Stony Brook University

Folding of α-Helical Proteins and Protein Design Using Non-coded Amino Acids

A Dissertation Presented

by

Benben Song

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University

December 2007

Stony Brook University

The Graduate School

Benben Song

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree,
Hereby recommend the acceptance of this dissertation.

Daniel P. Raleigh, Ph. D., Advisor
Department of Chemistry, Stony Brook University

Peter J. Tonge, Ph. D., Chairperson
Department of Chemistry, Stony Brook University

Carlos Simmerling, Ph. D., Third Member
Department of Chemistry, Stony Brook University

Carlos De Los Santos, Ph. D., Outside Member
Department of Pharmacology, Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin

Dean of the Graduate School

Abstract of the Dissertation

Folding of α-Helical Proteins and Protein Design Using Non-coded Amino Acids

by

Benben Song

Doctor of Philosophy
in
Chemistry

Stony Brook University

2007

A major goal of protein design is to enhance protein stability. Replacement of helix C-capping glycines which have positive φ angles by D-alanine is shown to be a new strategy to increase protein stability. A series of small helical proteins have been targeted and the substitutions stabilize the proteins by 0.6 to 1.2 kcal/mol. Such substitutions can decrease the conformational entropy of the unfolded state without introducing any strain into the native structure. D-valine and D-alanine variants were compared to test for desolvation effect and the results showed that the backbone desolvation has an unfavorable effect on protein stability but is not as large as the favorable entropy effect.

The villin headpiece, HP67, is a small globular protein that consists of two subdomains; the N-terminal subdomain and the C-terminal subdomain, which form an extended hydrophobic core. A salt bridge between E39 and K70 links the two subdomains and is buried in the hydrophobic core. The contribution of this salt bridge to the structure and stability of HP67 was studied by characterizing a series of variants (HP67E39Q, HP67E39L, HP67K70M and HP67E39QK70M). One-dimensional and ^{1}H-^{15}N HSQC NMR were used to characterize the structure of the variants and the N-terminal subdomain was found to be partially unfolded in all cases. Thus the buried salt bridge is indispensable to ensure a correct fold for the N-terminal subdomain.

The contribution of electrostatic interactions to protein folding and stability has been studied in the N-terminal domain of ribosomal protein L9 (NTL9) by analyzing the effects changing ionic strength upon stability and folding kinetics. The Leffler plot of $\Delta G^{\neq}$, the change of the activation free energy upon perturbation of ionic strength, vs. ΔG^{0}, the change of the free energy of folding, is linear (r^2= 0.918) with slope equal to 0.45. The relatively low value of the slope indicates that the ionic strength dependent interactions are modestly developed in the transition state.

A series of $^{13}C{=}^{18}O$ carbonyl-labeled HP36 proteins were synthesized to probe protein folding dynamics at specific sites by IR method. These proteins were prepared as part of a collaboration project with Dr. R. Brian Dyer of Los Alamos National Laboratory.

Table of Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

a_d	The y intercept for a post-transition line
a_n	The y intercept for a pre-transition line
b_d	The slope for a post-transition line
b_n	The slope for a pre-transition line
C_o	Concentration dependent scaling factor for NMR line-shape analysis
CD	Circular dichroism
DSE	The denatured state ensemble
EDTA	Ethylenediaminetetraacetic acid
F	The folded state
Fmoc	9-fluorenylmethyloxycarbonyl
Gnd HCl	Guanidine hydrochloride
h	Planck's constant
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single-quantum coherence
I	The ionic strength
IPTG	Isopropylthio-β-D-galactoside
K	Equilibrium constant for protein folding
k_{ex}	Exchange rate, $k_{ex} = k_f + k_u$
k_f	Folding rate
k_u	Unfolding rate
MALDI-TOF	Matrix assisted laser desorption and ionization time of flight
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
m	Slope of ΔG° vs. denaturant concentration
p_D	Population of unfolded (denatured) protein
p_N	Population of native protein
PAL	Polystyrene Fmoc support for peptide amides
PAL-PEG-PS	Polyethylene glycol polystyrene Fmoc support for peptide amides
ppm	Parts per million
REFER	Rate equilibrium free energy relationship
T_{2D}	Line-width of the denatured state resonance
T_{2N}	Line-width of the native state resonance
TEAP	Triethylamine phosphate
TFA	Trifloroacetic acid
T	Temperature
T_m	Midpoint of thermal denaturation transition
TOCSY	Total correlation spectroscopy
TSE	Transition state ensemble
TSP	3-(Trimethylsilyl)propionate
U	The unfolded state
UV	Ultraviolet
α_x	The value of $(\frac{\partial \Delta G^{\neq}}{\partial x})/(\frac{\partial \Delta G^{0}}{\partial x})$ where x represents a perturbation

β_T	Tanford beta value (= m_f/m_{eq})
θ_m	An alternative notation used for the Tanford β value
ΔC°_p	Heat capacity change of unfolding
$\Delta C^\ddagger_p$	Activation heat capacity change
$\Delta G^\circ_{D\text{-}N}$	Standard state free energy of unfolding
$\Delta G^\ddagger$	Activation free energy
ΔH°	Standard state enthalpy of unfolding
$\Delta H^\ddagger$	Activation enthalpy
ΔS°	Standard state entropy of unfolding
$\Delta S^\ddagger$	Activation entropy
ν_D	Frequency of unfolded (denatured) state resonance
ν_N	Frequency of native state resonance
[den]	Concentration of denaturant
[θ]	Mean residue ellipticity

Acknowledgments

I would like to thank my advisor Daniel Raleigh whose guidance, detailed suggestions and encouragement helped me in all the time of research and writing of this thesis. I would not accomplish all my work without his great ideas and patience. I like the opportunity to learn so many techniques in his lab and I believe I will benefit from the training I received here for my professional career.

I would like to express my gratitude to Professor Peter J. Tonge, Professor Carlos Simmerling and Professor Carlos de los Santos for serving as my committee members. I thank them for their valuable suggestions on my research projects.

I am grateful for the friendship with the past and present group-members. I had a wonderful time working with them: Sylvia Tracz, Jason Horng, Jae-hyun Cho, Andisheh Abedini, Burcu Anil, Yuefeng Tang, Ying Li, Yuan Bi, Ruchi Gupta, Bing Shan, Humeyra Taskent, Fanling Meng, Shifeng Xiao, Wenli Meng, Konstantine Aprilakis.

Last I would like to give my special thanks to my family members for their encouragement and unconditional support.

List of Publications

1. Anil, B.; **Song, B**.; Tang, Y. and Raleigh, D. P. (2004). Exploiting the right side of the Ramachandran plot: substitution of glycines by D-alanine can significantly increase protein stability. *J. Am. Chem. Soc. 126(41)*, 13194-5.

2. Brewer, S. H.; Song, B.; Raleigh, D. P. and Dyer, R. B. (2007). Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy. *Biochemistry, 46*, 3279-85.

3. Marek, P.; Abedini, A.; Song, B.; Kanungo, M.; Johnson, M. E.; Gupta, R.; Zaman, W.; Wong, S. S. and Raleigh, D. P. (2007). Aromatic interactions are not required for amyloid fibril formation by islet amyloid polypeptide but do influence the rate of fibril formation and fibril morphology. *Biochemistry, 46*, 3255-61.

