macromolecule A is placed the sample cell. The reference cell is filled with a solution with known thermal characteristics (usually water) (*260*). A solution of ligand B is placed in the injection syringe. In this configuration A is the titrand and B is the titrant.

During each ITC experiment a specific volume of the titrant is injected into the sample cell while the contents are rapidly stirred. At each injection of titrant into the cell, heat is absorbed or released depending on the nature of the molecular interactions taking place (*72, 73, 261*). An active feedback circuit maintains both the sample cell and the reference cell at the same temperature during each experiment (*72*).

The observable quantity measured during the experiment is the time dependent change in electrical power necessary to maintain equivalent temperatures in both the sample and reference cell (*72, 73, 261*). If the molecular interactions are exothermic, then the temperature in the sample cell will be higher than the reference cell, and power to the heaters is decreased to allow the temperature between the cells to be maintained (*73*). If the interactions are endothermic, the temperature in the sample cell is lower than the reference cell and the power to the heater is increased. The time between each injection can be varied; however sufficient time should be taken between each titration to allow for the temperature in the sample cells to equilibrate with respect to the reference cell (*72*).

During each titration a certain amount of heat is absorbed or released (*262*). This heat evolution is proportional to the number and type of molecular interactions which take place between the molecules within the sample cell, or the number of molecules of ligand which are bound by the macromolecule (*202, 261*). During initial injections, all of the ligand injected will

be bound up by the macromolecule. This will be observed as a large change in heat evolved and consequently a large signal recorded as the active feedback circuit compensates for the heat difference by changing the power to the sample heater (*73*). As the ligand concentration increases during subsequent injections, the individual molecules of the macromolecule A become saturated, leaving fewer binding events to occur, and consequently less heat evolution to take place. Eventually all binding sites are saturated, with further injections registering small heat changes resulting from dilution of the titrant into the sample cell volume (*72*).

Analyzing ITC Data

ITC data are collected as the power required to maintain equivalent temperatures in both the sample and reference cell (*73*). This value can be correlated to the amount of titrant present in the cell, plotted against time, and integrated to yield a titration curve. There are many models which could be used to fit titration data (*72, 73, 201*). The model used to fit a titration of IA_3 into YPrA is outlined below. Using Equation 5-1, we could write the association or binding constant, K as:

$$K = \frac{f}{(1-f)\cdot B_f} \tag{5-2}$$

where f is the fraction of sites on the enzyme that are occupied by the ligand B, and B_f is the concentration of free ligand. This equation assumes that all binding sites on the enzyme are equivalent. We can also write down the total concentration of the ligand B_T as:

$$B_T = B_f + n\cdot f\cdot A_T \tag{5-3}$$

for n, the number of binding sites or binding stoichiometry, and the total concentration of the macromolecule, A_T. Combining these equations we can solve for f.

$$f^2 - f\left[1 + \frac{B_T}{nA_T} + \frac{1}{nKA_T}\right] + \frac{K}{nA_T} = 0 \tag{5-4}$$

Additionally, we can write an equation for the total heat released into the solution (Q) in the sample cell volume V_0 in terms of the change in enthalpy due to binding (ΔH).

$$Q = n \cdot f \cdot A_T \cdot V_0 \cdot \Delta H \tag{5-5}$$

Solving the quadratic in Equation 5-4 for f and substituting this into Equation 5-5 results in Equation 5-6.

$$Q = \frac{1}{2}\left(nA_T V_o \Delta H\right)\left[\left(1 + \frac{B_T}{nA_T} + \frac{1}{nKA_T}\right) \pm \sqrt{\left(1 + \frac{B_T}{nA_T} + \frac{1}{nKA_T}\right)^2 - \frac{4B_T}{nA_T}}\right] \tag{5-6}$$

Using Equation 5-6 and given a binding stoichiometry (n), an association constant (K) and a change in enthalpy of binding (ΔH), we can calculate a heat content of the solution. A nonlinear least squares fit can produce a fit of Equation 5-6 to the data. Uncertainties in these values can be found using the bootstrapping method.

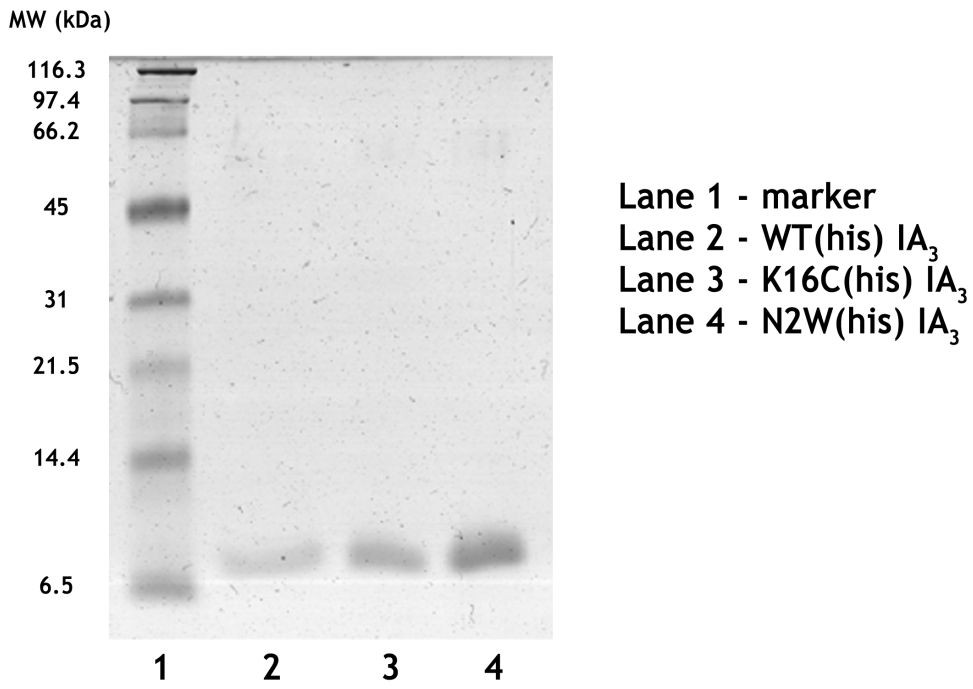
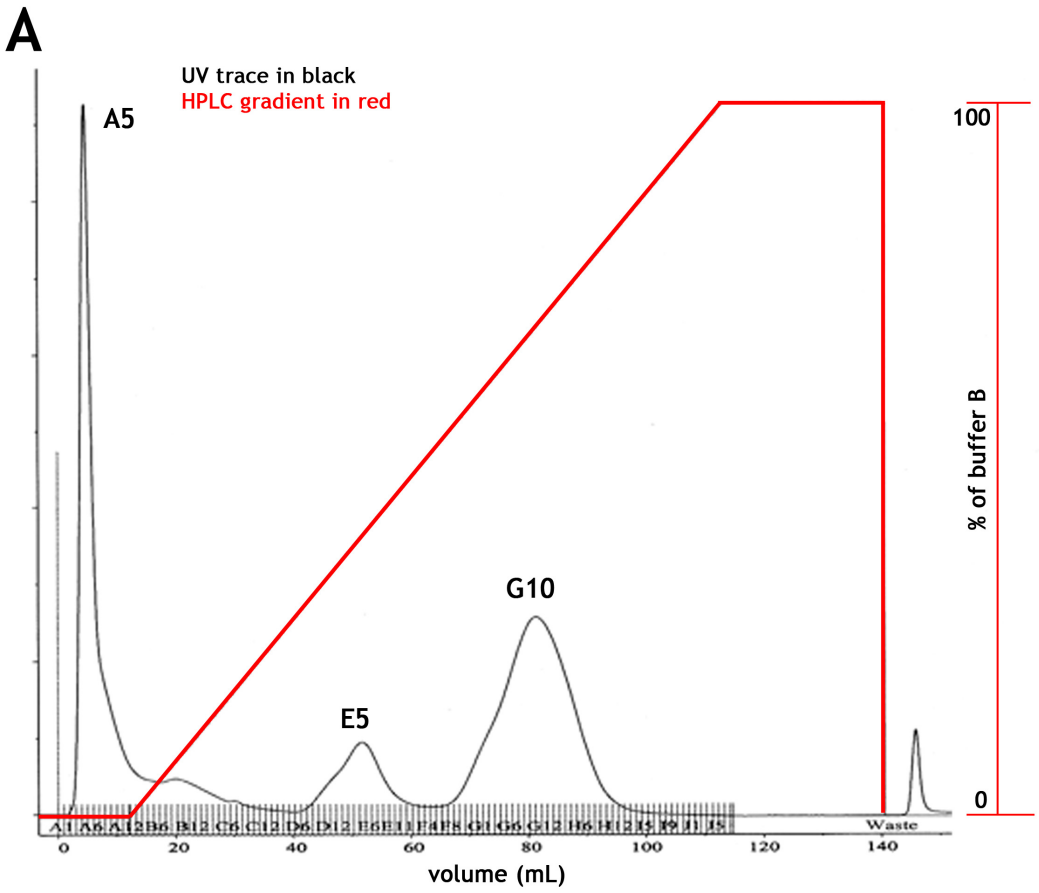


Figure 5-1: Representative 10% SDS-PAGE protein gel for several of the histidine tagged IA_3 proteins indicating similar molecular weights.



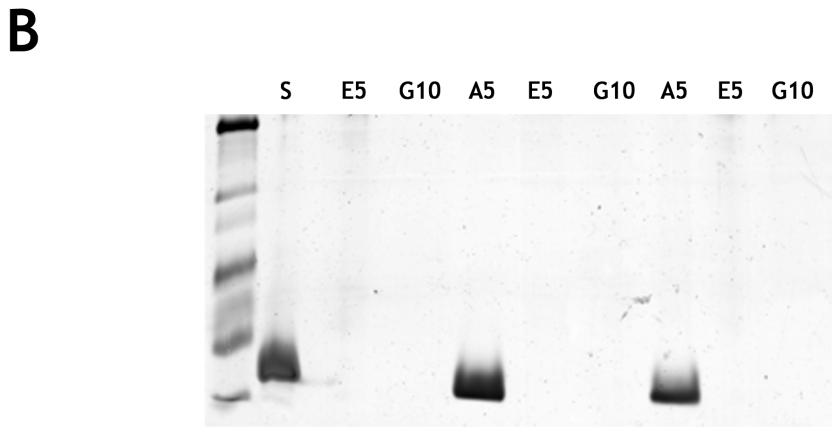


Figure 5-2: Anion exchange chromatography of WT IA_3 (without histidine tag). A) Results of purification. WT IA_3 is not trapped by the anion exchange column and can be found in fraction A5. B) Gel of fractions from the several purifications, with results similar to panel A. The protein standard is marked as S in the gel.

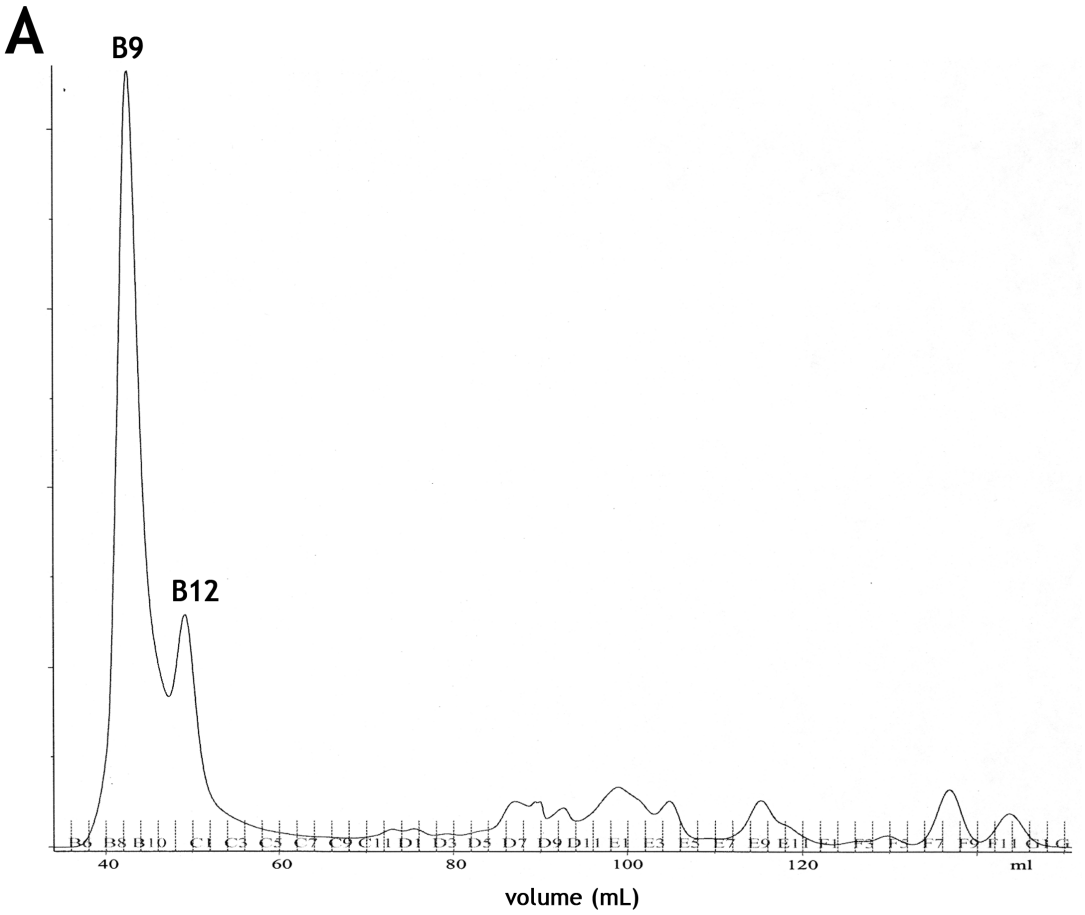


Figure 5-3: Purified IA_3 WT(his) protein elution profile from an S75 column elutes in fraction B12. A) Elution profile; this result is typical for any non-cysteine containing IA_3. B) The peak is confirmed by SDS-PAGE gel analysis. The protein standard is marked as S which matches with fraction B12 as the WT(his) IA_3.

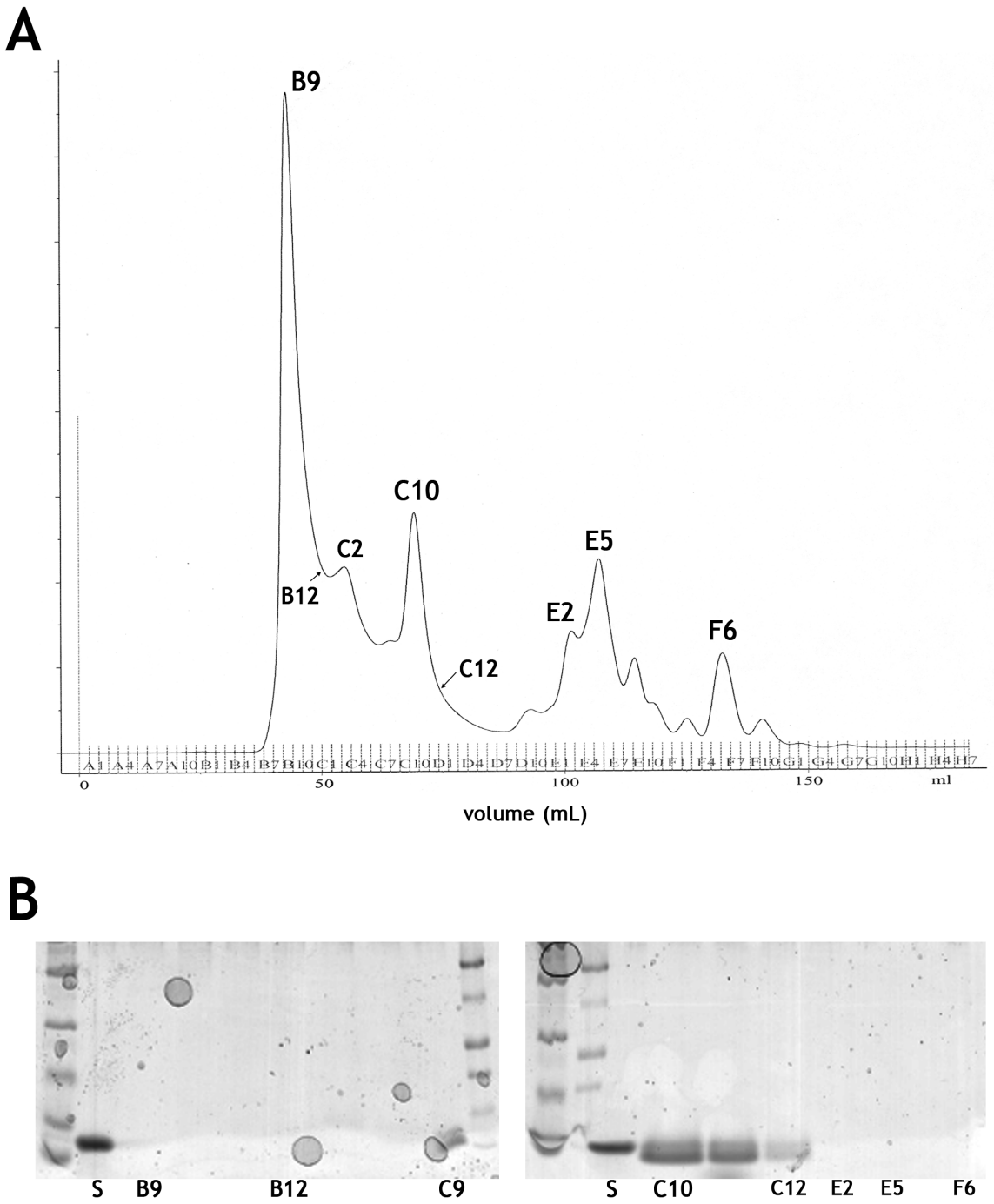
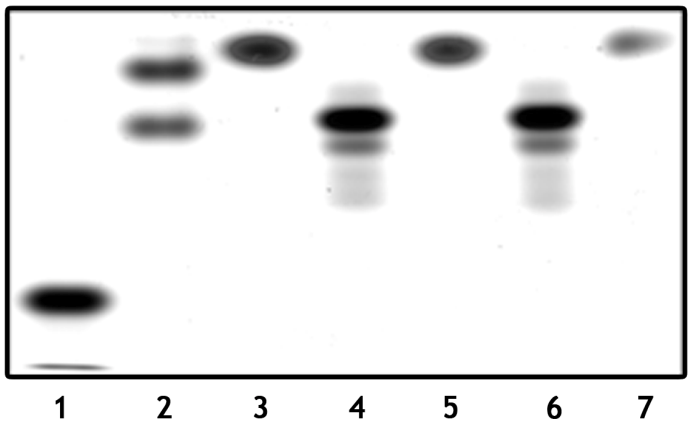


Figure 5-4: Elution profile of a cysteine mutant of IA_3. A) Cysteine mutants display anomalous retention times in an S75 column. Instead of eluting around fraction B12, these mutants elute around fraction C10. B) This is confirmed by SDS-PAGE gel analysis. The protein standard is marked as S.