4. Meng, F.; Abedini, A.; Song, B. and Raleigh, D. P. (2007). Amyloid Formation by Pro-Islet Amyloid Polypeptide Processing Intermediates: Examination of the Role of Protein Heparan Sulfate Interactions and Implications for Islet Amyloid Formation in Type 2 Diabetes. *Biochemistry, 46*, 12091-99.

5. Mareck, J; **Song, B**.; Brewer, S.; Belyea, J.; Dyer, B.; Raleigh, D. P. (2007). A Simple and Economical Method for the Production of ^{13}C, ^{18}O-labeled Fmoc-Amino Acids with High Level of Enrichment: Applications to Isotope-Edited IR Studies of Proteins. *Organic Letters,* 9, 4935-4937.

6. **Song, B**.; Cho, JH. & Raleigh, D. P. (2007). Ionic Strength Dependent Effects in Protein Folding: Analysis of Rate Equilibrium Free Energy Relationships and Their Interpretation. *Biochemistry,* 46, 14206-14214.

7. **Song, B**.; Makhatadze, G., Pappu, R. & Raleigh, D. P. (2007). Stabilization of Proteins by Replacement of Helix C-Capping Residues with D-amino acids. *To be submitted.*

1. Introduction

1.1 Protein structure

Proteins, which are polymers of L-amino acids, are very important functional molecules in biological organisms, responsible for such diverse activities as catalytic function, cell signaling and structural roles. Protein structure can be divided into four levels: primary structure, which is the amino acid sequence of the polypeptide chain; secondary structure, locally defined highly regular sub-structures; tertiary structure, which is the special three-dimensional structure for a single protein molecule; quaternary structure, which is the arrangement of several proteins molecules. To function properly, most proteins need to fold into a correct tertiary conformation although so called natively disordered proteins are an exception. We know that the protein secondary and tertiary structures are determined by the amino acid sequence, but the mechanism of folding is still a mystery (*1*).

The secondary structure of a region of a protein can be defined by the conformation of the polypeptide backbone, which is defined by the relative position in space of the connected alpha carbon, carboxyl carbon and amide nitrogen atoms. The φ and ψ angles are the dihedral angles between the atoms which are shown in figure 1-1. Not all of the φ and ψ angles are permitted for the peptide backbone. On the contrary, only a limited range is allowed because of steric clashes, and the allowed values are represented by the Ramachandran plot (figure 1-2). Since Glycine only has a hydrogen as its sidechain, it has a larger permitted range of φ and ψ angles.

α-helices and β-sheets are frequently occurring secondary structures. Both of them represent a way to forming hydrogen bonds in the peptide backbone. In an α-helix, the carbonyl oxygen of residue i forms a hydrogen bond with amide proton of residue i+4. The first and last four residues in a helix are different from the interior ones because they cannot make the intrahelical hydrogen bonds. The residues at the N-terminus of the helix are often found to be involved in side-chain to main-chain hydrogen bond interactions with residues outside of the helix while the residues at the C-terminus usually form additional hydrophobic interactions (*2, 3*). The first residue after the C-terminal residue in a helix is termed the C-capping position, and glycine is frequently found to be at this position. Glycine occupies this position because the φ and ψ values at the C-capping position are usually positive values and only glycine can easily adopt that conformation. The positive φ angle allows the glycine N-H to make two hydrogen bonds back to the helix. β-sheets are conformations of extended polypeptide chains that make hydrogen bonds with another parallel or antiparallel extended chain. The association of β-sheets is implicated in protein aggregation and fibril formation. Additional frequently observed secondary structures are turns and bends. They always occur at the position where the

direction of the main chain sharply changes. The secondary structure of a protein can be roughly estimated by far-ultraviolet circular dichroism spectroscopy. A double minimum at 208 and 222nm indicates α-helix structure whereas β-sheet structure has a single minimum at 217nm (*4*).

1.2 Protein stability

Protein stability is the free energy difference between the native state and the denatured state. In the denatured state, a protein samples a wide range of conformations and forms hydrogen bonds with solvent. There are a wide range of unfolded states. They can vary from relatively highly expanded states to states which have clusters of hydrophobic and aromatic residues and are compact, to even more structured states which contain significant secondary structure. For example, Wuthrich and co-workers showed that the denatured N-terminal region of 434-repressor protein has native-like structure even in strongly denaturing solution using NMR (*5*). Mok and coworkers also found that the N-terminal SH3 domain of drk has a compact unfolded state under physiological conditions due to specific interactions using long range NOEs (*6*). As the protein folds up, the interactions with solvent break down and other interactions within the protein build up. The main interactions are the hydrophobic effect, hydrogen bonds and salt-bridges. The energy of each interaction is small, but because of the large number of them, the total effect can be huge. However the effects are large in both the folded and unfolded states and the net stability is the small difference of two large numbers. Generally speaking, the hydrophobic effect is the tendency for nonpolar solute to separate from water molecules (*7*). It is the sum of entropic and enthalpic effects which have a different dependence on temperature. At room temperature, the entropy change is the driving force and it is negative for solvation which makes it unfavorable. Since proteins have larger solvent exposed area when they unfold, the entropy effect will favor the folded state. However at high temperature, the enthalpy change becomes the driving force. Estimates from protein thermal denaturation experiments indicate that the hydrophobic effect is the dominant force for protein folding. Salt bridges can be defined as a specific electrostatic interaction between two opposite charged residues in a protein that are located close to each other. Their contribution to protein stability is still a contentious topic (*8-14*). Some studies showed that removal of salt bridges resultes in destabilization of the protein. On the other hand some mutational studies suggest that the contribution is fairly small or even zero. One reason for this argument is that the desolvation penalty is so large that it offsets the favorable columbic interaction (*15*). Another important interaction, hydrogen bonds were thought to be the dominant force for protein folding in the early years of protein science (*8*). However, recent studies argue that hydrogen bonds do not contribute to protein stability as much as was assumed (*16*). The reason is that in the unfolded state, the peptide backbone forms hydrogen bonds with water molecules (*17, 18*). When a protein folds up, these hydrogen bonds are broken, and sometimes may not be completely compensated by the formation of intra-protein hydrogen bonds. So hydrogen bonds could

make a positive, zero or even negative contribution to protein stability. Many mutation studies apparently indicate that hydrogen bonds make a positive contribution to protein stability, however almost all of these studies do not take into account desolvation effects (*19*).

Protein stability can be measured experimentally. Most proteins unfold with increasing temperature or with the addition of denaturant, and many of them can fold back upon the reversal of these conditions. Thus there is equilibrium between the folded state and unfolded state and the change in environment can cause a shift between the different states. The Gibbs-Helmholtz equation describes the dependence of protein stability on temperature as follows:

$$\Delta G^0(T) = \Delta H^0(T_m)(1-\frac{T}{T_m}) - \Delta C_p{}^0[(T_m - T) + T\ln\frac{T}{T_m}] \qquad 1\text{-}1$$

where ΔG^0 is the free energy of unfolding, T_m is the melting temperature defined as the midpoint of thermal unfolding transition. T_m is formally defined for a monomeric protein by $\Delta G^0(T_m) = 0$. $\Delta H^0(T_m)$ is the enthalpy change upon unfolding at T_m and $\Delta C_p{}^0$ is the heat capacity. Equation 1-1 assumes that $\Delta C_p{}^0$ is independent of temperature. The thermal melting curve can be monitored by far-ultraviolet circular dichoism because both the α-helix and the β-sheet CD signals differ from the CD spectrum of an unfolded protein. The plot of CD signal versus temperature can be fit to the following equation:

$$\theta = \frac{(a_n + b_n T) + (a_d + b_d T)e^{-(\Delta G^0(T))/RT)}}{1 + e^{-(\Delta G^0 T))/RT}} \qquad 1\text{-}2$$

By using the Gibbs-Helmholtz equation to define $\Delta G^0(T)$, we can calculate protein stability. The fits give reliable values of T_m and $\Delta H^0(T_m)$ but not$\Delta C_p{}^0$ since it is related to the second derivative of the θ vs T plot.

The equilibrium between two states can also be perturbed by adding denaturant. In many cases, the free energy of unfolding has been found to be proportional to the concentration of the denaturant described by the following equation:

$$\Delta G^0 = \Delta G^0(H_2O) - m[denaturant] \qquad 1\text{-}3$$

where $\Delta G^0(H_2O)$ is the free energy of unfolding in the absence of denaturant and m is a constant for each protein which is proportional to the change of the exposed surface area upon unfolding (*20, 21*).

Thermal and chemical denaturation experiments monitored by CD can give the global stability of the protein, while H/D exchange measurements monitored by NMR spectroscopy can give more detailed information about each residue (*22, 23*). After dissolving the protein in D_2O, amide protons will exchange with the D_2O solvent. Thus

the signals from the amide protons will decay and finally disappear. The exchange rate can be calculated by fitting the normalized peak volumes to a first order exponential decay:

$$V(t) = V_0 e^{-(k_{ex}t)} + b \qquad 1\text{-}4$$

Where V(t) is normalized peak volume, k_{ex} is the exchange rate and b is a baseline correction term. The amide protons from solvent exposed residues will exchange quickly with D_2O, so their peaks decay fast. However, the amide protons from buried residues need to be exposed to solvent via unfolding of the protein before they can exchange with D_2O. So the exchange rates for those buried residues should be slower than the exposed residues. By comparing the exchange rates to the intrinsic amide proton exchange rate constants (k_{ch}), which is calculated from denatured peptide models, the free energy of the opening reaction (ΔG^0_{ex}) for the individual residue can be determined from the following equation:

$$\Delta G^0_{ex} = -RT \ln \frac{k_{ex}}{k_{ch}} \qquad 1\text{-}5$$

Besides the stability of individual residues, this method can also tell us which residues are on the surface and which residues are buried in the hydrophobic core. Equation 1-5 assumes the so-called EX2 limit (*24, 25*), in which the rate of folding is much faster than k_{ch}.

1.3 Protein folding

After being synthesized on the ribosome, proteins fold into their native structure on the order of seconds or even less for small proteins. Many diseases are related to the misfolding of proteins, such as Alzheimer's disease, Parkinson's disease and type II diabetes (*26-31*). Understanding of the mechanism of protein folding is very important to solve the mechanism of these diseases and find a way to cure them. Previous studies have shown that the protein folding process can not be a random search through all possible conformations because in that case it may take a huge amount of time (*32*). Several models have been proposed to understand the mechanism of protein folding. Many studies have suggested that protein folding is driven by the hydrophobic force. There are several different models of protein folding. Examples include the diffusion-collision model and the nucleation model (*33-41*). None of these models can explain the folding mechanism completely. How proteins fold to the correct structure by the information encoded in the primary sequence is still an unsolved problem.

Recently the 'funnel model' has been proposed as a new view of the folding mechanism (Figure 1-3) (*42-45*). The vertical axis represents the total free energy minus the contribution due to conformational entropy, $G\text{-}T\Delta S_{conformational}$. The native state is located at the bottom of the funnel. The width of the funnel is related to the conformational entropy of the protein. At the top of the funnel, the protein exists in a

large number of conformations with high enthalpy and entropy. For each protein the shape of the funnel is specific which is determined by the amino acids sequence. Instead of having completely smooth sides, the funnel has some roughness along the sides which gives rise to some local minima and maxima. An important point of this 'funnel model' is that the protein folding process can have many parallel pathways instead of one single pathway to reach the bottom of the funnel, the native state.

Two-state folding is the simplest mechanism, which many small single-domain globular proteins may adopt. In two-state folding, each pathway has only two minima, the unfolded state and native state, and one saddle point, the transition state, without any accumulation of intermediates. Cooperative folding and unfolding are very important features of two-state mechanism. However, for some proteins, in addition to the two minima mentioned above, other minima exist which correspond to the intermediate states. Folding pathways with one or more intermediates are called multi-state folding.

Folding kinetics can be studied by the stopped-flow method, which can determine the protein folding and unfolding rates, for reactions which occur on the order of milliseconds or slower. The folding and unfolding rates are found to be proportional to the denaturant concentration as follows:

$$\ln k_f([den] = \ln k_f(H_2O) - \frac{m_f}{RT}[den] \qquad 1\text{-}6$$

$$\ln k_u([den]) = \ln k_u(H_2O) + \frac{m_u}{RT}[den] \qquad 1\text{-}7$$

where m_f and m_u are constants of proportionality. The m_f–value is proportional to the difference of the surface area between the unfolded state and transition state, while m_u is related to the difference between the transition state and native state . The observed rate constant (k_{obs}) is the sum of k_f and k_u. If the protein undergoes two-state folding, the plot of $\ln k_{obs}$ versus the concentration of denaturant will show a V-shape, which is called a Chevron plot, because of the linear dependence of $\ln k_f$ and $\ln k_u$ on the denaturant concentration. However, if there are intermediates accumulating in the folding process, the plot will show curvature at low denaturant concentration (*46*). Thus a Chevron plot is a good method to determine if a protein undergoes two-state folding mechanism. The equilibrium m-value can be calculated by subtracting m_f from m_u. The Tanford β-value, β_T or θ_m, is equal to $m_f/(m_u-m_f)$ and represents the extent of surface buried in the transition state. For example, a β_T value of 0.5 means that 50% of the surface area burial that occurs upon folding is buried in the transition state.

1.4 ***Opportunities for using non-coded amino acids in protein design and in protein folding studies: D-Amino acids***

Gly with a positive φ-angle is commonly found at the C-Capping position at the end of an α-helix. A Gly with positive φ-angles allows the carbonyl of the Gly to form

bifrucated hydrogen bond to the two C-terminal amide hydrogens of the helix. Gly is the most flexible amino acid in the sense that it can adopt the most conformations in the unfolded state. So it leads to higher entropy than other amino acids in the unfolded state. In the native state, since each residue is fixed in a specific position, the difference between the entropy of Gly and other amino acids is not as big as in the unfolded state. Thus mutating Gly to other amino acids should decrease the entropy of the unfolded state, and enhance protein stability. However, previous studies showed that substitution of Gly with other amino acids often destabilized the protein (*47, 48*). This might because of the strain introduced in the native structure by the substitution since non-glycine L-amino acids with positive φ-angels are energetically unfavorable. Thus replacement of Gly which have positive φ-angles by D-amino acids could stabilize the protein by decreasing the entropy of the unfolded state without introducing any strain into the native structure. Some studies have been done that incorporate D-amino acid into β-turns in hairpins, showing that D-amino acids at the (i+2) position can promote the type II' β-turn formation (*49-51*). However no thermodynamic data was presented. Here, in this dissertation, thermodynamic and kinetic data will be presented for D-amino acid substitutions.