Lane	Sample
1	α-lactalbumin
2	carbonic anhydrase
3	K24C(his)
4	K16C(his)
5	K24C(his)
6	K16C(his)
7	WT(his)

Figure 5-5: A 7% native gel of several IA_3 proteins illustrating the differences between cysteine mutants and WT(his) IA_3.

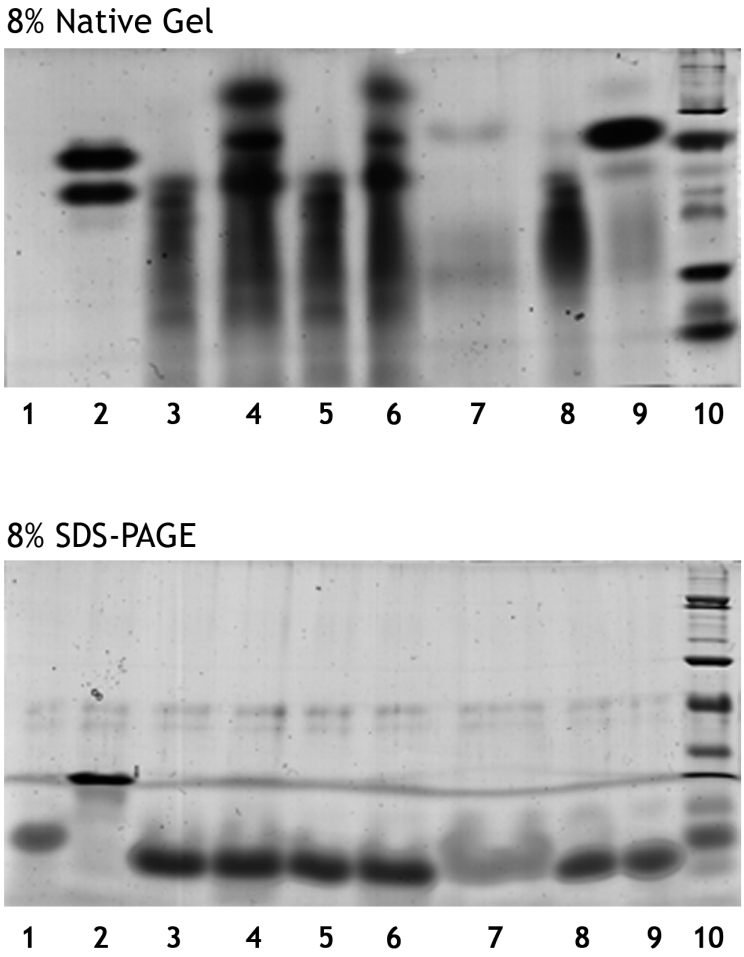


Figure 5-6: SDS-PAGE and native gel results for several IA_3 mutants illustrating odd mobility that can be seen in the native gel, while the SDS-PAGE gel shows similarity among the proteins.

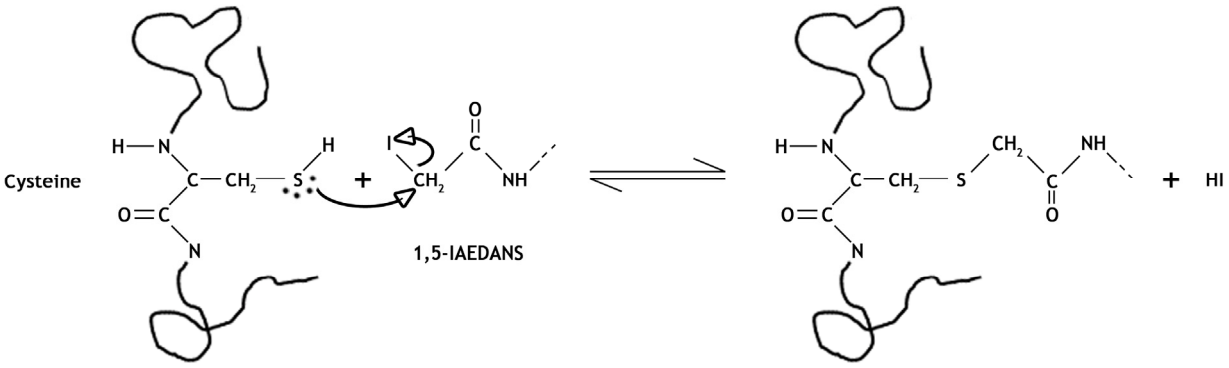


Figure 5-7: Nucleophilic substitution reaction which labels cysteine with a dansyl moiety.

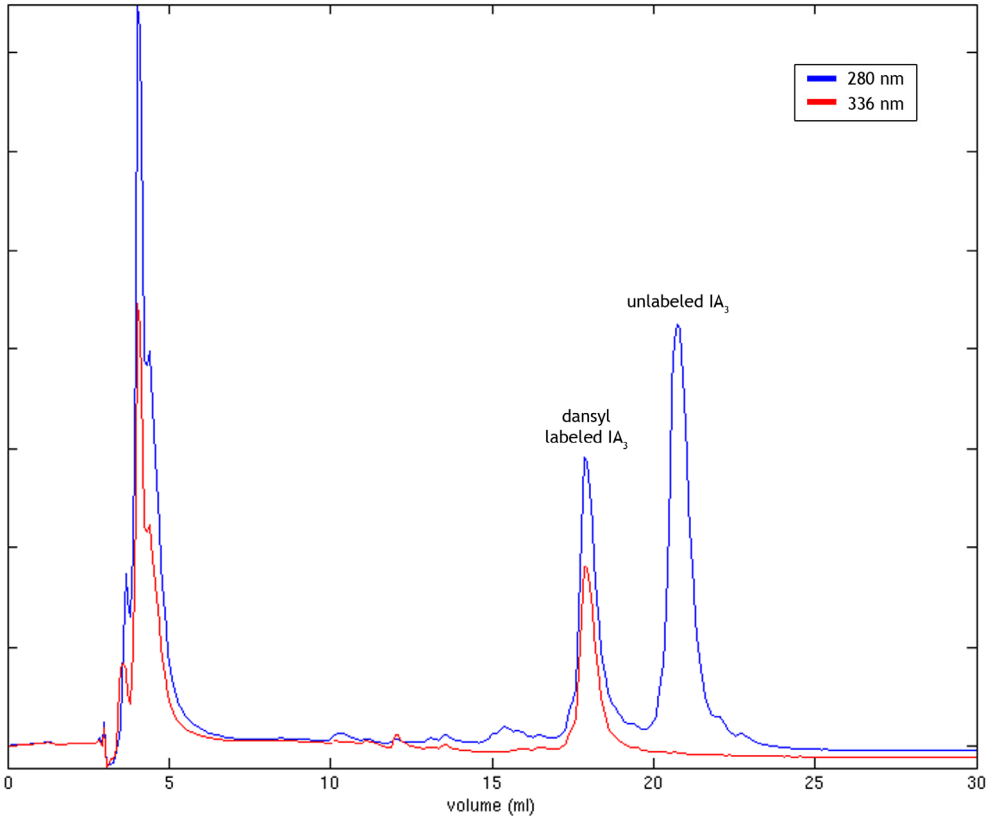


Figure 5-8: Purifying labeled IA_3 using HPLC over a C18 reverse phase column. The blue line is absorbance at 280 nm, the red line is absorbance at 337 nm for dansyl. The labeled protein elutes at 18 minutes, followed by the unlabeled protein. Elution peaks at 5 minutes are the unreacted components in solution or salts.

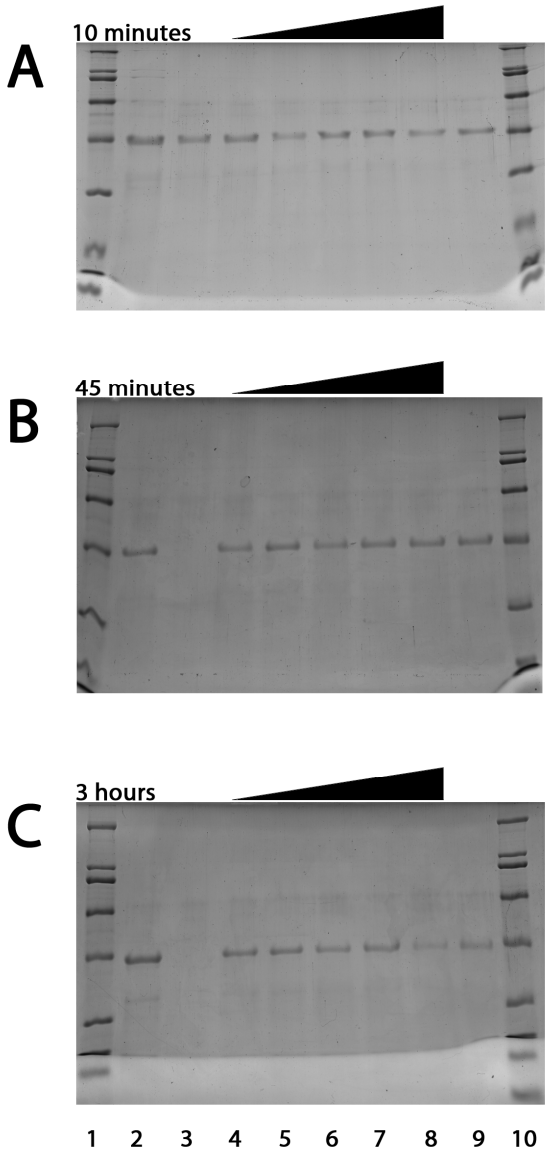


Figure 5-9: Gel analysis of YPrA temperature stability. YPrA was allowed to equilibrate for different amounts of time at different temperatures. The black ramp above each gel indicates the increase in temperature. A) A marker is in lanes 1 and 10, initial YPrA (before temperature equilibration) is in lanes 2, 3 and 9, lanes 4 through 8 have been allowed to equilibrate for 10 minutes at 4°C, 15°C, 25°C, 35°C, and 55°C respectively. B) A marker is in lanes 1 and 10, initial YPrA (before temperature equilibration) is in lanes 2 and 9, lane 3 is empty, lanes 4 through 8 have been allowed to equilibrate for 45 minutes at 4°C, 15°C, 25°C, 35°C, and 55°C respectively. C) A marker is in lanes 1 and 10, initial YPrA (before temperature equilibration) is in lanes 2 and 9, lane 3 is empty, lanes 4 through 8 have been allowed to equilibrate for 3 hours at 4°C, 15°C, 25°C, 35°C, and 55°C respectively.

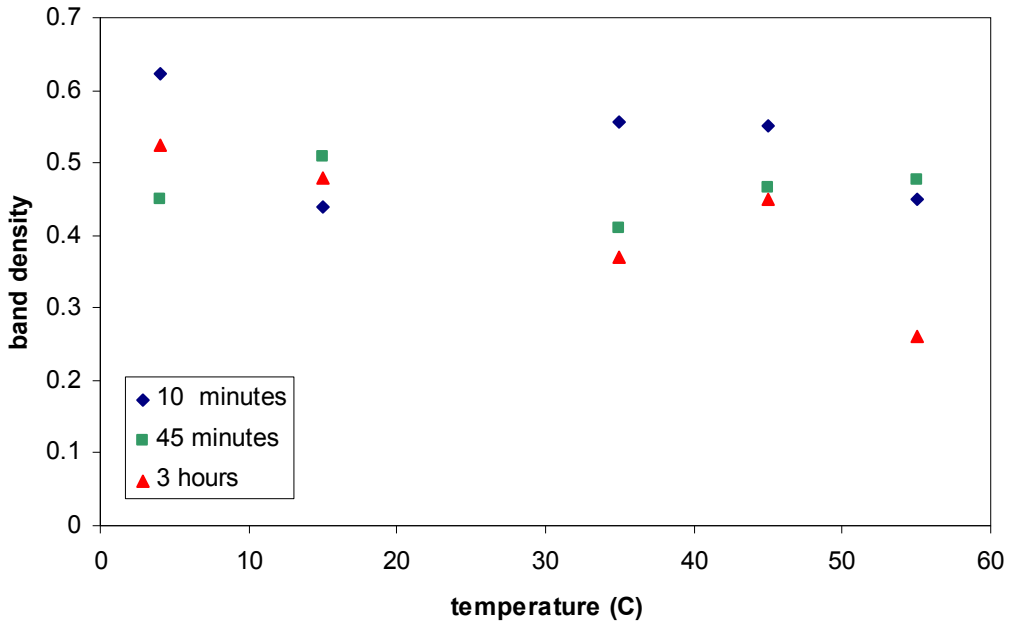


Figure 5-10: Densitometry results of the gels in Figure 5-9.

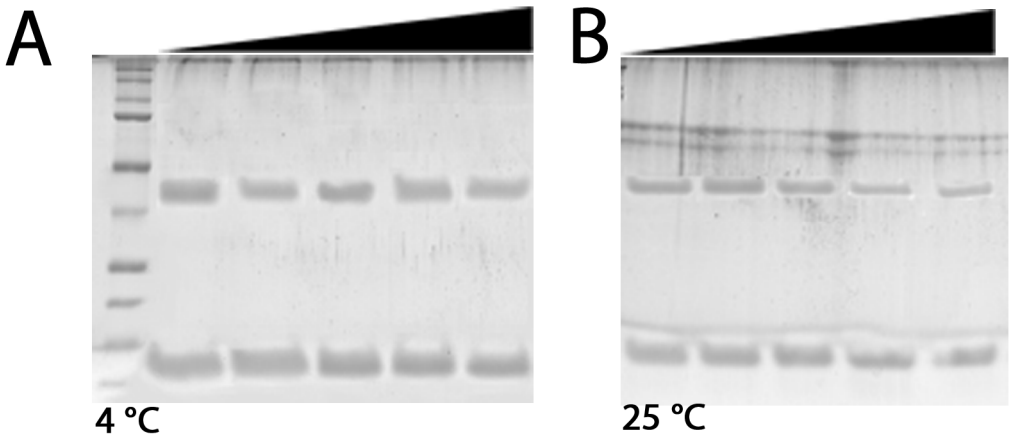


Figure 5-11: YPrA·IA_3 complex stability. Complexes of YPrA and IA_3 were allowed to equilibrate at two different temperatures for increasing amounts of time. Both panels show gels of the YPrA·IA_3 complex, with panel A at 4°C, while panel B was at 25°C. The complexes were allowed to equilibrate for 0, 30 minutes, 1 hour, 2 hours and 3 hours as denoted by the black wedge above the appropriate lanes of each gel.

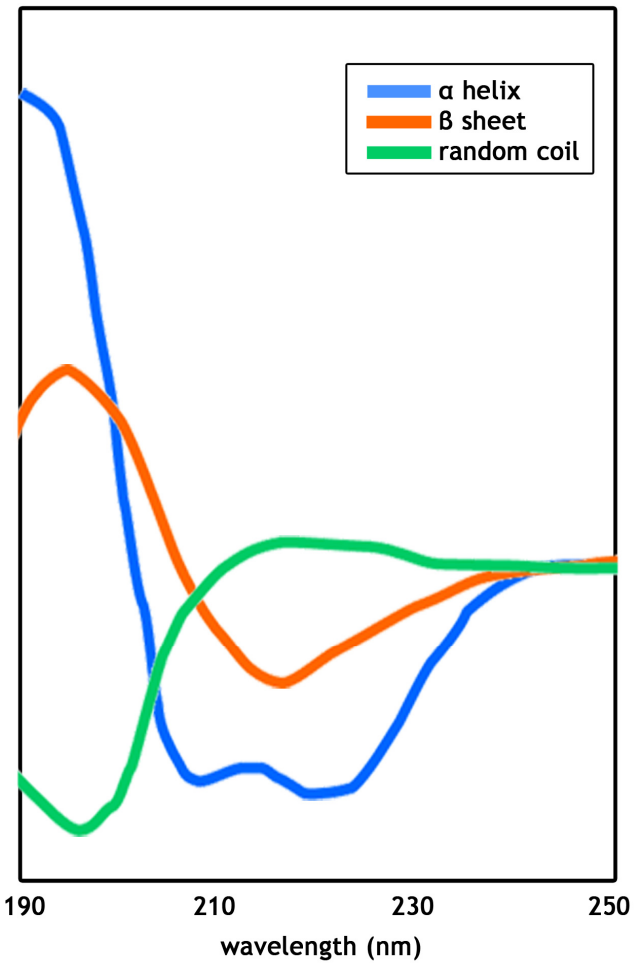


Figure 5-12: Typical CD spectra for various secondary structures.