Since D-amino acids are non-coded amino acids, the traditional method of bacterial expression can not be used to generate the proteins containing D-amino acids. Fortunately, recent advances in solid peptide synthesis and expressed protein ligation make it possible to incorporate non-coded amino acids to protein (*52-54*). So replacement of Gly which have positive φ-angles by D-amino acids can be a new strategy for protein design to enhance protein stability.

1.5 Electrostatic interactions and ionic strength effects on protein folding and stability

Since proteins are ployampholytes, electrostatic interactions among the charged moieties can contribute significantly to their stabilities and folding dynamics. However, these electrostatic interactions can be shield by the ions in solvent. Thus protein folding and stability can be strongly dependent on the solvent ionic strength. According to Debye-Huckel limiting law, both the stability and log of the folding rate will have a linear relationship to the square root of ionic strength (*55*). However, usually not all charge-charge interactions will be equally screened by salt even if they lie on the surface of the protein (*56*). And sometimes the chaotropic or kosmotropic behavior of added salts can strongly affect protein stability and folding rate (*57*). Thus the effects of ionic strength on protein folding and stability are complicated and it is difficult to deduce the relationship. In this dissertation, an analysis of ionic strength dependent effects was described using Rate Equilibrium Free Energy Relationships.

1.6 Proteins investigated in this dissertation

The chicken villin headpiece domain is a 76 residue, α-helical F-actin binding protein (*58*). The first 9 residues can be removed without affecting function or structure

of the protein (*58, 59*). The protein containing residues 10 to 76 is usually used for studies and is referred as HP67. HP67 consists of two subdomains, N-terminal subdoamin and C-terminal subdomain (figure 1-4). The C-terminal subdomain contains 35 residues, and is denoted HP35. A Met is incorporated at the N-terminus of the recombinant HP35 and the resulting construct is referred as HP36 (figure 1-4). HP36 contains three α-helices linked by turns and loops and can fold independently to a tertiary structure which is similar to the C-terminal domain of the intact HP67. HP36 is one of the smallest proteins which folds cooperatively and ultrafast via a two state mechanism. Thus makes it an attractive model for studying folding process (*60-65*).

The N-terminal subdomain of HP67 can not fold independently to its native structure in isolation. In intact HP67, the N-terminal subdomain has several loops, turns and a short α-helix, forming an extended hydrophobic core with the C-terminal subdomain. So it is obvious that interactions between the two subdomains are important for the protein to fold to the native conformation. Although HP35 can fold independently, it can not function properly without the N-terminal domain. Glu39 and Lys70 are found to be close to each other and both residues are buried in the hydrophobic core of HP67. Thus HP67 provides a good model to study the contribution of buried salt bridges to protein folding and stability.

For the protein design studies using D-amino acids, several small α-helical proteins have been probed. They are ubiquitin binding associated domain (UBA), the peripheral subunit-binding domain (PSBD), the GA module of an albumin binding domain (GA) and an engrailed homeodomain(*66*). UBA contains 45 residues and consists of three α-helices folded around a hydrophobic core. The PSBD domain is 41 residues, and has a simple topology, comprising two parallel α-helices and a short 3_{10}-helix. The GA protein is a 47-residue, three-helix bundle protein, while the EH domain is a 59-residue, three-helix bundle protein. All of these are fast-folding helix proteins and each has a glycine locating at a helix C-capping position. Ribbon diagrams of the structures are shown in figure 1-5 with the Ramachandran plot showing the φ and ψ angles of the glycines.

The N-terminal domain of ribosomal protein L9 (NTL9) is a small protein which can fold independently and the folding mechanism has been shown to be two-state (*67*). It has a simple topology, containing a three-stranded β-sheet sandwiched between the central α-helix and a short α-helix (figure 1-6). NTL9 lacks disulfide bonds and the folding process does not require any metal or cofactor bindings, making it an attractive target for folding studies. In addition, it has multiple charged residues on its surface which contribute to the biological function of the protein. Thus it is an ideal model to study the role of the electrostatic interactions in stability and folding kinetics.

1.7 The aims of this dissertation

This dissertation describes studies on protein design to enhance protein stability by incorporating non-coded amino acids, as well as studies of the contribution of

electrostatic interactions to protein stability and folding kinetics. Gly with positive φ-angles at C-capping positions have been replaced by D-Ala to decrease the entropy of unfolded state. This work is described in chapter 2. The studies on the contribution of a buried salt-bridge in HP67 to stability and folding dynamics are summarized in chapter 3. The results showed that the protein can not fold to the native structure without the salt-bridge. The contribution of electrostatic interactions to protein folding and stability has been studied in NTL9 by changing the solvent ionic strength. The effect is quite complicated and the results are described in chapter 4. The development of electrostatic interactions in transition state was also studied and compared to previous studies. The final chapter summarizes efforts in the synthesis of specifically labeled variants of HP36. These proteins were prepared as part of a collaboration project with Dr. R. Brian Dyer of Los Alamos National Laboratory.

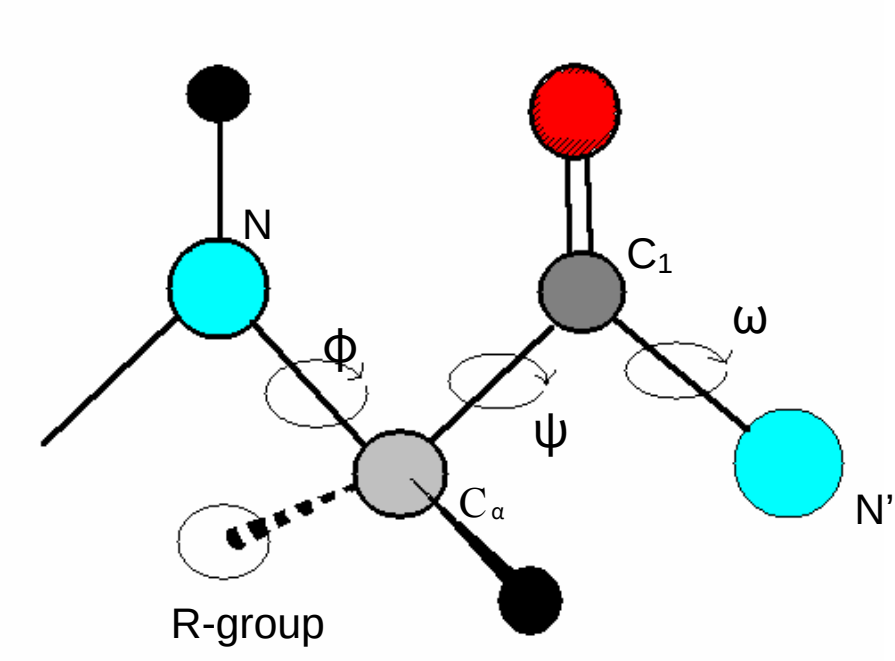


Figure 1-1 Definitions of the torsional angles φ and ψ.

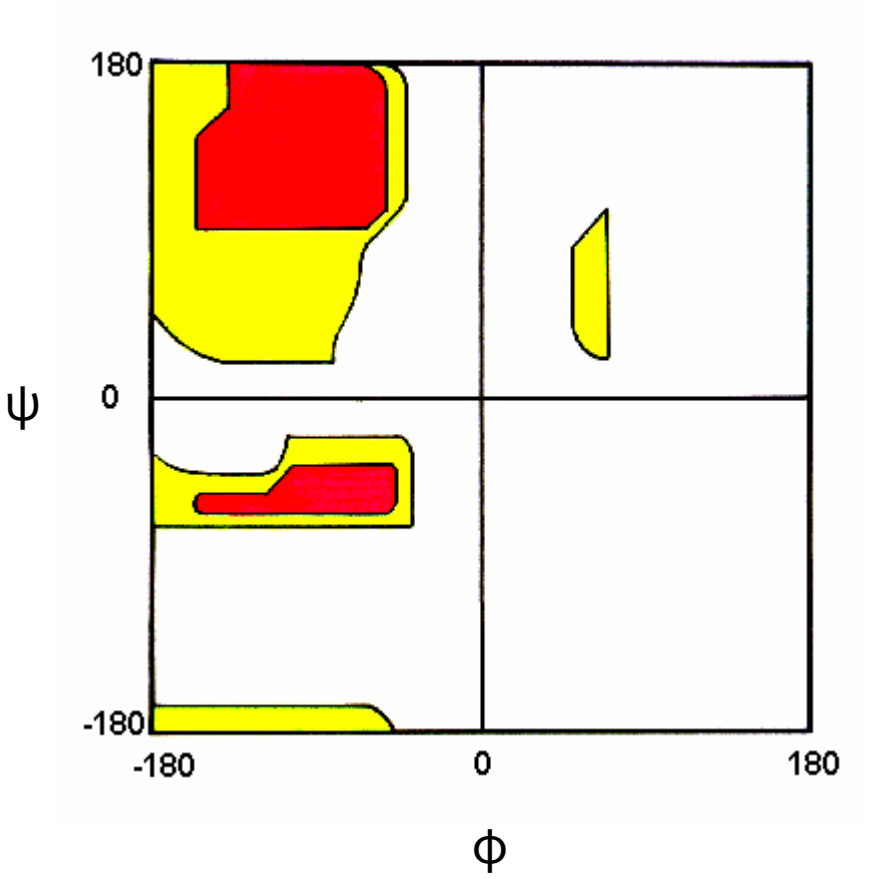


Figure 1-2 The Ramachandran plot for an L-amino acid illustrating the preferred conformations of the polypeptide chain. The red area is the highly preferred region; the yellow area is the allowed region; the other areas represent conformations which are higher in free energy.

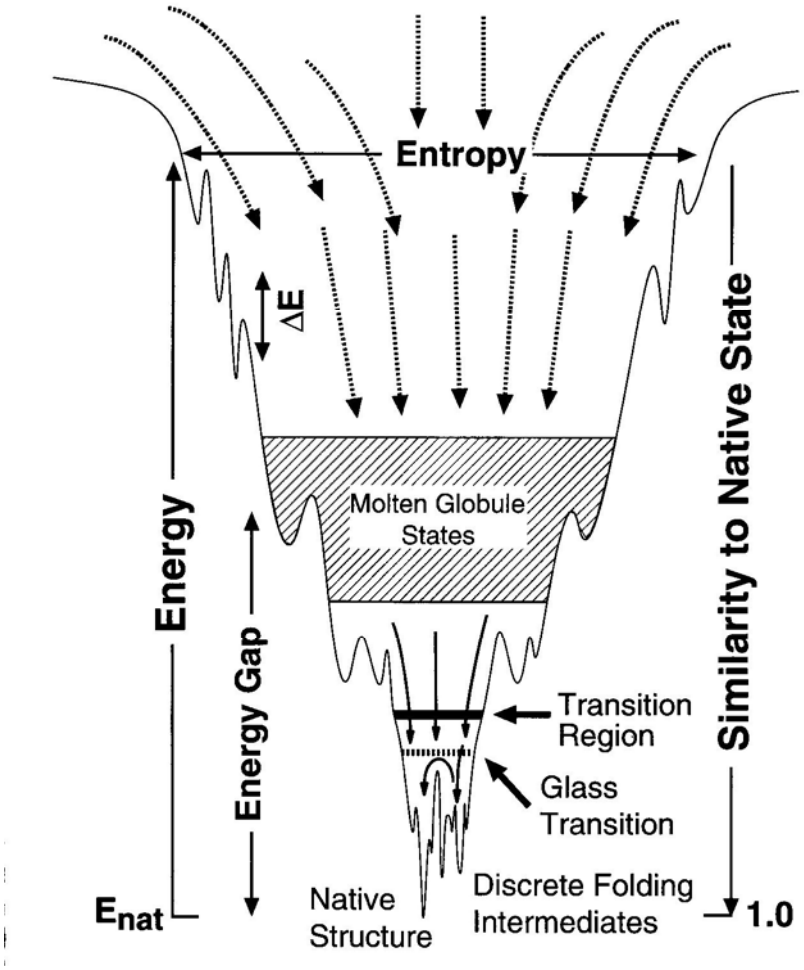
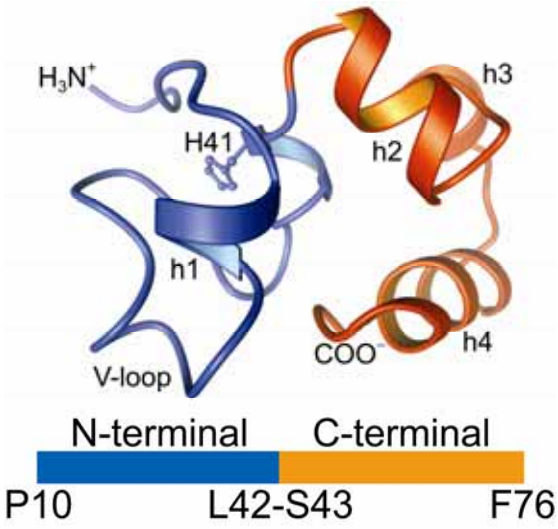


Figure 1-3 Cross section through a folding funnel (figure adapted from Wolynes) (*45*).

a



b

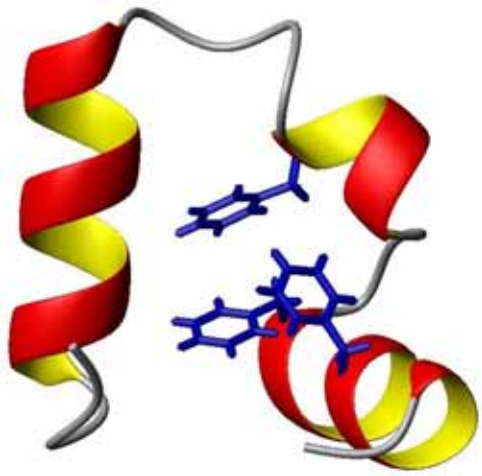


Figure 1-4 Ribbon diagram of the villin headpiece (a) HP67; the C-terminal subdomain is colored orange and the N-terminal subdomain is colored blue. (b) HP36; the orientation of figure (a) and (b) are different. The figures were created using the PDB files 1QQV and 1VII and the program MOLMOL.