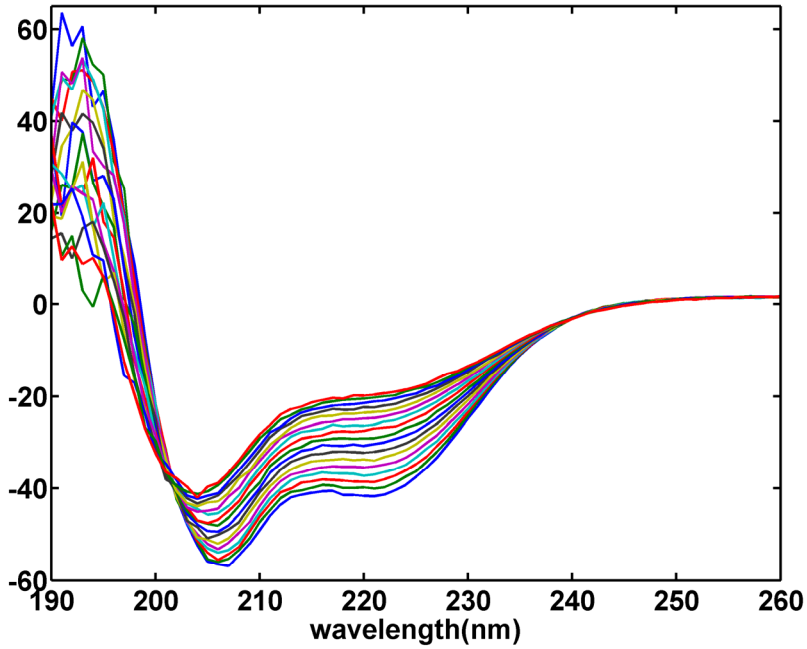


Figure 5-13: CD wavelength spectra of WT(his) IA_3 in 15% TFE at temperatures ranging from 5°C to 85°C.

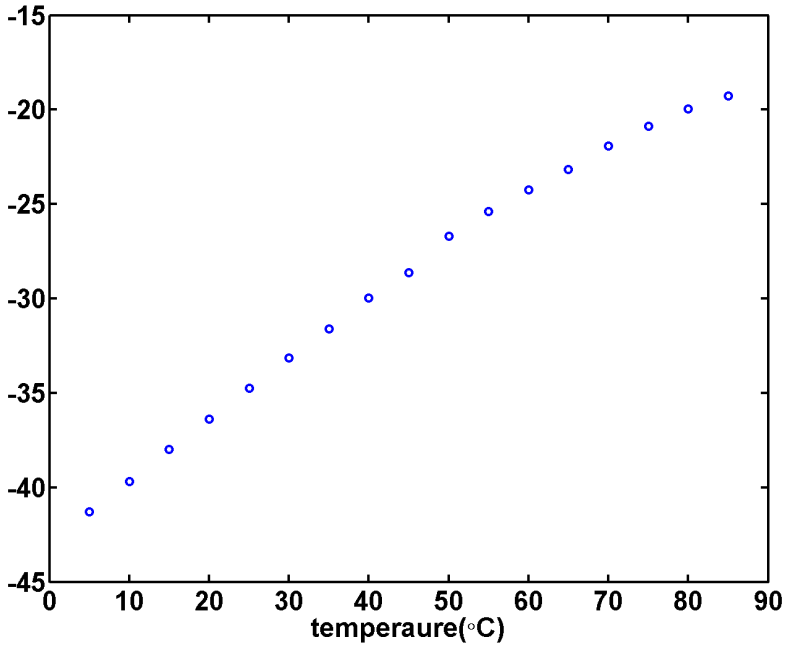


Figure 5-14: CD temperature scan at a wavelength of 222 nm of WT(his) IA_3 in 15% TFE indicating the change in CD signal as the temperature increases.

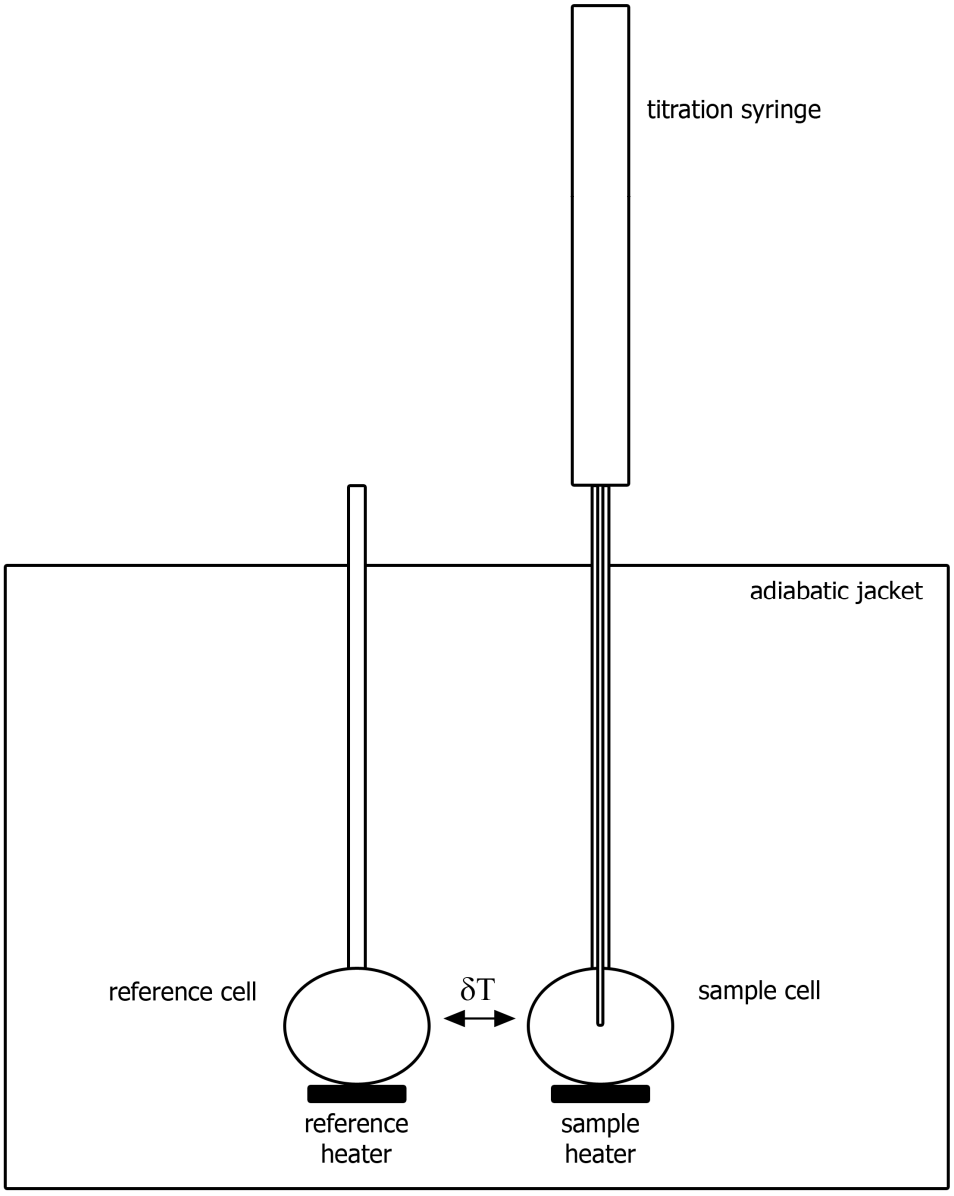


Figure 5-15: An ITC instrument.

CHAPTER 6
FUTURE DIRECTIONS AND CONCLUSIONS

Future Directions

The investigation of the folding and binding interactions of IA_3 with YPrA is not complete. This is mainly due to several interesting questions and observations that were raised during the course of my study which point toward other experiments. Time is always a factor in research. Satisfactorily resolving the experiments on hand always take precedence over pursuing new experimental studies. This does not stifle research, but rather makes an interested scientist yearn for more hours in a day, and more days in a week. This section on Future Directions is a short listing of several experiments which would shed more light on the interaction of IA_3 and YPrA, pushing our understanding of intrinsically disordered proteins forward.

N-terminal IA_3 Kinetics

In chapter 3 laser induced temperature jump spectroscopy highlighted a rate enhancement in the folding of IA_3 in the presence of YPrA. These studies logically lead to an interesting question. Does the C-terminus of the inhibitor cause this rate enhancement? Construction of a truncation mutant containing only the N-terminus and the N2W and dansylated K16C mutation would be necessary to investigate this question. A lack of rate enhancement from this N-terminal construct would provide direct evidence of the interaction of the C-terminus with YPrA, and an explanation of its function.

Calorimetry

Repetition of ITC experiments is needed to verify the initial ITC results in chapter 4. However these results indicated that the C-terminus of the inhibitor is interacting with the protease. Good measurements of the association constant were impossible since IA_3 has a high affinity for YPrA with a subnanomolar inhibition constant. Several authors have now shown that

association constants for tight binding inhibitors can be measured. A possible experiment would consist of the strong binding inhibitor IA_3 being titrated into YPrA which is saturated with a weak binding inhibitor. Weak binding inhibitors could be one of several weak binding IA_3 mutants or a small molecule inhibitor.

DSC experiments could provide more information about the energetics as well. Melting point temperatures of IA_3 in the presence of TFE would be useful in further verifying $TFE_{midpoint}$ values found in chapters 2 and 3. Additionally, melting curves of YPrA and various IA_3 mutants may provide clues as to the C-terminal interactions of IA_3 to YPrA.

Inactive YPrA

In our hands, we have never been successfully able to produce active, mature YPrA. While being able to make YPrA would be a boon both in terms of YPrA supply and cost, another interesting advantage would be in making catalytically inactive YPrA. Mutation of the catalytic aspartates (asp residues 32 and 215) within the active site of the protease would allow for surface interaction studies of peptides/proteins with YPrA without the fear of proteolytic degradation by YPrA. Examination of a truncation mutant containing only the C-terminus of IA_3 and its interaction with YPrA could provide complementary evidence to the studies undertaken in this dissertation. Additionally, other experimental studies of the interaction between catalytically inactive YPrA and IA_3 may provide more insight into the binding mechanism of the inhibitor.

C-terminal EPR Spectroscopy

Regions with low electron density in the crystal structure of the YPrA·IA_3 complex can be interpreted as dynamic regions or mobile regions. The C-terminus of IA_3 is not seen in the crystal structure, and additionally is not necessary for inhibition of the protease. However, we are proposing that the C-terminus interacts via non-specific charge interactions, weakly with the surface of YPrA. EPR experiments of the WT IA_3 with YPrA could provide more evidence of

the interaction of the C-terminus of the inhibitor with the protease. EPR labels systematically placed in each residue in the C-terminus of IA_3 may highlight which residues are nonspecifically interacting with YPrA. Mutating these residues may lead to a means of modulating the folding rate enhancement seen in chapter 3.

Conclusions

In this dissertation I have detailed several experiments to understand the folding and binding processes involved in IA_3 and its interaction with YPrA. Using IA_3 as a model system, I have investigated the benefits of intrinsic disorder. The work in this dissertation has focused on intrinsic disorder and the IA_3·YPrA system, and is summarized in this section.

NMR was used in chapter 2 to understand the residue level movements of the amino acids in IA_3 and their contribution to a folded IA_3 structure as a function of TFE concentration. This chapter highlighted that while a structure derived from NMR data would be useful, it is not needed to assess the residue movements of the protein in TFE. This was advantageous since solving a structure from NMR data at each TFE concentration would be time consuming. Instead relatively quick HSQC experiments were combined with a few 3D NMR experiments at interspersed TFE concentrations to derive a picture of the relative movements of the residues as a function of TFE.

Additionally, SVD was used in chapter 2 to emphasize data arising from the residue movements of the inhibitor above the noise in the data, and investigate how the folding transition varied with TFE and in which residues was this transition most prominent. Datasets were decomposed using SVD into two datasets, one which contained the relative residue movements and one which contained the TFE dependence.

Analysis of the residue movements indicated that while residues throughout the sequence of IA_3 showed an increased helical propensity with the addition of TFE, residues in the N-

terminus of the protein exhibited a greater transition towards an alpha-helical structure than residues in the C-terminus of the protein. The residues which demonstrated the greatest transition to an alpha-helical structure were also the same residues of IA_3 found in the crystal structure of the YPrA·IA_3 complex. This suggests that while the whole protein may be affected by the unstructured to structured transition, residues in the inhibitory N-terminus of the protein show more of this transition as a function of TFE concentration.

SVD analysis of the data was surprising in that three components of TFE dependence were identified in our data. Two SVD components were well fit by a two-state model of folding, which could be represented as an unfolded population of molecules transitioning to a folded population of molecules as the TFE concentration is raised. However, a weakly populated third component was present in the data as well, suggesting that a two-state description of the folding of IA_3 may not be entirely complete.

Taken together these data suggest that while all the residues in the protein are undergoing largely two-state transition, residues in the N-terminus of the protein exhibit this transition to a larger degree. Considering that the C-terminus of the protein is not needed for inhibition of YPrA, and that it is not found in the crystal structure of the IA_3·YPrA complex, it is possible that the third component may be related to the C-terminus of the inhibitor and its interactions with YPrA.

In chapter 3, the kinetics of folding of IA_3 was examined in TFE or with YPrA using FRET and laser induced temperature jump spectroscopy. The addition of fluorescence to the inhibitor, both by mutation and fluorescence labeling, provided a convenient way to assess folding, by taking advantage of the distance based resonance energy transfer using FRET.

Temperature jump spectroscopy was used to investigate the kinetics of the folding/unfolding transition.

We have so far assessed the folding rate constants of IA_3 in water for various temperatures. This is an important step in elucidating the folding/binding process IA_3 goes through in the presence of YPrA. Preliminary results indicate that the folding of the N-terminus of the protein is enhanced by the presence of YPrA. This suggests that the interaction of IA_3 with YPrA affects folding of the portion of IA_3 which is necessary for inhibition of the protease. This may be an indicator that the unstructured C-terminus of the protein is interacting with YPrA to promote folding of the N-terminus of IA_3. Examination to see if this effect is present in a truncated IA_3 containing only the N-terminus of the protein is necessary, however beyond the scope of this dissertation chapter. This was elaborated upon in the Future Directions section of this chapter.

Energetics of the interaction of interaction of IA_3 with YPrA was investigated in chapter 4 using ITC. ITC experiments allowed us to investigate the interaction of YPrA with several constructs of IA_3: an N-terminus only construct, WT IA_3, and WT IA_3 containing a C-terminal his6 addition. These IA_3 proteins differ only in their C-termini which display different charges at the experimental pH of 4.5. Non-specific interactions between the C-termini of these molecules and YPrA were deduced using ITC.

These results indicate that the C-terminus of the IA_3 molecule is indeed interacting with the surface of YPrA. The action of the unstructured C-terminus of IA_3 may be to enhance the meeting of the inhibitor N-terminus of IA_3 with partner residues in the flaps of YPrA. Weak, non-specific charge interactions between the C-terminus of IA_3 may not be lasting enough to be seen in the crystal structure of the complex, but are effectively raising the local concentration of

the N-terminus around the active site of YPrA. This work also provides evidence of the fly-casting effect due to the unstructured C-terminus of IA_3.

This dissertation has looked at the folding and binding of IA_3. IA_3 provides a unique model to investigate and understand the benefits of intrinsic disorder within a protein. Several biophysical techniques and analysis were used to assess the folding of IA_3 in TFE, or the coupled folding and binding of IA_3 in the presence of YPrA. The results of this dissertation indicate that this small 68 amino acid protein displays a rich and complex set of responses. The understood interaction between IA_3 and YPrA is inhibition due to binding of the N-terminus into YPrA, and intrinsic disorder in the inhibitor appears to enhance this inhibitory action. Understanding of the benefits of intrinsic disorder in IA_3 could lead to not only a general understanding of intrinsically disordered proteins, but the development of novel inhibitory proteins/peptides which are able to exploit disordered regions to enhance their inhibition.

APPENDIX
PROGRAMMING CODE

Why Programming?

My research has consisted of a fair amount of programming in a variety of scripts and languages. This is useful since it provides a way to more fully understand the mechanics of data analysis. Additionally, it leads away from a dependence on supplied programs which inevitably can act like black boxes for research (where data is stuffed into one end of the black box and numbers sprout out the other - leading to a low understanding of how the analysis was done).

It can be argued that this is akin to reinventing the wheel for some analyses. While this may be a fact, it is probably true that researchers on the cutting edge of science should have a strong grasp of any analysis of their data. Errors or inconsistencies may be missed or misinterpreted if there is a dependence on analysis software that is incorrectly programmed. One would hope that a software package to analyze data would always work correctly. This, sadly, is not a rule and can be seen in the various iterations a piece of software goes through over time.

Two main advantages of programming to a researcher are the ability to do complicated analyses, and to aid in repetitive computer tasks. Much data analysis can be done (and is often done) using Excel. While this provides a convenient way to quickly analyze data, it is also confining in the types of analysis that can be done. Complicated fitting, analysis, or advanced mathematical procedures are beyond its capabilities. While it is an excellent tool for the little that it does, reliance on Excel (or Excel like spreadsheet programs) certainly limits data analysis. The other advantage, programming to lower repetitive actions, is strictly a time saver. Spending an hour to code together a program to end a repetitive set of actions is time well spent. This is especially true in a field of study where time spent on repetitive actions means time away from other productive tasks such as data gathering or analysis.

One Program to Rule Them All?

A common problem when deciding to start programming is which language/script should be used? The answer to this question unfortunately varies depending on the situation and the problem needing to be answered. More often than not, however, there are multiple choices available in terms of programming languages/scripts to solve a particular problem. This does mean that a researcher needs to be flexible enough to learn and work with many different languages. There is a benefit in that there is a similarity of logic between languages, and with each new language/script that is learned, it becomes easier to learn the next one.

The scripts I have listed in this appendix are written in a variety of programs. Each one is particular to its situation. While there are other, and arguably better, ways (use of different languages, better coding style, coding construction) to achieve the same results, these programs served me well when I needed them. I have done my best to comment the code appropriately and write code which could be followed easily, however there are some bits which were added ad hoc. For unintelligible code I apologize. Here I have included several scripts/programs which I found very useful.

The Whole Shebang

bash Scripting

NMR indirect referencing

This bash script was used to indirectly reference the NMR values used in chapter 2.

```
#!/bin/bash

#
# indirect referencing
# a shell script (portable to any platform with standard bash and bc)
# okg 8/23/05
#

### USER CONFIGURABLE OPTIONS
#
# How much precision (digits) do you want
```

```
scale=12
#
# Chemical Shift Ratios
# current ratios can be found at:
# http://beta.bmrb.wisc.edu/ref_info/cshift.html
X1H=1.00           # 1H (DSS)
X15N=0.101329118   # 15N-1H (DSS)
X13C=0.251449530   # 13C-1H (DSS)
#
###### NO MORE CHANGES BELOW THIS LINE

### FUNCTIONS
# calc_freq() calculates the frequency based on
# Chemical Shift Ratio
# 1H DSS signal (in MHz)
# Nucleus signal (in MHz)
calc_freq()
{
XCSR=$1     # chemical shift ratio of interest
DSS=$2      # MHz value of DSS
Nuc=$3  # MHz value of Nucleus of interest
zeroNuc=`echo "${DSS}*${XCSR}" | bc`
#echo "scale=${scale} ; ((${Nuc}-${zeroNuc})*1000000)/${Nuc}"
NucPPM=`echo "scale=${scale} ; ((${Nuc}-${zeroNuc})*1000000)/${Nuc}" | bc`
echo $NucPPM
}
# exit_check() checks to see if input is null and exits if it is
exit_check()
{
if [ -z $1 ]
 then
   exit 0
fi
}
############

### CASE STATEMENTS
base=${0##*/}    # get the prgram basename
#
case $1 in
   help)  # need help?
echo "
 Returns frequencies in ppm from:
     1H frequency at DSS in MHz
     1H carrier frequency in MHz (sfo1)
     15N and 13C carrier frequencies (sfo2/3)
 usage:
     $base [option]
 options:
     with no option present, $base will function normally
     equations - equations for indirect referencing
     license   - display the license and quit
     help      - show this help then quit
"
exit 0;;   # exit cleanly
   license | lic )   # license information
echo "
```

```
This work was based on \"reference.exe\" -
a FORTRAN program by Dr. A. S. Edison.

\"${base}\" is licensed by O. Ganesh under the
Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License
http://creativecommons.org/licenses/by-nc-sa/2.5/legalcode
any derivative works based on this work shall reference the author
"
exit 0;;   # exits cleanly
   equation | equations | eq | eqs)
echo "
 Indirect Chemical Shift Referencing
     DSS  = 1H DSS frequency in MHz
     Nuc  = nucleus frequency in MHz
     Xcsr = Chemical Shift Ratio for nucleus of interest
 calculation of 0ppm freq:
     0ppm = DSS * Xcsr
 indirect referenced frequency (ppm):
     ind_ref = ((Nuc - 0ppm) * 1e6) / Nuc
"
exit 0;;
   *) ;;  # catch all for anything else (currently not used)
esac
###########

### THE PROGRAM (finally!)
echo " use '$base help' for help
 Returns frequencies (ppm) from:
     1H frequency at DSS in MHz
     1H carrier frequency in MHz (sfo1)
     15N and 13C carrier frequencies (sfo2/3)

Enter 1H DSS frequency in MHz: "
read DSS ; exit_check $DSS  # gets 1H-DSS freq and checks it
echo "Enter 1H carrier frequency in MHz: "
read H ; exit_check $H   # gets 1H freq and checks it
HPPM=`calc_freq $X1H $DSS $H` # calculates the 1H referenced frequency
echo "DSS: $DSS  1H: $H
indirect H1(ppm) = $HPPM
"

echo "Enter 15N frequency in MHz: "
read N ; exit_check $N # gets 15N freq and checks it
NPPM=`calc_freq $X15N $DSS $N` # calculates the 15N referenced frequency
echo "DSS: $DSS  15N: $N
indirect N15(ppm) = $NPPM
"

echo "Enter 13C frequency in MHz: "
read C ; exit_check $C # gets 13C freq and checks it
CPPM=`calc_freq $X13C $DSS $C` # calculates the 13C referenced frequency
echo "DSS: $DSS  13C: $C
indirect C13(ppm) = $CPPM
"

exit 0 # exit cleanly
```

Web Scraping Using Ruby (via bash)

Web scraping is the name given to any form of extracting data from a webpage/website over the internet. This is important and useful as scientific information/data is increasingly being moved online (examples of this are PubMed, the Protein Data Bank, or WormBase). Unfortunately, there is no common 'information format' which is readily accessible. Each database may have its own described format which it may or may not adhere to. Because of this, collecting, collating, and correlating large amounts of information for local analysis can become a tedious task.

There are many ways to scrape information from a website, however my tools of choice have become scRUBYt! and Hapricot. Hpricot is excellent at working through structured data such as HTML or XML, while scRUBYt! handles forms on websites well. Ruby is an evolving language and has odd characteristics which have not been ironed out. In this incomplete code snippet I am calling ruby via bash to scrape data from aprs.he.fi. The data are then parsed and put into an array for later work.

```
#!/bin/bash

# output for the user
echo "Gathering the data from aprs.he.fi"

# build ruby code
ruby_call="
require 'rubygems'
require 'scrubyt'
aprs_data = Scrubyt::Extractor.define do
  fetch 'http://aprs.he.fi'
  fill_textfield 'nick', 'me'
  submit
  fetch 'http://aprs.he.fi/?c=raw&call=N6WPV-3&limit=1000'
  tracking_info \"/html/body/span[@class='raw_line']\"
end
aprs_data.to_xml.write(\$stdout, 1)"

# call it out to ruby
# this is not clean because scRUBYt! is still beta
```

```
# hpricot would be better, but can't handle forms like scRUBYt!
raw_data=`echo "$ruby_call" | ruby`

# output for the user
tput cuu1; echo "Parsing the data and building arrays"

# parse out the dates
dates=`echo "$raw_data" | grep "tracking_info" | grep "A=0" | \
      sed "s/ GMT:/>/" | sed "s/A=/>/" | cut -d">" -f2`

# parse out the altitude
alt=`echo "$raw_data" | grep "tracking_info" | grep "A=0" | \
      sed "s/ GMT:/>/" | sed "s/A=/>/" | cut -d">" -f4 | \
      cut -d" " -f1`
# make an altitude array for the output
     altitude=( $(echo "$alt") )
```

Perl Programming

Sequence dependent chemical shifts

The chemical shift of a residue can be influenced by its nearest residue neighbors, up to two neighbors on either side. Calculating the correction factors can be done by hand; however, this is tedious and can be error-prone. This perl script automates the task, by reading in an amino acid sequence and generating the random coil shift values as well as the sequence dependent corrections.

```
#!/usr/bin/perl -w

#
# correction.pl takes amino acid sequence and
# generates the random coil shift values along
# with sequence dependent corrections.
# okg 8/28/05
# to run use 'correction.pl help' in UNIX/Mac/Linux
# or use 'perl correction.pl help' in Windows
#

use strict;
use warnings;
use complete_db;     # load the database module

# Initialize variables
my $primary_seq = ' ';
my (@primary_seqa, $aa, $aaHN, $aaHA, $aaNH, $aaCA, $aaCO,
$previi, $previiHN, $previiHA, $previiNH, $previiCA, $previiCO,
$previ, $previHN, $previHA, $previNH, $previCA, $previCO,
$nexti, $nextiHN, $nextiHA, $nextiNH, $nextiCA, $nextiCO,
```

```
$nextii, $nextiiHN, $nextiiHA, $nextiiNH, $nextiiCA, $nextiiCO,
$sumHN, $sumHA, $sumNH, $sumCA, $sumCO);
my $proteinfilename=$ARGV[0]; my $outputfile=$ARGV[1];

# help
sub show_help {
print STDERR <<EOF;
takes an amino acid sequence and calculates the
random coil chemical shifts and sequence dependent corrections
usage:
 UNIX, Mac, Linux
   $0 input_filename output_filename
 Windows
   perl $0 input_filename output_filename
options are:
   help    - shows this help then exits
   license - shows the license then exits
EOF
exit;
}

# license
sub show_license {
print STDERR <<EOF;

This work is licensed by O. Ganesh under the
Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License
http://creativecommons.org/licenses/by-nc-sa/2.5/legalcode
any derivative works based on this work shall reference the author

EOF
exit;
}

# catching the help or license options
# (Aargh!  PERL needs strong case statements!)
if ($#ARGV == 0) {
    if ($proteinfilename eq "help" || $proteinfilename eq "info") {
        show_help;
    } elsif ($proteinfilename eq "license" || $proteinfilename eq "lic") {
        show_license;
    }
}
if ($#ARGV < 1) {
    show_help;
}

# open the inputfile for reading, or exit
unless ( open(PROTEINFILE, $proteinfilename) ) {
    print "Cannot open file \"$proteinfilename\".\n";
    print "(maybe it's mistyped?)\n";
    exit;
}

# open the outputfile for writing, or exit
unless ( open(OUTFILE, ">$outputfile") ) {
    print "Cannot open file \"$outputfile\" to write to.\n";
```

```
    print "(maybe it's  write-protected?)\n";
    exit;
}

# initial titles for the output file
print OUTFILE "AA Atom RC_val seq_dep_i-2 seq_dep_i-1 ";
print OUTFILE "seq_dep_i+1 seq_dep_i+2 sum_of_all_corrections\n";

# read in the sequence from the protein file
open(PROTEINFILE, $proteinfilename);
@primary_seqa = <PROTEINFILE>;
$primary_seq = join( '', @primary_seqa);
# Remove whitespace, clean up the input sequence, upper-case it
$primary_seq =~ s/\s//g;
chomp $primary_seq;
$primary_seq = uc $primary_seq;

# get the sequence length
my $seq_length= length($primary_seq);

# send the sequence to the screen for the user
print "$primary_seq\nsequence length is $seq_length residues\n";

for(my $i=0; $i < $seq_length ; $i += 1) {

    # take care of the random coil values
    $aa = substr($primary_seq,$i,1);
    $aaHN =  randcoilvals($aa . 'HN');
    $aaHA =  randcoilvals($aa . 'HA');
    $aaNH =  randcoilvals($aa . 'NH');
    $aaCA =  randcoilvals($aa . 'CA');
    $aaCO =  randcoilvals($aa . 'CO');

    # take care of the i-2 sequence dependence
    if ( $i > 1 ) {
       $previi = substr($primary_seq,$i-2,1);
    } else {
       $previi = '-';
    }
    $previiHN =seq_dep_previi($previi . 'HN');
    $previiHA =seq_dep_previi($previi . 'HA');
    $previiNH =seq_dep_previi($previi . 'NH');
    $previiCA =seq_dep_previi($previi . 'CA');
    $previiCO =seq_dep_previi($previi . 'CO');

    # take care of the i-1 sequence dependence
    if ( $i > 0 ) {
       $previ = substr($primary_seq,$i-1,1);
    } else {
       $previ = '-';
    }
    $previHN =seq_dep_previ($previ . 'HN');
    $previHA =seq_dep_previ($previ . 'HA');
    $previNH =seq_dep_previ($previ . 'NH');
    $previCA =seq_dep_previ($previ . 'CA');
    $previCO =seq_dep_previ($previ . 'CO');
```

```
    # take care of the i+1 sequence dependence
    if ( $i + 2 > $seq_length ) {
       $nexti = '-';
    } else {
       $nexti  = substr($primary_seq,$i+1,1);
    }
    $nextiHN =seq_dep_nexti($nexti . 'HN');
    $nextiHA =seq_dep_nexti($nexti . 'HA');
    $nextiNH =seq_dep_nexti($nexti . 'NH');
    $nextiCA =seq_dep_nexti($nexti . 'CA');
    $nextiCO =seq_dep_nexti($nexti . 'CO');

    # take care of the i+2 sequence dependence
    if ( $i + 3 > $seq_length ) {
       $nextii = '-';
    } else {
       $nextii  = substr($primary_seq,$i+2,1);
    }
    $nextiiHN =seq_dep_nextii($nextii . 'HN');
    $nextiiHA =seq_dep_nextii($nextii . 'HA');
    $nextiiNH =seq_dep_nextii($nextii . 'NH');
    $nextiiCA =seq_dep_nextii($nextii . 'CA');
    $nextiiCO =seq_dep_nextii($nextii . 'CO');

# sum up all the values
# (RC + sd_i-2 + sd_i-1 + sd_i+1 + sd_i+2)
$sumHN = $aaHN + $previiHN + $previHN + $nextiHN + $nextiiHN ;
$sumHA = $aaHA + $previiHA + $previHA + $nextiHA + $nextiiHA ;
$sumNH = $aaNH + $previiNH + $previNH + $nextiNH + $nextiiNH ;
$sumCA = $aaCA + $previiCA + $previCA + $nextiCA + $nextiiCA ;
$sumCO = $aaCO + $previiCO + $previCO + $nextiCO + $nextiiCO ;

# and the nice output written to the output file
print OUTFILE "$aa HN $aaHN $previiHN $previHN $nextiHN $nextiiHN $sumHN\n";
print OUTFILE "  CO $aaCO $previiCO $previCO $nextiCO $nextiiCO $sumCO\n";
print OUTFILE "  NH $aaNH $previiNH $previNH $nextiNH $nextiiNH $sumNH\n";
print OUTFILE "  Ca $aaCA $previiCA $previCA $nextiCA $nextiiCA $sumCA\n";
print OUTFILE "  Ha $aaHA $previiHA $previHA $nextiHA $nextiiHA $sumHA\n";

# updating output for the user
print ".";

}

# close the output file
close (OUTFILE);

# final instructions for the user
print "\ncomplete.\nOpen \"$outputfile\" in Gnumeric/OO/Excel (space
delimited) for results\n";
exit;
```