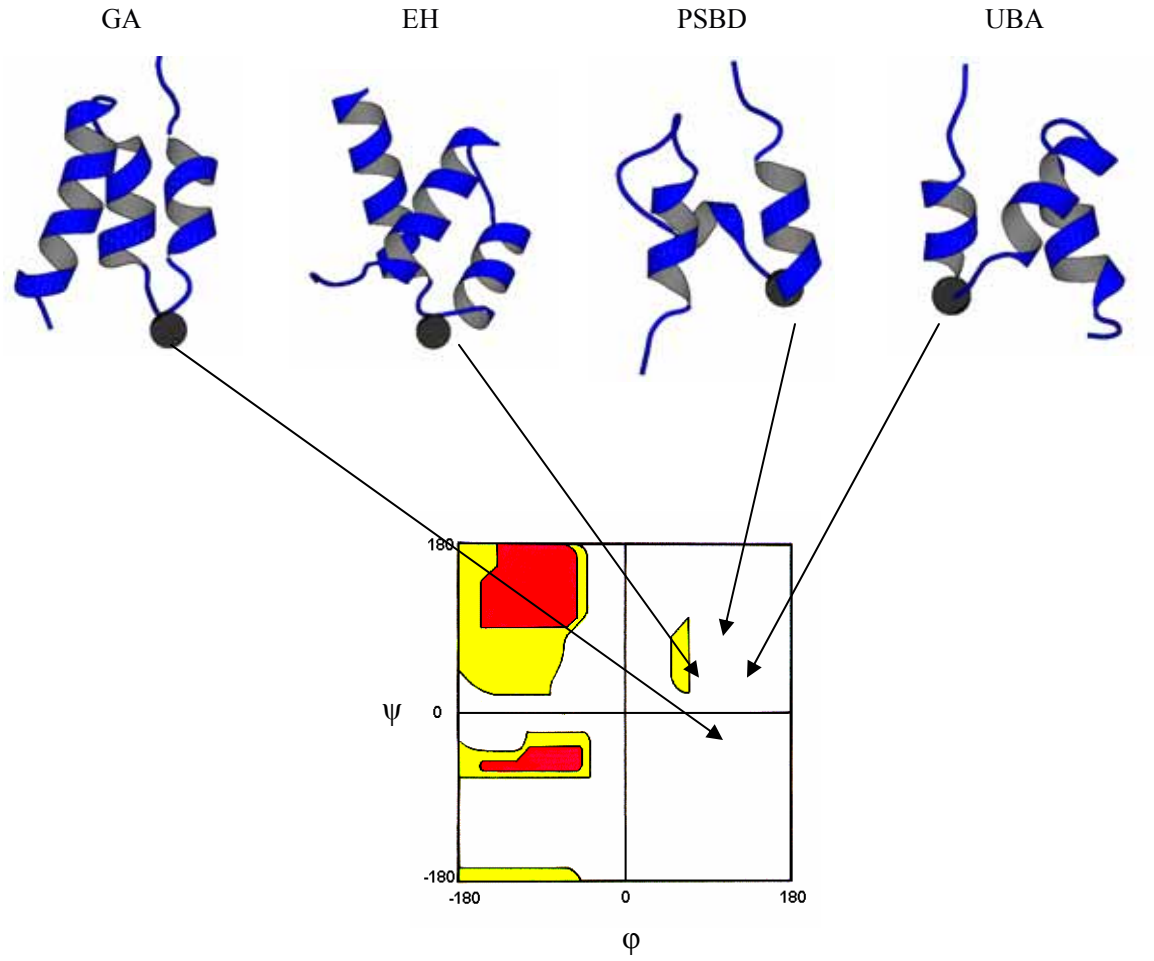


Figure 1-5 Ribbon diagram of GA, EH, PSBD and UBA. The glycine at the helix C-capping position is shown as a sphere. The arrow represents the respective φ and ψ values for these glycines. The figure was generated using the program Molscript and the PDB files 2PDD, 1ENH, 1PRB and 1UBA.

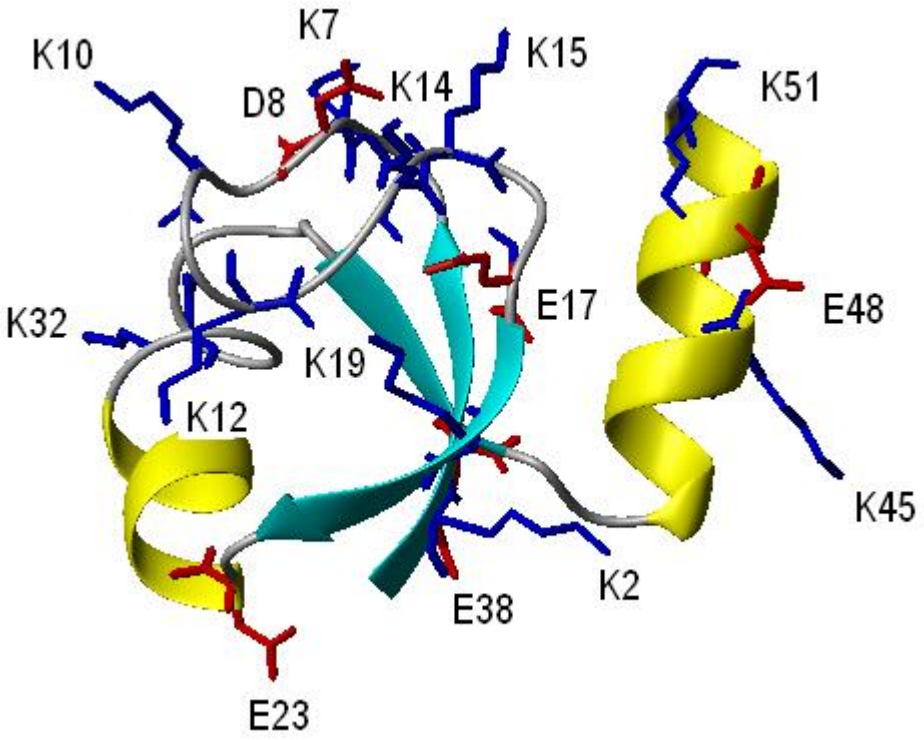


Figure 1-6 A ribbon diagram of NTL9. The acidic and basic residues are shown in stick format. The diagram was constructed using the pdb file 2HBB and the program MolMol. The last few residues of NTL9 are disordered in the X-ray structure. Thus the position of Glu54 and the C-terminus are not shown.