Database of random coil values

The random coil values needed for the previous script are stored here. I decided to keep this database separate from the main program since it would ease in updating these values if they changed.

```
# This is the complete database of random coil values
# and sequence dependent effects on chemical shift values
#
# it is a module for easy updating
# (so no one has to mess with the main program)
#
# To use this module, add the following line (without the comment)
# to your code, making sure the module is in the same directory as
# the program you are using it from (or another place where Perl can find
# it).
#
# use complete_db.pl;

# random coil values
#
# Schwarzinger S, Kroon GJ, Foss TR, Wright PE, Dyson HJ.
# J Biomol NMR. 2000 Sep;18(1):43-8.
#
sub randcoilvals {
    my($aa) = @_;
     my(%aa_val) = (
     "ACA" => "52.8",
     "ACO" => "178.5",
     "AHA" => "4.35",
     "AHN" => "8.35",
     "ANH" => "125",
     "CCA" => "55.6",
     "CCO" => "175.5",
     "CHA" => "4.76",
     "CHN" => "8.54",
     "CNH" => "118.7",
     "DCA" => "53",
     "DCO" => "175.9",
     "DHA" => "4.82",
     "DHN" => "8.56",
     "DNH" => "119.1",
     "ECA" => "56.1",
     "ECO" => "176.8",
     "EHA" => "4.42",
     "EHN" => "8.4",
     "ENH" => "120.2",
     "FCA" => "58.1",
     "FCO" => "176.6",
     "FHA" => "4.65",
     "FHN" => "8.31",
     "FNH" => "120.7",
     "GCA" => "45.4",
```

```
"GCO" => "174.9",
"GHA" => "4.02",
"GHN" => "8.41",
"GNH" => "107.5",
"HCA" => "55.4",
"HCO" => "175.1",
"HHA" => "4.79",
"HHN" => "8.56",
"HNH" => "118.1",
"ICA" => "61.6",
"ICO" => "177.1",
"IHA" => "4.21",
"IHN" => "8.17",
"INH" => "120.4",
"KCA" => "56.7",
"KCO" => "177.4",
"KHA" => "4.36",
"KHN" => "8.36",
"KNH" => "121.6",
"LCA" => "55.5",
"LCO" => "178.2",
"LHA" => "4.38",
"LHN" => "8.28",
"LNH" => "122.4",
"MCA" => "55.8",
"MCO" => "177.1",
"MHA" => "4.52",
"MHN" => "8.42",
"MNH" => "120.3",
"NCA" => "53.3",
"NCO" => "176.1",
"NHA" => "4.79",
"NHN" => "8.51",
"NNH" => "119",
"PCA" => "63.7",
"PCO" => "177.8",
"PHA" => "4.45",
"PHN" => "0",
"PNH" => "0",
"QCA" => "56.2",
"QCO" => "176.8",
"QHA" => "4.38",
"QHN" => "8.44",
"QNH" => "120.5",
"RCA" => "56.5",
"RCO" => "177.1",
"RHA" => "4.38",
"RHN" => "8.39",
"RNH" => "121.2",
"SCA" => "58.7",
"SCO" => "175.4",
"SHA" => "4.51",
"SHN" => "8.43",
"SNH" => "115.5",
"TCA" => "62",
"TCO" => "175.6",
"THA" => "4.43",
```

```
     "THN" => "8.25",
     "TNH" => "112",
     "VCA" => "62.6",
     "VCO" => "177",
     "VHA" => "4.16",
     "VHN" => "8.16",
     "VNH" => "119.3",
     "WCA" => "57.6",
     "WCO" => "177.1",
     "WHA" => "4.7",
     "WHN" => "8.22",
     "WNH" => "122.1",
     "YCA" => "58.3",
     "YCO" => "176.7",
     "YHA" => "4.58",
     "YHN" => "8.26",
     "YNH" => "120.9",
    );
    if(exists $aa_val{$aa}) {
        return $aa_val{$aa};  # looks like everything went well, continue on
    }else{
        # something's not right!
        # instead of causing more problems, it's probably better to stop
        print STDERR "unexpected value: \"$aa\"!!\n";
        exit;
    }
}

# sequence dependent corrections (i-2 amino acid)
#
# Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ.
# J Am Chem Soc. 2001 Apr 4;123(13):2970-8.
#
sub seq_dep_previi {
    my($aa) = @_;
     my(%aa_val) = (
     "ACA" => "-0.02",
     "ACO" => "-0.11",
     "AHA" => "-0.02",
     "AHN" => "-0.01",
     "ANH" => "-0.12",
     "CCA" => "-0.03",
     "CCO" => "-0.08",
     "CHA" => "-0.01",
     "CHN" => "0",
     "CNH" => "-0.06",
     "DCA" => "-0.03",
     "DCO" => "-0.08",
     "DHA" => "-0.02",
     "DHN" => "-0.02",
     "DNH" => "-0.12",
     "ECA" => "-0.01",
     "ECO" => "-0.09",
     "EHA" => "-0.02",
     "EHN" => "-0.01",
     "ENH" => "-0.06",
     "FCA" => "-0.07",
```

```
"FCO" => "-0.27",
"FHA" => "-0.06",
"FHN" => "-0.03",
"FNH" => "-0.18",
"GCA" => "0",
"GCO" => "0",
"GHA" => "0",
"GHN" => "0",
"GNH" => "0",
"HCA" => "-0.05",
"HCO" => "-0.1",
"HHA" => "-0.03",
"HHN" => "-0.01",
"HNH" => "-0.12",
"ICA" => "-0.07",
"ICO" => "-0.2",
"IHA" => "-0.03",
"IHN" => "-0.01",
"INH" => "-0.18",
"KCA" => "-0.01",
"KCO" => "-0.08",
"KHA" => "-0.02",
"KHN" => "0",
"KNH" => "-0.06",
"LCA" => "-0.01",
"LCO" => "-0.13",
"LHA" => "-0.04",
"LHN" => "0",
"LNH" => "-0.06",
"MCA" => "0",
"MCO" => "-0.08",
"MHA" => "-0.02",
"MHN" => "0",
"MNH" => "-0.06",
"NCA" => "-0.06",
"NCO" => "-0.09",
"NHA" => "-0.01",
"NHN" => "-0.01",
"NNH" => "-0.18",
"PCA" => "-0.22",
"PCO" => "-0.47",
"PHA" => "-0.01",
"PHN" => "-0.04",
"PNH" => "-0.18",
"QCA" => "-0.02",
"QCO" => "-0.05",
"QHA" => "-0.01",
"QHN" => "-0.01",
"QNH" => "-0.06",
"RCA" => "0",
"RCO" => "-0.06",
"RHA" => "-0.02",
"RHN" => "0",
"RNH" => "-0.06",
"SCA" => "0",
"SCO" => "-0.08",
"SHA" => "-0.01",
```

```
     "SHN" => "0",
     "SNH" => "-0.06",
     "TCA" => "-0.01",
     "TCO" => "-0.08",
     "THA" => "-0.01",
     "THN" => "0.01",
     "TNH" => "-0.06",
     "VCA" => "-0.07",
     "VCO" => "-0.2",
     "VHA" => "-0.02",
     "VHN" => "-0.01",
     "VNH" => "-0.24",
     "WCA" => "-0.02",
     "WCO" => "-0.26",
     "WHA" => "-0.08",
     "WHN" => "-0.08",
     "WNH" => "0",
     "YCA" => "-0.07",
     "YCO" => "-0.28",
     "YHA" => "-0.05",
     "YHN" => "-0.04",
     "YNH" => "-0.24",
     "-CA" => "0",
     "-CO" => "0",
     "-HA" => "0",
     "-HN" => "0",
     "-NH" => "0",
    );
    if(exists $aa_val{$aa}) {
        return $aa_val{$aa};
    }else{
        # something's not right!
        # instead of causing more problems, it's probably better to stop
        print STDERR "unexpected value: \"$aa\"!!\n";
        exit;
    }
}

# sequence dependent corrections (i-1 amino acid)
#
# Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ.
# J Am Chem Soc. 2001 Apr 4;123(13):2970-8.
#
sub seq_dep_previ {
    my($aa) = @_;
     my(%aa_val) = (
     "ACA" => "-0.17",
     "ACO" => "-0.77",
     "AHA" => "-0.03",
     "AHN" => "-0.05",
     "ANH" => "-0.33",
     "CCA" => "-0.07",
     "CCO" => "-0.51",
     "CHA" => "0.02",
     "CHN" => "-0.02",
     "CNH" => "-0.26",
     "DCA" => "0",
```

```
"DCO" => "-0.58",
"DHA" => "-0.01",
"DHN" => "-0.03",
"DNH" => "-0.2",
"ECA" => "-0.08",
"ECO" => "-0.48",
"EHA" => "-0.02",
"EHN" => "-0.03",
"ENH" => "-0.2",
"FCA" => "-0.23",
"FCO" => "-0.83",
"FHA" => "-0.09",
"FHN" => "-0.12",
"FNH" => "-0.49",
"GCA" => "0",
"GCO" => "0",
"GHA" => "0",
"GHN" => "0",
"GNH" => "0",
"HCA" => "-0.09",
"HCO" => "-0.65",
"HHA" => "-0.06",
"HHN" => "-0.04",
"HNH" => "-0.55",
"ICA" => "-0.2",
"ICO" => "-0.58",
"IHA" => "-0.02",
"IHN" => "-0.06",
"INH" => "-0.14",
"KCA" => "-0.11",
"KCO" => "-0.5",
"KHA" => "-0.02",
"KHN" => "-0.03",
"KNH" => "-0.2",
"LCA" => "-0.1",
"LCO" => "-0.5",
"LHA" => "-0.03",
"LHN" => "-0.03",
"LNH" => "-0.14",
"MCA" => "0.1",
"MCO" => "-0.41",
"MHA" => "-0.01",
"MHN" => "-0.02",
"MNH" => "-0.2",
"NCA" => "-0.03",
"NCO" => "-0.66",
"NHA" => "-0.01",
"NHN" => "-0.03",
"NNH" => "-0.26",
"PCA" => "-2",
"PCO" => "-2.84",
"PHA" => "0.11",
"PHN" => "-0.18",
"PNH" => "-0.32",
"QCA" => "-0.06",
"QCO" => "-0.48",
"QHA" => "-0.02",
```

```
     "QHN" => "-0.02",
     "QNH" => "-0.14",
     "RCA" => "-0.07",
     "RCO" => "-0.49",
     "RHA" => "-0.02",
     "RHN" => "-0.02",
     "RNH" => "-0.14",
     "SCA" => "-0.08",
     "SCO" => "-0.4",
     "SHA" => "0.02",
     "SHN" => "-0.03",
     "SNH" => "-0.03",
     "TCA" => "-0.04",
     "TCO" => "-0.19",
     "THA" => "0.05",
     "THN" => "0",
     "TNH" => "-0.03",
     "VCA" => "-0.21",
     "VCO" => "-0.57",
     "VHA" => "-0.01",
     "VHN" => "-0.05",
     "VNH" => "-0.14",
     "WCA" => "-0.17",
     "WCO" => "-0.85",
     "WHA" => "-0.1",
     "WHN" => "-0.13",
     "WNH" => "-0.26",
     "YCA" => "-0.22",
     "YCO" => "-0.85",
     "YHA" => "-0.1",
     "YHN" => "-0.11",
     "YNH" => "-0.43",
     "-CA" => "0",
     "-CO" => "0",
     "-HA" => "0",
     "-HN" => "0",
     "-NH" => "0",
    );
    if(exists $aa_val{$aa}) {
        return $aa_val{$aa};
    }else{
        # something's not right!
        # instead of causing more problems, it's probably better to stop
        print STDERR "unexpected value: \"$aa\"!!\n";
        exit;
    }
}

# sequence dependent corrections (i+1 amino acid)
#
# Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ.
# J Am Chem Soc. 2001 Apr 4;123(13):2970-8.
#
sub seq_dep_nexti {
    my($aa) = @_;
     my(%aa_val) = (
     "ACA" => "0.06",
```

```
"ACO" => "-0.07",
"AHA" => "-0.03",
"AHN" => "0.07",
"ANH" => "-0.57",
"CCA" => "0.1",
"CCO" => "-0.28",
"CHA" => "0",
"CHN" => "0.2",
"CNH" => "3.07",
"DCA" => "0.25",
"DCO" => "-0.13",
"DHA" => "-0.02",
"DHN" => "0.14",
"DNH" => "0.86",
"ECA" => "0.05",
"ECO" => "-0.2",
"EHA" => "-0.02",
"EHN" => "0.15",
"ENH" => "1.51",
"FCA" => "0.06",
"FCO" => "-0.25",
"FHA" => "-0.08",
"FHN" => "0.1",
"FNH" => "2.78",
"GCA" => "0",
"GCO" => "0",
"GHA" => "0",
"GHN" => "0",
"GNH" => "0",
"HCA" => "0.02",
"HCO" => "-0.22",
"HHA" => "0.01",
"HHN" => "0.2",
"HNH" => "1.68",
"ICA" => "-0.01",
"ICO" => "-0.18",
"IHA" => "-0.02",
"IHN" => "0.17",
"INH" => "4.87",
"KCA" => "-0.02",
"KCO" => "-0.18",
"KHA" => "-0.01",
"KHN" => "0.14",
"KNH" => "1.57",
"LCA" => "0.03",
"LCO" => "-0.13",
"LHA" => "-0.05",
"LHN" => "0.14",
"LNH" => "1.05",
"MCA" => "-0.06",
"MCO" => "-0.18",
"MHA" => "-0.01",
"MHN" => "0.15",
"MNH" => "1.57",
"NCA" => "0.23",
"NCO" => "-0.1",
"NHA" => "-0.02",
```

```
     "NHN" => "0.13",
     "NNH" => "0.87",
     "PCA" => "0.02",
     "PCO" => "-0.09",
     "PHA" => "-0.03",
     "PHN" => "0.19",
     "PNH" => "0.87",
     "QCA" => "0.04",
     "QCO" => "-0.18",
     "QHA" => "-0.01",
     "QHN" => "0.15",
     "QNH" => "1.62",
     "RCA" => "-0.01",
     "RCO" => "-0.19",
     "RHA" => "-0.02",
     "RHN" => "0.15",
     "RNH" => "1.62",
     "SCA" => "0.13",
     "SCO" => "-0.15",
     "SHA" => "0",
     "SHN" => "0.16",
     "SNH" => "2.55",
     "TCA" => "0.12",
     "TCO" => "-0.13",
     "THA" => "0",
     "THN" => "0.14",
     "TNH" => "2.78",
     "VCA" => "-0.02",
     "VCO" => "-0.18",
     "VHA" => "-0.02",
     "VHN" => "0.17",
     "VNH" => "4.34",
     "WCA" => "0.03",
     "WCO" => "-0.3",
     "WHA" => "-0.15",
     "WHN" => "0.04",
     "WNH" => "3.19",
     "YCA" => "0.06",
     "YCO" => "-0.24",
     "YHA" => "-0.08",
     "YHN" => "0.09",
     "YNH" => "3.01",
     "-CA" => "0",
     "-CO" => "0",
     "-HA" => "0",
     "-HN" => "0",
     "-NH" => "0",
    );
    if(exists $aa_val{$aa}) {
        return $aa_val{$aa};
    }else{
        # something's not right!
        # instead of causing more problems, it's probably better to stop
        print STDERR "unexpected value: \"$aa\"!!\n";
        exit;
    }
}
```

```
# sequence dependent corrections (i+2 amino acid)
#
# Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ.
# J Am Chem Soc. 2001 Apr 4;123(13):2970-8.
#
sub seq_dep_nextii {
    my($aa) = @_;
     my(%aa_val) = (
     "ACA" => "0.01",
     "ACO" => "-0.02",
     "AHA" => "0",
     "AHN" => "-0.1",
     "ANH" => "-0.15",
     "CCA" => "-0.01",
     "CCO" => "-0.07",
     "CHA" => "0",
     "CHN" => "-0.07",
     "CNH" => "0",
     "DCA" => "-0.01",
     "DCO" => "-0.04",
     "DHA" => "-0.01",
     "DHN" => "-0.11",
     "DNH" => "-0.29",
     "ECA" => "0.01",
     "ECO" => "-0.03",
     "EHA" => "0",
     "EHN" => "-0.07",
     "ENH" => "-0.12",
     "FCA" => "0.01",
     "FCO" => "-0.1",
     "FHA" => "-0.04",
     "FHN" => "-0.37",
     "FNH" => "-0.46",
     "GCA" => "0",
     "GCO" => "0",
     "GHA" => "0",
     "GHN" => "0",
     "GNH" => "0",
     "HCA" => "0.01",
     "HCO" => "-0.07",
     "HHA" => "0.01",
     "HHN" => "0",
     "HNH" => "0.17",
     "ICA" => "0.02",
     "ICO" => "-0.02",
     "IHA" => "-0.01",
     "IHN" => "-0.09",
     "INH" => "0",
     "KCA" => "0.02",
     "KCO" => "-0.03",
     "KHA" => "0",
     "KHN" => "-0.06",
     "KNH" => "-0.06",
     "LCA" => "0.02",
     "LCO" => "-0.01",
     "LHA" => "-0.01",
```

```
"LHN" => "-0.08",
"LNH" => "-0.06",
"MCA" => "0.01",
"MCO" => "-0.02",
"MHA" => "0",
"MHN" => "-0.06",
"MNH" => "-0.06",
"NCA" => "0.01",
"NCO" => "-0.03",
"NHA" => "-0.01",
"NHN" => "-0.07",
"NNH" => "-0.17",
"PCA" => "0.04",
"PCO" => "-0.02",
"PHA" => "-0.01",
"PHN" => "-0.12",
"PNH" => "-0.17",
"QCA" => "0.01",
"QCO" => "-0.03",
"QHA" => "0",
"QHN" => "-0.06",
"QNH" => "-0.06",
"RCA" => "0.02",
"RCO" => "-0.03",
"RHA" => "0",
"RHN" => "-0.06",
"RNH" => "-0.06",
"SCA" => "0",
"SCO" => "-0.06",
"SHA" => "-0.01",
"SHN" => "-0.08",
"SNH" => "-0.17",
"TCA" => "0",
"TCO" => "-0.05",
"THA" => "-0.01",
"THN" => "-0.06",
"TNH" => "-0.12",
"VCA" => "0.01",
"VCO" => "-0.03",
"VHA" => "-0.01",
"VHN" => "-0.08",
"VNH" => "-0.06",
"WCA" => "-0.08",
"WCO" => "-0.17",
"WHA" => "-0.16",
"WHN" => "-0.62",
"WNH" => "-0.64",
"YCA" => "-0.01",
"YCO" => "-0.13",
"YHA" => "-0.04",
"YHN" => "-0.42",
"YNH" => "-0.52",
"-CA" => "0",
"-CO" => "0",
"-HA" => "0",
"-HN" => "0",
"-NH" => "0",
```

```
    );
    if(exists $aa_val{$aa}) {
        return $aa_val{$aa};
    }else{
        # something's not right!
        # instead of causing more problems, it's probably better to stop
        print STDERR "unexpected value: \"$aa\"!!\n";
        exit;
    }
}

1;
```

Tcl/Tk

NMRView panel placement script

This panel placement script is an amazing time saver. Its purpose is to rearrange the windows of NMRView to accommodate my workflow. This may not seem like much. However, after moving panels around to a desired configuration several times, the benefits of this script become immediately obvious. Tk commands can be run directly from NMRView through the Tk console.

```
#
# panel placement
#
# okg 6/14/05
# because I got sick and tired of moving them every time
#
# you can call winfo from inside nmrview's TkConsole
# to get the geometry settings.   However, winfo will
# only give you the current settings, and will not let
# you reset the window.   Instead you have to call to
# the windowmanager, wm.   It's actually easier to handle
# everything using wm, since it almost uses the same
# calls that winfo does.   To get the window positions
# use:
# wm geometry .widgetName
# To figure out which widget you want, list em:
# winfo children .
# When you are done, set them up with:
# wm geometry .widgetName heightxwidth+xposition+yposition
#
# This Tcl script will load up the settings that I like
# to use it, pop open the panels, then load up the script
# from the TkConsole.   Here's the settings that I like:

# Assign > Peak panel
wm geometry .peakAnal 468x180+1123+30
# Peak Attributes panel
```

```
wm geometry .peakAttr 206x225+1384+241
# View and Cursor menu
# (be careful of tearoff order)
wm geometry .tearoff1 87x172+1187+241
wm geometry .tearoff2 80x110+1289+241
# Attributes panel
wm geometry .specAttr 408x233+1180+497
# Plot panel
wm geometry .plot 451x188+1138+761
# TkCon console
wm geometry .tkcon 80x10+4+994

# tell the user it's done
echo "Panels moved! Have a nice day! -okg"
```

NMRView assignment helper script

This script is another NMRView time saver. When making residue assignments of crosspeaks, several spectra need to be consulted simultaneously in a linked way. This script will try to match that peak in other spectra, including spectra which are along different dimensions.

```
#! /usr/bin/tclsh
#
# auto peak stuff
# helps unify views of CbCa(CO)NH and HNCaCb with HNCO
#
# okg 11/22/05
#
# NMRView TCL Commands can be found at:
# http://www.hhmi.umbc.edu/toolkit/analysis/nmrview3/html/nv_win.html#nv_win
#
# to run this use:
# source assignment_helper_0tfe.tcl
#
#This work ("assignment_helper") is licensed by O. Ganesh under the
#Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License
#http://creativecommons.org/licenses/by-nc-sa/2.5/legalcode
#any derivative works based on this work shall reference the author
#

# make sure the hnco window is active
nv_win act .ia3_notfe_hnco.0
# get the Proton value
set Hline [cross1x]
# get the Nitrogen value
set Nline [cross1y]

# set the HNCaCb as the active window
nv_win act .ia3_notfe_cbcanh.0
# reset the spectra (if needed)
nv_win full
# move the black cursor to top leftmost edge
```

```
# make absolutely sure that the axes for the HNCaCb are:
#    x:HN  y:CACB  z:15N
cross1 [expr $Hline + .13] 17.22
# move the red cursor to bottom rightmost edge
# make absolutely sure that the axes for the CbCa(CO)NH are:
#    x:HN  y:CACB  z:15N
cross2 [expr $Hline - .13] 67.9
# expand the region
nv_win expand
# set the right z plane
z $Nline

# set the CbCaCONH as the active window
nv_win act .ia3_notfe_cbcaconh.0
# reset the spectra (if needed)
nv_win full
# move the black cursor to top leftmost edge
# make absolutely sure that the axes for the HNCaCb are:
#    x:HN  y:CACB  z:15N
cross1 [expr $Hline + .13] 17.22
# move the red cursor to bottom rightmost edge
# make absolutely sure that the axes for the CbCa(CO)NH are:
#    x:HN  y:CACB  z:15N
cross2 [expr $Hline - .13] 67.9
# expand the region
nv_win expand
# set the right z plane
z $Nline

# now recenter all Proton lines
# set the CbCaCONH as the active window
nv_win act .ia3_notfe_cbcaconh.0
# set Proton axis
cross1 $Hline 28.0
# set the HNCaCb as the active window
nv_win act .ia3_notfe_cbcanh.0
# set Proton axis
cross1 $Hline 28.0
# set the HNCO as the active window
nv_win act .ia3_notfe_hnco.0
# set Proton and Nitrogen axis
cross1 $Hline $Nline

# finally, just set the CbCaCONH as the active window
nv_win act .ia3_notfe_cbcaconh.0
```

MATLAB Scripts

Curve fitting

Much of the analysis that I did was using MATLAB. Fitting functions to data is a relatively straight-forward task, however learning how to do this well is slightly mysterious. Here I have listed a small script which details a working exponential fitter. I owe a debt of thanks to Dr. Steve Hagen for sharing his wealth of MATLAB knowledge with me (especially with respect to data fitting).

```
function [ Err ] = linearfit( mb_start, xdata, ydata, drawplot)
%Linear fit to data
%should be used as:
%[m,b]=fminsearch('linear',[1,1],[],time_OD,RT_od)
%good thing to remember, the variable after the function name is ALWAYS the
one that's optimized
%(in this case it's [1,1]=mb_start), if you want to pass other info along,
then you have to use
%[] followed by the data list.
%  To use this function as a plotter, use:
%linear(mb_fit, xdata, ydata,1)
m=mb_start(1);
b=mb_start(2);

if (exist('drawplot') ==1 & exist('mb_start') == 1)
    figure;
    yplot=mb_start(1)*xdata+mb_start(2);
    plot(xdata, ydata, 'o', xdata, yplot, '-');
    title(sprintf('Exp fit Example: y_f_i_t=%.3f*x+%.3f', ...
    m,b));
    xlabel('x')
    ylabel('y')
else

y_fit=m*xdata+b;

Err=norm(ydata-y_fit);
End
```

Bootstrapping code

Bootstrapping can be use to estimate errors in fitting. Data can be bootstrapped in MATLAB directly by calling the bootstrap command. This small example script will take a dataset and resample it, then submit the resampled dataset to a fitter.

```
function [ answers ] = EF_Tu_bootstrapper( input1, input2, input3 )
```

```
% Generalized bootstrapping function
%
% remember, bootstrap works by resampling the data set(s)
% what this will do is minimize the function based on the
% new resampled dataset.
%
% So, in reality what this is doing is finding the best values for dH and
% dS based on a resampled dataset, and using the best values on the full
% set as a starting point.   So, what you get out is an optimized answer
% over a resampled dataset.   This may be a better answer then the original.
% usage:
% rhos1000=bootstrp(1000,'EF_Tu_bootstrapper', temp_s50_, cd_fixed_s50_,
err_t_s50_);
% remember, THIS function works by resampling temp_s50_,
cd_fixed_s50_,err_t_s50_
% and submitting those resampled datasets to be refit using the constants
% as starting points.
%
% If you called the bootstrp function directly on the original function
% then what you are really doing is testing how robust or fragile a certain
% answer is.
% EXAMPLE (not this function):
% rhos100=bootstrp(100,@(SV1,SV2) CD_T([-110 -.1],temps,SV1,SV2),SV1,SV2);
%
% to view your results, use
% hist(rhos1000(:,1),60); % to see the best values for the first

consts=[-89.42 75.32 70.97 2];

temp_s=input1;
cd_fixed_s=input2;
err_t_s=input3;

[consts]=fminsearch('eftu_sig_fit_weighted',consts,[],temp_s,cd_fixed_s,err_t
_s);

%output
answers=[consts(3)];
% the results will be in [midpoint_val]
```

LIST OF REFERENCES

1. Anfinsen, C. (1993) Christian B. Anfinsen, Nobel Lecture, December 11, 1972, in *Nobel Lectures, Chemistry 1971-1980* (Frängsmyr, T., and Forsén, S., Eds.), World Scientific Publishing Co., Singapore.

2. Anfinsen, C. B. (1972) The formation and stabilization of protein structure, *The Biochemical journal 128*, 737-749.

3. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science (New York, N.Y 181*, 223-230.

4. Alberts, B. (2002) *Molecular Biology of the Cell*, 4 ed., Garland Science, New York.

5. Dyson, H. J., and Wright, P. E. (2006) According to current textbooks, a well-defined three-dimensional structure is a prerequisite for the function of a protein. Is this correct?, *IUBMB life 58*, 107-109.

6. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, *Journal of molecular biology 293*, 321-331.

7. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nature reviews 6*, 197-208.

8. Ganesh, O. K., Green, T. B., Edison, A. S., and Hagen, S. J. (2006) Characterizing the residue level folding of the intrinsically unstructured IA3, *Biochemistry 45*, 13585-13596.

9. Green, T. B., Ganesh, O., Perry, K., Smith, L., Phylip, L. H., Logan, T. M., Hagen, S. J., Dunn, B. M., and Edison, A. S. (2004) IA3, an aspartic proteinase inhibitor from Saccharomyces cerevisiae, is intrinsically unstructured in solution, *Biochemistry 43*, 4071-4081.

10. Li, M., Phylip, L. H., Lees, W. E., Winther, J. R., Dunn, B. M., Wlodawer, A., Kay, J., and Gustchina, A. (2000) The aspartic proteinase from Saccharomyces cerevisiae folds its own inhibitor into a helix, *Nature structural biology 7*, 113-117.

11. Phylip, L. H., Lees, W. E., Brownsey, B. G., Bur, D., Dunn, B. M., Winther, J. R., Gustchina, A., Li, M., Copeland, T., Wlodawer, A., and Kay, J. (2001) The potency and specificity of the interaction between the IA3 inhibitor and its target aspartic proteinase from Saccharomyces cerevisiae, *The Journal of biological chemistry 276*, 2023-2030.

12. Ramachandran, G. N., Ramakrishnan, C., and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations, *Journal of molecular biology 7*, 95-99.

13. Ramachandran, G. N., Venkatachalam, C. M., and Krimm, S. (1966) Stereochemical criteria for polypeptide and protein chain conformations. 3. Helical and hydrogen-bonded polypeptide chains, *Biophysical journal 6*, 849-872.

14. Ramakrishnan, C., and Ramachandran, G. N. (1965) Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units, *Biophysical journal 5*, 909-933.

15. Avbelj, F., and Baldwin, R. L. (2003) Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of phi, *Proceedings of the National Academy of Sciences of the United States of America 100*, 5742-5747.

16. Brant, D. A., and Flory, P. J. (1965) The Configuration of Random Polypeptide Chains. II. Theory, *J. Am. Chem. Soc. 87*, 2791-2800.

17. Tanford, C., Kawahara, K., and Lapanje, S. (1966) Proteins in 6-M guanidine hydrochloride. Demonstration of random coil behavior, *The Journal of biological chemistry 241*, 1921-1923.

18. Eker, F., Griebenow, K., Cao, X., Nafie, L. A., and Schweitzer-Stenner, R. (2004) Preferred peptide backbone conformations in the unfolded state revealed by the structure analysis of alanine-based (AXA) tripeptides in aqueous solution, *Proceedings of the National Academy of Sciences of the United States of America 101*, 10054-10059.

19. Motta, A., Reches, M., Pappalardo, L., Andreotti, G., and Gazit, E. (2005) The preferred conformation of the tripeptide Ala-Phe-Ala in water is an inverse gamma-turn: implications for protein folding and drug design, *Biochemistry 44*, 14170-14178.

20. 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., and Obradovic, Z. (2001) Intrinsically disordered protein, *Journal of molecular graphics & modelling 19*, 26-59.

21. Romero, P., Obradovic, Z., and Dunker, A. K. (2004) Natively disordered proteins: functions and predictions, *Applied bioinformatics 3*, 105-113.

22. Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions?, *Proteins 41*, 415-427.

23. Tompa, P. (2002) Intrinsically unstructured proteins, *Trends in biochemical sciences 27*, 527-533.

24. Liu, J., and Rost, B. (2003) NORSp: Predictions of long regions without regular secondary structure, *Nucleic acids research 31*, 3833-3835.

25. Liu, J., Tan, H., and Rost, B. (2002) Loopy proteins appear conserved in evolution, *Journal of molecular biology 322*, 53-64.

26. Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., and Russell, R. B. (2003) Protein disorder prediction: implications for structural proteomics, *Structure 11*, 1453-1459.

27. Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T., and Dahlquist, F. W. (1997) The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28, *Nature structural biology 4*, 285-291.

28. Daughdrill, G. W., Hanely, L. J., and Dahlquist, F. W. (1998) The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations, *Biochemistry 37*, 1076-1082.

29. Fletcher, C. M., McGuire, A. M., Gingras, A. C., Li, H., Matsuo, H., Sonenberg, N., and Wagner, G. (1998) 4E binding proteins inhibit the translation factor eIF4E without folded structure, *Biochemistry 37*, 9-15.

30. Fletcher, C. M., and Wagner, G. (1998) The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein, *Protein Sci 7*, 1639-1642.

31. Pfander, R., Neumann, L., Zweckstetter, M., Seger, C., Holak, T. A., and Tampe, R. (1999) Structure of the active domain of the herpes simplex virus protein ICP47 in water/sodium dodecyl sulfate solution determined by nuclear magnetic resonance spectroscopy, *Biochemistry 38*, 13692-13698.

32. Takuma, K., Yan, S. S., Stern, D. M., and Yamada, K. (2005) Mitochondrial dysfunction, endoplasmic reticulum stress, and apoptosis in Alzheimer's disease, *Journal of pharmacological sciences 97*, 312-316.

33. Hilser, V. J., and Thompson, E. B. (2007) Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins, *Proceedings of the National Academy of Sciences of the United States of America 104*, 8311-8315.

34. Singh, G. P., Ganapathi, M., and Dash, D. (2007) Role of intrinsic disorder in transient interactions of hub proteins, *Proteins 66*, 761-765.

35. Kourie, J. I. (2001) Mechanisms of amyloid beta protein-induced modification in ion transport systems: implications for neurodegenerative diseases, *Cellular and molecular neurobiology 21*, 173-213.

36. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function, *Biochemistry 41*, 6573-6582.

37. Dunker, A. K., Brown, C. J., and Obradovic, Z. (2002) Identification and functions of usefully disordered proteins, *Advances in protein chemistry 62*, 25-49.

38. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks, *The FEBS journal 272*, 5129-5148.

39. Radivojac, P., Iakoucheva, L. M., Oldfield, C. J., Obradovic, Z., Uversky, V. N., and Dunker, A. K. (2007) Intrinsic disorder and functional proteomics, *Biophysical journal 92*, 1439-1456.

40. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling, *J Mol Recognit 18*, 343-384.

41. Reeves, R., and Beckerbauer, L. (2001) HMGI/Y proteins: flexible regulators of transcription and chromatin structure, *Biochimica et biophysica acta 1519*, 13-29.

42. Elton, T. S., and Reeves, R. (1986) Purification and postsynthetic modifications of Friend erythroleukemic cell high mobility group protein HMG-I, *Analytical biochemistry 157*, 53-62.

43. Sgarra, R., Lee, J., Tessari, M. A., Altamura, S., Spolaore, B., Giancotti, V., Bedford, M. T., and Manfioletti, G. (2006) The AT-hook of the chromatin architectural transcription factor high mobility group A1a is arginine-methylated by protein arginine methyltransferase 6, *The Journal of biological chemistry 281*, 3764-3772.

44. Cattaruzzi, G., Altamura, S., Tessari, M. A., Rustighi, A., Giancotti, V., Pucillo, C., and Manfioletti, G. (2007) The second AT-hook of the architectural transcription factor HMGA2 is determinant for nuclear localization and function, *Nucleic acids research 35*, 1751-1760.

45. Fedele, M., Battista, S., Manfioletti, G., Croce, C. M., Giancotti, V., and Fusco, A. (2001) Role of the high mobility group A proteins in human lipomas, *Carcinogenesis 22*, 1583-1591.

46. American Heart Association. (2007) Cardiovascular Disease Statistics, American Heart Association, Dallas, Texas.

47. Center for Disease Control. (2007) Cardiovascular Disease Fact Sheet, Center for Disease Control.

48. Merk & Co. Inc. (2007) The Merk Manual Medical Library, (Laboratories, M. R., Ed.) http://www.merck.com/mmhe/index.html ed., Merk Research Laboratories, Whitehouse Station, N.J.

49. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S., and Dunker, A. K. (1998) Thousands of proteins likely to have long disordered regions, *Pacific Symposium on Biocomputing*, 437-448.

50. Dunker, A. K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., and Villafranca, J. E. (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations, *Pacific Symposium on Biocomputing*, 473-484.

51. Romero, Obradovic, and Dunker, K. (1997) Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family, *Genome Inform Ser Workshop Genome Inform 8*, 110-124.

52. Cheng, Y., LeGall, T., Oldfield, C. J., Dunker, A. K., and Uversky, V. N. (2006) Abundance of intrinsic disorder in protein associated with cardiovascular disease, *Biochemistry 45*, 10448-10460.

53. Demchenko, A. P. (2001) Recognition between flexible protein molecules: induced and assisted folding, *J Mol Recognit 14*, 42-61.

54. Hubbard, S. J. (1998) The structural aspects of limited proteolysis of native proteins, *Biochimica et biophysica acta 1382*, 191-206.

55. Hubbard, S. J., Beynon, R. J., and Thornton, J. M. (1998) Assessment of conformational parameters as predictors of limited proteolytic sites in native protein structures, *Protein engineering 11*, 349-359.

56. Hubbard, S. J., Campbell, S. F., and Thornton, J. M. (1991) Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors, *Journal of molecular biology 220*, 507-530.

57. Hubbard, S. J., Eisenmenger, F., and Thornton, J. M. (1994) Modeling studies of the change in conformation required for cleavage of limited proteolytic sites, *Protein Sci 3*, 757-768.

58. Hubbard, S. J., Thornton, J. M., and Campbell, S. F. (1992) Substrate recognition by proteinases, *Faraday discussions*, 13-23.

59. Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., and Zambonin, M. (1986) Correlation between sites of limited proteolysis and segmental mobility in thermolysin, *Biochemistry 25*, 1847-1851.

60. Fontana, A., Polverino de Laureto, P., and De Filippis, V. (1993) *Molecular aspects of proteolysis of globular proteins.*, Elsevier, Amsterdam.

61. Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997) Probing the partly folded states of proteins by limited proteolysis, *Folding & design 2*, R17-26.

62. Novotny, J., and Bruccoleri, R. E. (1987) Correlation among sites of limited proteolysis, enzyme accessibility and segmental mobility, *FEBS letters 211*, 185-189.

63. Spolaore, B., Bermejo, R., Zambonin, M., and Fontana, A. (2001) Protein interactions leading to conformational changes monitored by limited proteolysis: apo form and fragments of horse cytochrome c, *Biochemistry 40*, 9460-9468.

64. Kriwacki, R. W., Wu, J., Tennant, L., Wright, P. E., and Siuzdak, G. (1997) Probing protein structure using biochemical and biophysical methods. Proteolysis, matrix-assisted laser desorption/ionization mass spectrometry, high-performance liquid chromatography and size-exclusion chromatography of p21Waf1/Cip1/Sdi1, *Journal of chromatography 777*, 23-30.

65. Markus, G. (1965) Protein substrate conformation and proteolysis, *Proceedings of the National Academy of Sciences of the United States of America 54*, 253-258.

66. Squire, P. G. (1964) A Relationship between the Molecular Weights of Macromolecules and Their Elution Volumes Based on a Model for Sephadex Gel Filtration, *Archives of biochemistry and biophysics 107*, 471-478.

67. Rodbard, D., and Chrambach, A. (1970) Unified theory for gel electrophoresis and gel filtration, *Proceedings of the National Academy of Sciences of the United States of America 65*, 970-977.

68. Ward, D. N., and Arnott, M. S. (1965) Gel Filtration of Proteins, with Particular Reference to the Glycoprotein, Luteinizing Hormone, *Analytical biochemistry 12*, 296-302.

69. Hecht, E. (1998) *Optics*, 3rd ed., Addison Wesley, Longman, Massachusetts.

70. Fasman, G. D. (1996) *Circular Dichroism and the Conformational Analysis of Biomolecules*, Plenum Press, New York.

71. Greenfield, N. J. (2004) Analysis of circular dichroism data, *Methods in enzymology 383*, 282-317.

72. Pierce, M. M., Raman, C. S., and Nall, B. T. (1999) Isothermal titration calorimetry of protein-protein interactions, *Methods (San Diego, Calif 19*, 213-221.

73. Jelesarov, I., and Bosshard, H. R. (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition, *J Mol Recognit 12*, 3-18.

74. Lakowicz, J. R. (1999) *Principles of Fluorescence Spectroscopy*, 2nd edition ed., Plenum Publishing Corporation, New York.

75. Vivian, J. T., and Callis, P. R. (2001) Mechanisms of tryptophan fluorescence shifts in proteins, *Biophysical journal 80*, 2093-2109.

76. Chen, Y., and Barkley, M. D. (1998) Toward understanding tryptophan fluorescence in proteins, *Biochemistry 37*, 9976-9982.

77. Stryer, L. (1968) Fluorescence spectroscopy of proteins, *Science (New York, N.Y 162*, 526-533.

78. Goldberg, M. E., York, S., and Stryer, L. (1968) Fluorescence studies of substrate and subunit interactions of the beta-2 protein of Escherichia coli tryptophan synthetase, *Biochemistry 7*, 3662-3667.

79. Lackowicz, J. (1999) *Principles of fluorescence spectroscopy*, Kluwer Academic/Plenum Publishers, New York.

80. Wolf, D. E. (2007) Fundamentals of fluorescence and fluorescence microscopy, *Methods in cell biology 81*, 63-91.

81. Cardullo, R. A., and Parpura, V. (2003) Fluorescence resonance energy transfer microscopy: theory and instrumentation, *Methods in cell biology 72*, 415-430.