1.7 References

1. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science 181*, 223-230.
2. Richardson, J. S., and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of alpha helices, *Science 240*, 1648-1652.
3. Aurora, R., Srinivasan, R., and Rose, G. D. (1994) Rules for alpha-helix termination by glycine, *Science 264*, 1126-1130.
4. Compton, L. A., and Johnson, W. C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication, *Anal. Biochem. 155*, 155-167.
5. Neri, D., Billeter, M., Wider, G., and Wuthrich, K. (1992) NMR determination of residual structure in a urea-denatured protein, the 434-repressor, *Science 257*, 1559-1563.
6. Mok, Y. K., Kay, C. M., Kay, L. E., and Forman-Kay, J. (1999) NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions, *J. Mol. Biol. 289*, 619-638.
7. Chandler, D. (2005) Interfaces and the driving force of hydrophobic assembly, *Nature 437*, 640-647.
8. Mirsky, A. E., and Pauling, L. (1936) On the structure of native, denatured, and coagulated proteins, *Proc. Natl. Acad. Sci. U S A 22*, 439-447.
9. Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A. R. (1990) Estimating the contribution of engineered surface electrostatic interactions to protein stability by using double-mutant cycles, *Biochemistry 29*, 9343-9352.
10. Oliveberg, M., and Fersht, A. R. (1996) Formation of electrostatic interactions on the protein-folding pathway, *Biochemistry 35*, 2726-2737.
11. Cho, J. H., and Raleigh, D. P. (2006) Electrostatic interactions in the denatured state and in the transition state for protein folding: effects of denatured state interactions on the analysis of transition state structure, *J. Mol. Biol. 359*, 1437-1446.
12. Luisi, D. L., Snow, C. D., Lin, J. J., Hendsch, Z. S., Tidor, B., and Raleigh, D. P. (2003) Surface salt bridges, double-mutant cycles, and protein stability: An experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein L9, *Biochemistry 42*, 7050-7060.
13. Sanchez-Ruiz, J. M., and Makhatadze, G. I. (2001) To charge or not to charge?, *Trends Biotechnol. 19*, 132-135.
14. Perl, D., and Schmid, F. X. (2001) Electrostatic stabilization of a thermophilic cold shock protein, *J. Mol. Biol. 313*, 343-357.
15. Hendsch, Z. S., and Tidor, B. (1994) Do salt bridges stabilize proteins? A continuum electrostatic analysis, *Protein Sci. 3*, 211-226.

16. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation, *Adv Protein Chem. 14*, 1-63.
17. Yang, A. S., and Honig, B. (1995) Free energy determinants of secondary structure formation: I. alpha-Helices, *J. Mol. Biol. 252*, 351-365.
18. Baldwin, R. L. (2003) In search of the energetic role of peptide hydrogen bonds, *J. Biol. Chem. 278*, 17581-17588.
19. Myers, J. K., and Pace, C. N. (1996) Hydrogen bonding stabilizes globular proteins, *Biophys. J. 71*, 2033-2039.
20. Tanford, C. (1964) Isothermal unfolding of globular proteins in aqueous urea solutions, *J. Am. Chem. Soc. 86*, 10.
21. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138-2148.
22. Hoshino, M., Katou, H., Hagihara, Y., Hasegawa, K., Naiki, H., and Goto, Y. (2002) Mapping the core of the beta(2)-microglobulin amyloid fibril by H/D exchange, *Nat. Struct. Biol. 9*, 332-336.
23. Whittemore, N. A., Mishra, R., Kheterpal, I., Williams, A. D., Wetzel, R., and Serpersu, E. H. (2005) Hydrogen-deuterium (H/D) exchange mapping of Abeta 1-40 amyloid fibril secondary structure using nuclear magnetic resonance spectroscopy, *Biochemistry 44*, 4434-4441.
24. Vugmeyster, L., Kuhlman, B., and Raleigh, D. P. (1998) Amide proton exchange measurements as a probe of the stability and dynamics of the N-terminal domain of the ribosomal protein L9: comparison with the intact protein, *Protein Sci. 7*, 1994-1997.
25. Hvidt, A., and Nielsen, S. O. (1966) Hydrogen exchange in proteins, *Adv Protein Chem. 21*, 287-386.
26. Selkoe, D. J. (2004) Cell biology of protein misfolding: The examples of Alzheimer's and Parkinson's diseases, *Nat. Cell Biol. 6*, 1054-1061.
27. Aguzzi, A., and Haass, C. (2003) Games played by rogue proteins in prion disorders and Alzheimer's disease, *Science 302*, 814-818.
28. Dauer, W., and Przedborski, S. (2003) Parkinson's disease: Mechanisms and models, *Neuron 39*, 889-909.
29. Jaikaran, E. T., and Clark, A. (2001) Islet amyloid and type 2 diabetes: from molecular misfolding to islet pathophysiology, *Biochim. Biophys. Acta. 1537*, 179-203.
30. Dobson, C. M. (2004) Experimental investigation of protein folding and misfolding, *Methods 34*, 4-14.
31. Dumoulin, M., and Dobson, C. M. (2004) Probing the origins, diagnosis and treatment of amyloid diseases using antibodies, *Biochimie 86*, 589-600.
32. Levinthal, C. (1968) Are there pathways for protein folding?, *J. Chim. Phys. 65*, 2.

33. Karplus, M., and Weaver, D. L. (1994) Protein folding dynamics: The diffusion-collision model and experimental data, *Protein Sci. 3*, 650-668.
34. Wetlaufer, D. B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins, *Proc. Natl. Acad. Sci. U S A 70*, 697-701.
35. Dill, K. A. (1985) Theory for the folding and stability of globular proteins, *Biochemistry 24*, 1501-1509.
36. Dill, K. A. (1990) Dominant forces in protein folding, *Biochemistry 29*, 7133-7155.
37. Dill, K. A. (1990) The meaning of hydrophobicity, *Science 250*, 297-298.
38. Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D., and Chan, H. S. (1995) Principles of protein folding--a perspective from simple exact models, *Protein Sci. 4*, 561-602.
39. Rackovsky, S., and Scheraga, H. A. (1977) Hydrophobicity, hydrophilicity, and the radial and orientational distributions of residues in native proteins, *Proc. Natl. Acad. Sci. U S A 74*, 5248-5251.
40. Fersht, A. R. (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications, *Proc. Natl. Acad. Sci. U. S. A. 92*, 10869-10873.
41. Fersht, A. R. (1997) Nucleation mechanisms in protein folding, *Curr. Opin. Struct. Biol. 7*, 3-9.
42. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis, *Proteins 21*, 167-195.
43. Leopold, P. E., Montal, M., and Onuchic, J. N. (1992) Protein folding funnels: A kinetic approach to the sequence-structure relationship, *Proc. Natl. Acad. Sci. U. S. A. 89*, 8721-8725.
44. Shoemaker, B. A., Wang, J., and Wolynes, P. G. (1999) Exploring structures in protein folding funnels with free energy functionals: the transition state ensemble, *J. Mol. Biol. 287*, 675-694.
45. Wolynes, P. G., Onuchic, J. N., and Thirumalai, D. (1995) Navigating the folding routes, *Science 267*, 1619-1620.
46. Horng, J. C., Tracz, S. M., Lumb, K. J., and Raleigh, D. P. (2005) Slow folding of a three-helix protein via a compact intermediate, *Biochemistry 44*, 627-634.
47. Karplus, P. A. (1996) Experimentally observed conformation-dependent geometry and hidden strain in proteins, *Protein Sci. 5*, 1406-1420.
48. Nicholson, H., Soderlind, E., Tronrud, D. E., and Matthews, B. W. (1989) Contributions of left-handed helical residues to the structure and stability of bacteriophage T4 lysozyme, *J. Mol. Biol. 210*, 181-193.
49. Imperiali, B., Moats, R. A. Fisher, S. L. & Prins, T. J. (1992) A conformational study of peptides with the general structure Ac-L-Xaa-Pro-D-Xaa-L-Xaa-NH_2: spectroscopic evidence for a peptide with significant beta-turn character in water

and in dimethyl sulfoxide, *J. Am. Chem. Soc. 114*, 3182-3188.