82. Caylor, C. L., Dobrianov, I., Lemay, S. G., Kimmer, C., Kriminski, S., Finkelstein, K. D., Zipfel, W., Webb, W. W., Thomas, B. R., Chernov, A. A., and Thorne, R. E. (1999) Macromolecular impurities and disorder in protein crystals, *Proteins 36*, 270-281.

83. Dobrianov, I., Kriminski, S., Caylor, C. L., Lemay, S. G., Kimmer, C., Kisselev, A., Finkelstein, K. D., and Thorne, R. E. (2001) Dynamic response of tetragonal lysozyme crystals to changes in relative humidity: implications for post-growth crystal treatments, *Acta crystallographica 57*, 61-68.

84. Loris, R., Marianovsky, I., Lah, J., Laeremans, T., Engelberg-Kulka, H., Glaser, G., Muyldermans, S., and Wyns, L. (2003) Crystal structure of the intrinsically flexible addiction antidote MazE, *The Journal of biological chemistry 278*, 28252-28257.

85. Stryer, L. (1968) Implications of X-ray crystallographic studies of protein structure, *Annual review of biochemistry 37*, 25-50.

86. Keeler, J. (2005) *Understanding NMR Spectroscopy*, John Wiley And Sons, Chicago.

87. Wutrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.

88. Shortle, D. R. (1996) Structural analysis of non-native states of proteins by NMR methods, *Current opinion in structural biology 6*, 24-30.

89. Vucetic, S., Obradovic, Z., Vacic, V., Radivojac, P., Peng, K., Iakoucheva, L. M., Cortese, M. S., Lawson, J. D., Brown, C. J., Sikes, J. G., Newton, C. D., and Dunker, A. K. (2005) DisProt: a database of protein disorder, *Bioinformatics (Oxford, England) 21*, 137-140.

90. Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z., and Dunker, A. K. (2007) DisProt: the Database of Disordered Proteins, *Nucleic acids research 35*, D786-793.

91. Chou, P. Y., and Fasman, G. D. (1974) Prediction of protein conformation, *Biochemistry 13*, 222-245.

92. Chou, P. Y., and Fasman, G. D. (1978) Empirical predictions of protein conformation, *Annual review of biochemistry 47*, 251-276.

93. Chou, P. Y., and Fasman, G. D. (1978) Prediction of the secondary structure of proteins from their amino acid sequence, *Advances in enzymology and related areas of molecular biology 47*, 45-148.

94. Kyngas, J., and Valjakka, J. (1998) Unreliability of the Chou-Fasman parameters in predicting protein secondary structure, *Protein engineering 11*, 345-348.

95. Nishikawa, K. (1983) Assessment of secondary-structure prediction of proteins. Comparison of computerized Chou-Fasman method with others, *Biochimica et biophysica acta 748*, 285-299.

96. Garner, E., Cannon, P., Romero, P., Obradovic, Z., and Dunker, A. K. (1998) Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization, *Genome Inform Ser Workshop Genome Inform 9*, 201-213.

97. Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M., and Iakoucheva, L. M. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes, *PLoS computational biology 2*, e100.

98. Idicula-Thomas, S., and Balaji, P. V. (2005) Understanding the relationship between the primary structure of proteins and their amyloidogenic propensity: clues from inclusion body formation, *Protein Eng Des Sel 18*, 175-180.

99. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C. J., and Dunker, A. K. (2003) Predicting intrinsic disorder from amino acid sequence, *Proteins 53 Suppl 6*, 566-572.

100. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Proteins 42*, 38-48.

101. Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., and Dunker, A. K. (2005) Comparing and combining predictors of mostly disordered proteins, *Biochemistry 44*, 1989-2000.

102. Vucetic, S., Brown, C. J., Dunker, A. K., and Obradovic, Z. (2003) Flavors of protein disorder, *Proteins 52*, 573-584.

103. Radivojac, P., Obradovic, Z., Smith, D. K., Zhu, G., Vucetic, S., Brown, C. J., Lawson, J. D., and Dunker, A. K. (2004) Protein flexibility and intrinsic disorder, *Protein Sci 13*, 71-80.

104. Ackerman, M. S., and Shortle, D. (2002) Robustness of the long-range structure in denatured staphylococcal nuclease to changes in amino acid sequence, *Biochemistry 41*, 13791-13797.

105. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-range interactions within a nonnative protein, *Science (New York, N.Y 295*, 1719-1722.

106. Shortle, D., and Ackerman, M. S. (2001) Persistence of native-like topology in a denatured protein in 8 M urea, *Science (New York, N.Y 293*, 487-489.

107. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins, *Current opinion in structural biology 12*, 54-60.

108. Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism, *Proceedings of the National Academy of Sciences of the United States of America 97*, 8868-8873.

109. Levy, Y., Onuchic, J. N., and Wolynes, P. G. (2007) Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition, *Journal of the American Chemical Society 129*, 738-739.

110. Levy, Y., Cho, S. S., Onuchic, J. N., and Wolynes, P. G. (2005) A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes, *Journal of molecular biology 346*, 1121-1145.

111. Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature 447*, 1021-1025.

112. Barrett, A. J., Rawlings, N. D., and Woessner, J. F. (2003) *The Handbook of Proteolytic Enzymes*, 2nd edition ed., Academic Press, New York.

113. Blundell, T. L., Jenkins, J. A., Sewell, B. T., Pearl, L. H., Cooper, J. B., Tickle, I. J., Veerapandian, B., and Wood, S. P. (1990) X-ray analyses of aspartic proteinases. The three-dimensional structure at 2.1 A resolution of endothiapepsin, *Journal of molecular biology 211*, 919-941.

114. Sali, A., Veerapandian, B., Cooper, J. B., Moss, D. S., Hofmann, T., and Blundell, T. L. (1992) Domain flexibility in aspartic proteinases, *Proteins 12*, 158-170.

115. Coates, L., Erskine, P. T., Crump, M. P., Wood, S. P., and Cooper, J. B. (2002) Five atomic resolution structures of endothiapepsin inhibitor complexes: implications for the aspartic proteinase mechanism, *Journal of molecular biology 318*, 1405-1415.

116. Coates, L., Erskine, P. T., Wood, S. P., Myles, D. A., and Cooper, J. B. (2001) A neutron Laue diffraction study of endothiapepsin: implications for the aspartic proteinase mechanism, *Biochemistry 40*, 13149-13157.

117. Dash, C., Phadtare, S., Deshpande, V., and Rao, M. (2001) Structural and mechanistic insight into the inhibition of aspartic proteases by a slow-tight binding inhibitor from an extremophilic Bacillus sp.: correlation of the kinetic parameters with the inhibitor induced conformational changes, *Biochemistry 40*, 11525-11532.

118. Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., and Hoover, D. J. (1992) Direct observation by X-ray analysis of the tetrahedral "intermediate" of aspartic proteinases, *Protein Sci 1*, 322-328.

119. Mechler, B., Muller, M., Muller, H., Meussdoerffer, F., and Wolf, D. H. (1982) In vivo biosynthesis of the vacuolar proteinases A and B in the yeast Saccharomyces cerevisiae, *The Journal of biological chemistry 257*, 11203-11206.

120. Meussdoerffer, F., Tortora, P., and Holzer, H. (1980) Purification and properties of proteinase A from yeast, *The Journal of biological chemistry 255*, 12087-12093.

121. van den Hazel, H. B., Kielland-Brandt, M. C., and Winther, J. R. (1993) The propeptide is required for in vivo formation of stable active yeast proteinase A and can function even when not covalently linked to the mature region, *The Journal of biological chemistry 268*, 18002-18007.

122. Klionsky, D. J., Banta, L. M., and Emr, S. D. (1988) Intracellular sorting and processing of a yeast vacuolar hydrolase: proteinase A propeptide contains vacuolar targeting information, *Molecular and cellular biology 8*, 2105-2116.

123. Hirsch, H. H., Schiffer, H. H., and Wolf, D. H. (1992) Biogenesis of the yeast vacuole (lysosome). Proteinase yscB contributes molecularly and kinetically to vacuolar hydrolase-precursor maturation, *European journal of biochemistry / FEBS 207*, 867-876.

124. van den Hazel, H. B., Kielland-Brandt, M. C., and Winther, J. R. (1992) Autoactivation of proteinase A initiates activation of yeast vacuolar zymogens, *European journal of biochemistry / FEBS 207*, 277-283.

125. Kato, M., Kuzuhara, Y., Maeda, H., Shiraga, S., and Ueda, M. (2006) Analysis of a processing system for proteases using yeast cell surface engineering: conversion of precursor of proteinase A to active proteinase A, *Applied microbiology and biotechnology 72*, 1229-1237.

126. Afting, E. G., Lynen, A., Hinze, H., and Holzer, H. (1976) Effects of yeast proteinase A, proteinase B and carboxypeptidase Y on yeast phosphofructokinase, *Hoppe-Seyler's Zeitschrift fur physiologische Chemie 357*, 1771-1777.

127. Ammerer, G., Hunter, C. P., Rothman, J. H., Saari, G. C., Valls, L. A., and Stevens, T. H. (1986) PEP4 gene of Saccharomyces cerevisiae encodes proteinase A, a vacuolar enzyme required for processing of vacuolar precursors, *Molecular and cellular biology 6*, 2490-2499.

128. Woolford, C. A., Daniels, L. B., Park, F. J., Jones, E. W., Van Arsdell, J. N., and Innis, M. A. (1986) The PEP4 gene encodes an aspartyl protease implicated in the posttranslational regulation of Saccharomyces cerevisiae vacuolar hydrolases, *Molecular and cellular biology 6*, 2500-2510.

129. Zubenko, G. S., Park, F. J., and Jones, E. W. (1983) Mutations in PEP4 locus of Saccharomyces cerevisiae block final step in maturation of two vacuolar hydrolases, *Proceedings of the National Academy of Sciences of the United States of America 80*, 510-514.

130. Dreyer, T., Valler, M. J., Kay, J., Charlton, P., and Dunn, B. M. (1985) The selectivity of action of the aspartic-proteinase inhibitor IA3 from yeast (Saccharomyces cerevisiae), *The Biochemical journal 231*, 777-779.

131. Lenney, J. F. (1975) Three yeast proteins that specifically inhibit yeast proteases A, B, and C, *Journal of bacteriology 122*, 1265-1273.

132. Parr, C. L., Keates, R. A., Bryksa, B. C., Ogawa, M., and Yada, R. Y. (2007) The structure and function of Saccharomyces cerevisiae proteinase A, *Yeast (Chichester, England) 24*, 467-480.

133. Winterburn, T. J., Wyatt, D. M., Phylip, L. H., Berry, C., Bur, D., and Kay, J. (2006) Adaptation of the behaviour of an aspartic proteinase inhibitor by relocation of a lysine residue by one helical turn, *Biological chemistry 387*, 1139-1142.

134. Winterburn, T. J., Wyatt, D. M., Phylip, L. H., Bur, D., Harrison, R. J., Berry, C., and Kay, J. (2007) Key features determining the specificity of aspartic proteinase inhibition by the helix-forming IA3 polypeptide, *The Journal of biological chemistry 282*, 6508-6516.

135. Winterburn, T. J., Phylip, L. H., Bur, D., Wyatt, D. M., Berry, C., and Kay, J. (2007) N-terminal extension of the yeast IA3 aspartic proteinase inhibitor relaxes the strict intrinsic selectivity, *The FEBS journal 274*, 3685-3694.

136. Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S., and Richardson, D. C. (1999) Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms, *Journal of molecular biology 285*, 1711-1733.

137. Golub, G. H., and Reinsch, C. (1970) Singular value decomposition and least squares solution, *Numerische Mathematik 14*, 403-420.

138. Pastore, A., and Saudek, V. (1990) The relationship between chemical shift and secondary structure in proteins, *J. Magn. Reson. 90*, 165-176.

139. Williamson, M. P. (1990) Secondary-structure dependent chemical shifts in proteins, *Biopolymers 29*, 1423-1431.

140. Matsuura, H., Shimotakahara, S., Sakuma, C., Tashiro, M., Shindo, H., Mochizuki, K., Yamagishi, A., Kojima, M., and Takahashi, K. (2004) Thermal unfolding of ribonuclease T1 studied by multi-dimensional NMR spectroscopy, *Biological chemistry 385*, 1157-1164.

141. Sadqi, M., Fushman, D., and Munoz, V. (2006) Atom-by-atom analysis of global downhill protein folding, *Nature 442*, 317-321.

142. Zuiderweg, E. R. (2002) Mapping protein-protein interactions in solution by NMR spectroscopy, *Biochemistry 41*, 1-7.

143. Rajagopal, P., Waygood, E. B., Reizer, J., Saier, M. H., Jr., and Klevit, R. E. (1997) Demonstration of protein-protein interaction specificity by NMR chemical shift mapping, *Protein Sci 6*, 2624-2627.

144. Baber, J. L., Levens, D., Libutti, D., and Tjandra, N. (2000) Chemical shift mapped DNA-binding sites and 15N relaxation analysis of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, *Biochemistry 39*, 6022-6032.

145. Lee, A. L., Volkman, B. F., Robertson, S. A., Rudner, D. Z., Barbash, D. A., Cline, T. W., Kanaar, R., Rio, D. C., and Wemmer, D. E. (1997) Chemical shift mapping of the RNA-binding interface of the multiple-RBD protein sex-lethal, *Biochemistry 36*, 14306-14317.

146. Henry, E. M., and Hofrichter, J. (1992) Singular value decomposition: application to the analysis of experimental data, *Methods Enzymol. 210*, 129-192.

147. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J Biomol NMR 5*, 67-81.

148. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, *J Biomol NMR 2*, 661-665.

149. Kay, L., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc. 114*, 10663-10665.

150. Marion, D., Ikura, M., Tschudin, R., and Bax, A. (1989) Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins., *J. Magn. Reson. 85*, 393-399.

151. Grzesiek, S., and Bax, A. (1992) An efficient experiment for sequential backbone assignment of medium sized isotopically enriched proteins, *J. Magn. Reson. 99*, 201-207.

152. Grzesiek, S., and Bax, A. (1992) Correlating Backbone Amide and Side Chain Resonances in Larger Proteins by Multiple Relayed Triple Resonance NMR, *J. Am. Chem. Soc. 114*, 6291-6293.

153. Grzesiek, S., and Bax, A. (1992) Improved 3D Triple-Resonance NMR Techniques Applied to a 31 kDa Protein, *J. Magn. Reson. 96*, 432-440.

154. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) Three-Dimensional Triple-Resonance NMR-Spectroscopy of Isotopically Enriched Proteins, *J. Magn. Reson. 89*, 496-514.

155. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog. Nuc. Magn. Res. Spec. 34*, 93-158.

156. Palmer, A. G., Cavanagh, J., Wright, P. E., and Rance, M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy, *J. Magn. Reson. (1969) 93*, 151-170.

157. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR 6*, 277-293.

158. Johnson, B., and Blevins, R. (1994) NMR View: A computer program for the visualization and analysis of NMR data, *J. Biomol. NMR 4*, 603 - 614.

159. Friedrichs, M. (1995) A model-free algorithm for the removal of baseline artifacts, *J. Biomol. NMR 5*, 147-153.

160. Schwarzinger, S., Kroon, G. J., Foss, T. R., Wright, P. E., and Dyson, H. J. (2000) Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView, *J Biomol NMR 18*, 43-48.

161. Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts, *Journal of the American Chemical Society 123*, 2970-2978.

162. Golub, G. H., and Van Loan, C. F. (1996) *Matrix Computations*, 3 ed., Johns Hopkins University Press, Baltimore.

163. Nash, J. C. (1990) *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*, 2 ed., Adam Hilger.

164. Press, W., Flannery, B., Teukolsky, S., and Vetterling, W. (1992) *Numerical Recipes in C : The Art of Scientific Computing*, 2 ed., Cambridge University Press.

165. Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants, *Biochemistry 27*, 8063-8068.

166. Dobson, C. M., and Fersht, A. R. (1996) *Protein Folding*, Cambridge University Press.

167. Demchenko, A. P. (2001) Concepts and misconcepts in the analysis of simple kinetics of protein folding, *Curr Protein Pept Sci 2*, 73-98.

168. Asakura, T., Kazuhiro, T., Demura, M., and Williamson, M. P. (1995) The relationship between amide proton chemical shifts and secondary structure in proteins, *J Biomol NMR 6*, 227-236.

169. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure, *Journal of molecular biology 222*, 311-333.

170. Wishart, D. S., and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination, *Methods in enzymology 239*, 363-392.

171. Wishart, D. S., and Case, D. A. (2001) Use of chemical shifts in macromolecular structure determination, *Methods in enzymology 338*, 3-34.

172. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry 31*, 1647-1651.

173. Buck, M. (1998) Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins, *Q Rev Biophys 31*, 297-355.

174. Fink, A. L. (2005) Natively unfolded proteins, *Current opinion in structural biology 15*, 35-41.

175. Matsuo, H., Walters, K. J., Teruya, K., Tanaka, T., T., G. G., Lippard, S. J., Kyogoku, Y., and Wagner, G. (1999) Identification by NMR Spectroscopy of Residues at Contact Surfaces in Large, Slowly Exchanging Macromolecular Complexes., *J. Am. Chem. Soc. 121*, 9903-9904.

176. Edison, A. S., Espinoza, E., and Zachariah, C. (1999) Conformational ensembles: the role of neuropeptide structures in receptor binding, *J Neurosci 19*, 6318-6326.

177. Dossey, A. T., Reale, V., Chatwin, H., Zachariah, C., Debono, M., Evans, P. D., and Edison, A. S. (2006) NMR Analysis of Caenorhabditis elegans FLP-18 Neuropeptides: Implications for NPR-1 Activation, *Biochemistry 45*, 7586-7597.

178. Santiveri, C. M., Pantoja-Uceda, D., Rico, M., and Jimenez, M. A. (2005) Beta-hairpin formation in aqueous solution and in the presence of trifluoroethanol: a (1)H and (13)C nuclear magnetic resonance conformational study of designed peptides, *Biopolymers 79*, 150-162.

179. Kemmink, J., and Creighton, T. E. (1995) Effects of trifluoroethanol on the conformations of peptides representing the entire sequence of bovine pancreatic trypsin inhibitor, *Biochemistry 34*, 12630-12635.

180. Bolin, K. A., Pitkeathly, M., Miranker, A., Smith, L. J., and Dobson, C. M. (1996) Insight into a random coil conformation and an isolated helix: structural and dynamical characterisation of the C-helix peptide from hen lysozyme, *Journal of molecular biology 261*, 443-453.

181. Kentsis, A., and Sosnick, T. R. (1998) Trifluoroethanol promotes helix formation by destabilizing backbone exposure: desolvation rather than native hydrogen bonding defines the kinetic pathway of dimeric coiled coil folding, *Biochemistry 37*, 14613-14622.

182. Shiraki, K., Nishikawa, K., and Goto, Y. (1995) Trifluoroethanol-induced stabilization of the alpha-helical structure of beta-lactoglobulin: implication for non-hierarchical protein folding, *Journal of molecular biology 245*, 180-194.

183. Luidens, M. K., Figge, J., Breese, K., and Vajda, S. (1996) Predicted and trifluoroethanol-induced alpha-helicity of polypeptides, *Biopolymers 39*, 367-376.

184. Iovino, M., Falconi, M., Marcellini, A., and Desideri, A. (2001) Molecular dynamics simulation of the antimicrobial salivary peptide histatin-5 in water and in trifluoroethanol: a microscopic description of the water destructuring effect, *J Pept Res 58*, 45-55.

185. Petsko, G. A., and Ringe, D. R. (2003) *Protein Structure and Function*, 1 ed., New Science Press, New York.

186. Hansen, J. C., Lu, X., Ross, E. D., and Woody, R. W. (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains, *The Journal of biological chemistry 281*, 1853-1856.

187. Forster, T. (1993) *Annalen der Physik (english translation)*, American Institute of Physics, New York.

188. Stryer, L. (1978) Fluorescence energy transfer as a spectroscopic ruler, *Annual review of biochemistry 47*, 819-846.

189. Van der Meer, B. W., Coker, G., and Chen, S. (1994) *Resonance Energy Transfer: Theory and Data*, VCH Publishers, New York.

190. Hansen, S. B., Radic, Z., Talley, T. T., Molles, B. E., Deerinck, T., Tsigelny, I., and Taylor, P. (2002) Tryptophan fluorescence reveals conformational changes in the acetylcholine binding protein, *The Journal of biological chemistry 277*, 41299-41302.

191. Nannepaga, S. J., Gawalapu, R., Velasquez, D., and Renthal, R. (2004) Estimation of helix-helix association free energy from partial unfolding of bacterioopsin, *Biochemistry 43*, 550-559.

192. Dunn, B. M., Scarborough, P. E., Davenport, R., and Swietnicki, W. (1994) Analysis of proteinase specificity by studies of peptide substrates. The use of UV and fluorescence spectroscopy to quantitate rates of enzymatic cleavage, *Methods Mol Biol 36*, 225-243.

193. Green, T. B. (2003) *Structure-function studies of IA3, an inhibitor of yeast proteinase A*, University of Florida, Gainesville.

194. Qui, L. (2003) *Laser induced temperature jump investigations of fast protein folding dynamics*, Diss. University of Florida, Gainesville.

195. Qiu, L., Pabit, S. A., Roitberg, A. E., and Hagen, S. J. (2002) Smaller and faster: the 20-residue Trp-cage protein folds in 4 micros, *Journal of the American Chemical Society 124*, 12952-12953.

196. Ababou, A., and Ladbury, J. E. (2006) Survey of the year 2004: literature on applications of isothermal titration calorimetry, *J Mol Recognit 19*, 79-89.

197. Ababou, A., and Ladbury, J. E. (2007) Survey of the year 2005: literature on applications of isothermal titration calorimetry, *J Mol Recognit 20*, 4-14.

198. Cliff, M. J., Gutierrez, A., and Ladbury, J. E. (2004) A survey of the year 2003 literature on applications of isothermal titration calorimetry, *J Mol Recognit 17*, 513-523.

199. Cliff, M. J., and Ladbury, J. E. (2003) A survey of the year 2002 literature on applications of isothermal titration calorimetry, *J Mol Recognit 16*, 383-391.

200. Hytonen, V. P., Maatta, J. A., Niskanen, E. A., Huuskonen, J., Helttunen, K. J., Halling, K. K., Nordlund, H. R., Rissanen, K., Johnson, M. S., Salminen, T. A., Kulomaa, M. S., Laitinen, O. H., and Airenne, T. T. (2007) Structure and characterization of a novel chicken biotin-binding protein A (BBP-A), *BMC Struct Biol 7*, 8.

201. Leavitt, S., and Freire, E. (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry, *Current opinion in structural biology 11*, 560-566.

202. Privalov, P. L., and Dragan, A. I. (2007) Microcalorimetry of biological macromolecules, *Biophys Chem 126*, 16-24.

203. Murphy, K. P. (1999) Predicting binding energetics from structure: looking beyond DeltaG degrees, *Med Res Rev 19*, 333-339.

204. Cooper, A. (2000) Heat capacity of hydrogen-bonded networks: an alternative view of protein folding thermodynamics, *Biophys Chem 85*, 25-39.

205. Prabhu, N., and Sharp, K. (2006) Protein-solvent interactions, *Chem Rev 106*, 1616-1623.

206. Prabhu, N. V., and Sharp, K. A. (2005) Heat capacity in proteins, *Annu Rev Phys Chem 56*, 521-548.

207. Sturtevant, J. M. (1977) Heat capacity and entropy changes in processes involving proteins, *Proceedings of the National Academy of Sciences of the United States of America 74*, 2236-2240.

208. Zolotnitsky, G., Cogan, U., Adir, N., Solomon, V., Shoham, G., and Shoham, Y. (2004) Mapping glycoside hydrolase substrate subsites by isothermal titration calorimetry, *Proceedings of the National Academy of Sciences of the United States of America 101*, 11275-11280.

209. Gomez, J., Hilser, V. J., Xie, D., and Freire, E. (1995) The heat capacity of proteins, *Proteins 22*, 404-412.

210. Gribenko, A. V., Keiffer, T. R., and Makhatadze, G. I. (2006) Amino acid substitutions affecting protein dynamics in eglin C do not affect heat capacity change upon unfolding, *Proteins 64*, 295-300.

211. Bakk, A., Hoye, J. S., and Hansen, A. (2001) Heat capacity of protein folding, *Biophysical journal 81*, 710-714.

212. Kumar, S., Tsai, C. J., and Nussinov, R. (2003) Temperature range of thermodynamic stability for the native state of reversible two-state proteins, *Biochemistry 42*, 4864-4873.

213. Yang, A. S., Sharp, K. A., and Honig, B. (1992) Analysis of the heat capacity dependence of protein folding, *Journal of molecular biology 227*, 889-900.

214. Cooper, A. (2005) Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics, *Biophys Chem 115*, 89-97.

215. Makhatadze, G. I., and Privalov, P. L. (1990) Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect, *Journal of molecular biology 213*, 375-384.

216. Makhatadze, G. I., and Privalov, P. L. (1993) Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration, *Journal of molecular biology 232*, 639-659.

217. Privalov, P. L., and Makhatadze, G. I. (1990) Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects, *Journal of molecular biology 213*, 385-391.

218. Privalov, P. L., and Makhatadze, G. I. (1993) Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration, *Journal of molecular biology 232*, 660-679.

219. Kaspar, S., Perozzo, R., Reinelt, S., Meyer, M., Pfister, K., Scapozza, L., and Bott, M. (1999) The periplasmic domain of the histidine autokinase CitA functions as a highly specific citrate receptor, *Mol Microbiol 33*, 858-872.

220. Prorok, M., and Castellino, F. J. (1998) Thermodynamics of binding of calcium, magnesium, and zinc to the N-methyl-D-aspartate receptor ion channel peptidic inhibitors, conantokin-G and conantokin-T, *The Journal of biological chemistry 273*, 19573-19578.

221. Yokota, A., Takenaka, H., Oh, T., Noda, Y., and Segawa, S. (1998) Thermodynamics of the reconstitution of tuna cytochrome c from two peptide fragments, *Protein Sci 7*, 1717-1727.

222. Saheki, T., and Holzer, H. (1975) Proteolytic activities in yeast, *Biochimica et biophysica acta 384*, 203-214.

223. Saheki, T., Matsuda, Y., and Holzer, H. (1974) Purification and characterization of macromolecular inhibitors of proteinase A from yeast, *European journal of biochemistry / FEBS 47*, 325-332.

224. Goldberg, A. L. (2003) Protein degradation and protection against misfolded or damaged proteins, *Nature 426*, 895-899.

225. Gardner, R. G., Nelson, Z. W., and Gottschling, D. E. (2005) Degradation-mediated protein quality control in the nucleus, *Cell 120*, 803-815.

226. Hoffmann, R., and Valencia, A. (2003) Protein interaction: same network, different hubs, *Trends Genet 19*, 681-683.

227. Dosztanyi, Z., Chen, J., Dunker, A. K., Simon, I., and Tompa, P. (2006) Disorder and sequence repeats in hub proteins and their implications for network evolution, *J Proteome Res 5*, 2985-2995.

228. Hengen, P. (1995) Purification of His-Tag fusion proteins from Escherichia coli, *Trends in biochemical sciences 20*, 285-286.

229. Gaberc-Porekar, V., and Menart, V. (2005) Potential for using histidine tags in purification of proteins at large scale, *Chemical Engineering and Technology 28*, 1306-1314.

230. Shortle, D. (1996) The denatured state (the other half of the folding equation) and its role in protein stability, *Faseb J 10*, 27-34.

231. Benjwal, S., Verma, S., Rohm, K. H., and Gursky, O. (2006) Monitoring protein aggregation during thermal unfolding in circular dichroism experiments, *Protein Sci 15*, 635-639.

232. Sakharkar, M. K., Kangueane, P., Sakharkar, K. R., and Zhong, Z. (2006) Huge proteins in the human proteome and their participation in hereditary diseases, *In Silico Biol 6*, 275-279.

233. Tan, T., Frenkel, D., Gupta, V., and Deem, M. (2005) Length, protein-protein interactions, and complexity, *Physica A: Statistical Mechanics and its Applications 350*, 52-62.

234. Wang, D., Hsieh, M., and Li, W. H. (2005) A general tendency for conservation of protein length across eukaryotic kingdoms, *Mol Biol Evol 22*, 142-147.

235. Bryan, J. K. (1977) Molecular weights of protein multimers from polyacrylamide gel electrophoresis, *Analytical biochemistry 78*, 513-519.

236. Davis, B. J. (1964) Disc Electrophoresis. Ii. Method and Application to Human Serum Proteins, *Ann N Y Acad Sci 121*, 404-427.

237. Bier, M. (1959) *Electrophoresis: theory, methods, and applications*, Academic Press, New York.

238. Raymond, S. (1964) Acrylamide Gel Electrophoresis, *Ann N Y Acad Sci 121*, 350-365.

239. Raymond, S., and Nakamichi, M. (1962) Gel electrophoresis, *Clin Chem 8*, 471-474.

240. Raymond, S., and Nakamichi, M. (1962) Electrophoresis in synthetic gels. I. Relation of gel structure to resolution, *Analytical biochemistry 3*, 23-30.

241. Raymond, S., Nakamichi, M., and Aurell, B. (1962) Acrylamide gel as an electrophoresis medium, *Nature 195*, 697-698.

242. Smithies, O. (1995) Early days of gel electrophoresis, *Genetics 139*, 1-4.

243. McMurry (2003) *Organic Chemistry*, 6th edition ed., Brooks Cole, New York.

244. Bulheller, B. M., Rodger, A., and Hirst, J. D. (2007) Circular and linear dichroism of proteins, *Phys Chem Chem Phys 9*, 2020-2035.

245. Ciulli, A., and Abell, C. (2005) Biophysical tools to monitor enzyme-ligand interactions of enzymes involved in vitamin biosynthesis, *Biochem Soc Trans 33*, 767-771.

246. Lewis, E. A., and Murphy, K. P. (2005) Isothermal titration calorimetry, *Methods Mol Biol 305*, 1-16.

247. Bianconi, M. L. (2007) Calorimetry of enzyme-catalyzed reactions, *Biophys Chem 126*, 59-64.

248. Holdgate, G. A. (2001) Making cool drugs hot: isothermal titration calorimetry as a tool to study binding energetics, *Biotechniques 31*, 164-166, 168, 170 passim.

249. Ladbury, J. E., and Chowdhry, B. Z. (1996) Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions, *Chem Biol 3*, 791-801.

250. Bakk, A. (2004) Is it always possible to distinguish two- and three-state systems by evaluating the van't Hoff enthalpy?, *Phys Biol 1*, 152-158.

251. Bakk, A., Hoye, J. S., and Hansen, A. (2002) Apolar and polar solvation thermodynamics related to the protein unfolding process, *Biophysical journal 82*, 713-719.

252. Feig, A. L. (2007) Applications of isothermal titration calorimetry in RNA biochemistry and biophysics, *Biopolymers*.

253. Buurma, N. J., and Haq, I. (2007) Advances in the analysis of isothermal titration calorimetry data for ligand-DNA interactions, *Methods (San Diego, Calif 42*, 162-172.

254. Seelig, J. (2004) Thermodynamics of lipid-peptide interactions, *Biochimica et biophysica acta 1666*, 40-50.

255. Livingstone, J. R. (1996) Antibody characterization by isothermal titration calorimetry, *Nature 384*, 491-492.

256. Lonhienne, T. G., and Winzor, D. J. (2004) A potential role for isothermal calorimetry in studies of the effects of thermodynamic non-ideality in enzyme-catalyzed reactions, *J Mol Recognit 17*, 351-361.

257. Perozzo, R., Folkers, G., and Scapozza, L. (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects, *J Recept Signal Transduct Res 24*, 1-52.

258. Velazquez-Campoy, A., Leavitt, S. A., and Freire, E. (2004) Characterization of protein-protein interactions by isothermal titration calorimetry, *Methods Mol Biol 261*, 35-54.

259. Ward, W. H., and Holdgate, G. A. (2001) Isothermal titration calorimetry in drug discovery, *Prog Med Chem 38*, 309-376.

260. Ramos, R., Gaisford, S., Buckton, G., Royall, P. G., Yff, B. T., and O'Neill, M. A. (2005) A comparison of chemical reference materials for solution calorimeters, *Int J Pharm 299*, 73-83.

261. Privalov, G. P., and Privalov, P. L. (2000) Problems and prospects in microcalorimetry of biological macromolecules, *Methods in enzymology 323*, 31-62.

262. Weber, P. C., and Salemme, F. R. (2003) Applications of calorimetric methods to drug discovery and the study of protein interactions, *Current opinion in structural biology 13*, 115-121.

263. Gaucher, E. A., Govindarajan, S., and Ganesh, O. K. (2008) Palaeotemperature trend for Precambrian life inferred from resurrected proteins, *Nature 451*, 704-707.

BIOGRAPHICAL SKETCH

Omjoy Kumar Ganesh was born in 1976 in Ann Arbor, Michigan to T.S. and Lalitha Ganesh. He was raised in the cereal city of the U.S., Battle Creek, Michigan until the age of 11. In 1987, his family moved to Gainesville, Florida to escape the cold of the north. In Gainesville, he attended Howard W. Bishop Middle School, followed by F.W. Buchholz High School. It was in high school that he realized his interest in science and math, sometimes annoying teachers with "how" and "why" questions.

In the summer of 1993 , Omjoy (who now went by the nickname OJ) was accepted into the Student Science Training Program, a residential research program at the University of Florida (UF) run by the Center for Precollegiate Education and Training. During this time he worked with Dr. William G. Vernetson at the UF Training Reactor on a project using nuclear activation analysis to investigate aluminum levels and their relationship with Alzheimer's disease.

After graduating from Buchholz High School in 1994, OJ was accepted to and attended UF. The institution was perfect for him, as the university was large enough to afford him the opportunity to experience many different subjects, and meet a wide variety of people to enrich (and sometimes hurt) his life in many ways. After a year focused on math and English, OJ decided to major in chemistry. A year after that decision, he felt that this did not suit him, and he moved to psychology, which after one semester suited him even less. A chance decision to take a physics 3 course (which was not part of the general education requirements) proved pivotal.

The course, taught by Dr. Gary Ihas, introduced OJ to 'strange' concepts such as subatomic particles, relativity, and quantum physics. This sparked OJ's interest and led to his earning a bachelor's degree in physics. During that time he had the privilege to be instructed by Dr. David Tanner, Dr. Yasumasa Takano, and Dr. Dimitri Maslov. Additionally, he worked on two projects, briefly in computational astrophysics with Dr. Robert Bucheler, and more extensively

on atomic force microscopy using carbon nanotubes with Dr. Arthur Hebard and Dr. Andrew Rinzler.

Graduating from UF in 1999, OJ went to teach mathematics and physics at Santa Fe Community College. But after 2 years he decided to continue his education, by entering graduate school, through the Interdisciplinary Program in Biomedical Sciences at UF. Rotating through the labs of Dr. Benner, Dr. Purich, and Dr. Walter, he eventually settled into the labs of Dr. Hagen and Dr. Edison to work on the intrinsically unstructured protein IA_3. This arrangement worked well, since it provided him with the stability of a physics culture while allowing him to explore new territory in biochemistry.

During his graduate work at UF, OJ has coauthored three papers, two in the journal *Biochemistry* (*8, 9*) and one in *Nature* (*263*). In 2007, he was a corecipient of the Boyce award. In that same year he also chaired a lecture panel on the uses of NMR and MRI at UF. More importantly, during his years in graduate school OJ married his lovely wife Katia, made many long lasting friendships, and grew academically into an excellent scientist. While very few things in this day and age have long lasting effects, the knowledge that OJ gained from this experience will last him a lifetime, and will be passed on to others.