50. Richardson, J. S., Richardson, D. C., Tweedy, N. B., Gernert, K. M., Quinn, T. P., Hecht, M. H., Erickson, B. W., Yan, Y., McClain, R. D., Donlan, M. E., et al. (1992) Looking at proteins: representations, folding, packing, and design. Biophysical Society National Lecture, 1992, *Biophys. J. 63*, 1185-1209.
51. Haque, T. S., & Gellman, S. H. (1997) Insights on β-hairpin stability in aqueous solution from peptides with enforced type I' and type II' b-turns, *J. Am. Chem. Soc. 119*, 2303-2304.
52. Schnolzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S. B. (1992) In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences, *Int. J. Pept. Protein Res. 40*, 180-193.
53. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) Synthesis of proteins by native chemical ligation, *Science 266*, 776-779.
54. Muir, T. W. (2003) Semisynthesis of proteins by expressed protein ligation, *Annu. Rev. Biochem. 72*, 249-289.
55. de Los Rios, M. A., and Plaxco, K. W. (2005) Apparent Debye-Huckel electrostatic effects in the folding of a simple, single domain protein, *Biochemistry 44*, 1243-1250.
56. Lee, K. K., Fitch, C. A., and Garcia-Moreno, E. B. (2002) Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein, *Protein Sci. 11*, 1004-1016.
57. Broering, J. M., and Bommarius, A. S. (2005) Evaluation of Hofmeister effects on the kinetic stability of proteins, *J. Phys. Chem. B 109*, 20612-20619.
58. Vardar, D., Buckley, D. A., Frank, B. S., and McKnight, C. J. (1999) NMR structure of an F-actin-binding "headpiece" motif from villin, *J. Mol. Biol. 294*, 1299-1310.
59. McKnight, C. J., Doering, D. S., Matsudaira, P. T., and Kim, P. S. (1996) A thermostable 35-residue subdomain within villin headpiece, *J. Mol. Biol. 260*, 126-134.
60. De Mori, G. M., Colombo, G., and Micheletti, C. (2005) Study of the Villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics, *Proteins 58*, 459-471.
61. Duan, Y., Wang, L., and Kollman, P. A. (1998) The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation, *Proc. Natl. Acad. Sci. U. S. A. 95*, 9897-9902.
62. Fernandez, A., Shen, M. Y., Colubri, A., Sosnick, T. R., Berry, R. S., and Freed, K. F. (2003) Large-scale context in protein folding: villin headpiece, *Biochemistry 42*, 664-671.
63. Jang, S., Kim, E., Shin, S., and Pak, Y. (2003) Ab initio folding of helix bundle proteins using molecular dynamics simulations, *J. Am. Chem. Soc. 125*, 14841-14846.

64. Shen, M. Y., and Freed, K. F. (2002) All-atom fast protein folding simulations: the villin headpiece, *Proteins 49*, 439-445.
65. Zagrovic, B., Snow, C. D., Khaliq, S., Shirts, M. R., and Pande, V. S. (2002) Native-like mean structure in the unfolded ensemble of small proteins, *J. Mol. Biol. 323*, 153-164.
66. Kalia, Y. N., Brocklehurst, S. M., Hipps, D. S., Appella, E., Sakaguchi, K., and Perham, R. N. (1993) The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of Bacillus stearothermophilus, *J. Mol. Biol. 230*, 323-341.
67. Kuhlman, B., Boice, J. A., Fairman, R., and Raleigh, D. P. (1998) Structure and stability of the N-terminal domain of the ribosomal protein L9: Evidence for rapid two-state folding, *Biochemistry 37*, 1025-1032.

2. Stabilization of Proteins by Replacement of Helix C-Capping Residues with D-amino acids

Abstract

A new strategy to enhance protein stability is described in this chapter. Replacement of helix C-capping glycines which have positive φ angles by D-Ala is shown to increase protein stability. A series of small helical proteins, peripheral subunit-binding domain, the ubiquitin binding associated domain, the GA module of an albumin binding domain and an engrailed homeodomain have been targeted. The substitution stabilized the proteins by 0.6 to 1.2 kcal/mol. Such substitutions can decrease the conformational entropy of the unfolded state without introducing any strain into the native structure. Native state strain is the main source of the destabilization observed in previous studies involving L-amino acids. Analysis of the folding and unfolding rates suggested the stabilization mainly comes from effects upon the unfolded state. Differential desolvation of the peptide backbone upon a Gly to D-Ala substitution might also contribute to the change in protein stability, but the effect is modest when the residue is fully exposed on surface. D-Val and D-Ala variants of PSBD were compared to test for desolvation effects. The results showed that the backbone desolvation effect has an unfavorable effect on protein stability but it is not as large as the favorable entropy effect.

2.1 Introduction

An important goal of protein design is to enhance protein stability. The loss of conformational entropy upon folding is one of the main unfavorable features that needs to be overcome to reach the native state. In principle, protein stability can be enhanced by destabilizing the unfolded state by decreasing its conformational entropy (*1, 2*). Glycine experiences the fewest local conformational restrictions and thus is expected to suffer the largest loss of backbone conformational entropy upon folding. Hence replacing glycines by amino acids with a side chain is a potentially useful strategy to enhance stability (*1, 3, 4*). However, glycines often adopt conformations with positive φ-angles which are unfavorable for L-amino acids (*5*). Hence substitution by L-amino acids can introduce significant strain into the native state. In this case the decrease of unfolded state entropy can be partially offset by strain in the native state, and the substitution typically results in only a modest increase or even a decrease in stability (*1, 6, 7*). Changes in solvation caused by introduction of the new methyl group can also contribute to changes in stability and could partially obscure the expected entropic effects (*8*). Addition of a sidechain will

partially shield the peptide backbone and this will alter the free energy of the chain. This effect will only make a contribution to ΔG^0 of folding if the effect is different in the folded and unfolded states. Consider first, a glycine whose peptide group is fully exposed to solvent in the native state. Addition of a sidechain will change the solvent exposure of this site by the same amount in the folded and unfolded states and solvation effects will not contribute to ΔG^0 of folding. In contrast consider a glycine which is buried in the folded state or partially buried. Addition of the sidechain will have a small effect upon native state solvation but will have a significant effect upon solvation in the unfolded state. In this case solvation effects will make a contribution to ΔG^0 of folding. An attractive alternative strategy is to substitute glycines that adopt positive values of φ by D-amino acids since D-amino acids can adopt positive φ-angles without introducing strain in the native state. Recent advances in solid phase synthesis and native ligation methodology make it possible to incorporate D-amino acids to proteins (*9-11*).

Glycines with positive φ-angles are often found at C-capping sites of helices and in type I' and II' β-turns (*8, 12, 13*). Previous studies have examined the consequences of a D-Ala for Gly substitution in the protein NTL9 at two non-helix capping sites and found a stability increase ranging from 1.30 to 1.87 kcal/mol, while in ubiquitin virtually no significant change in stability was observed for a D-Ala variant (*2*). Thus, while it is clear that substitution of glycines with positive φ-angles by D-Ala can lead to increases in stability, it is not at all clear what the magnitude of the effect is likely to be or why it appears strongly context dependent. In order to address these issues I studied the effect of replacing C-capping glycines by D-Ala, in four globular proteins. The peripheral subunit-binding domain (PSBD), the ubiquitin binding associated domain (UBA), the GA module of an albumin binding domain (GA) and an engrailed homeodomain (EH) (*14*), were chosen for these studies. The PSBD domain is 41 residues, and has a simple topology, comprising two parallel α-helices and a short 3_{10}-helix (*14*). UBA contains 45 residues and consists of three α-helices folded around the hydrophobic core (*15*). The GA protein is a 47-residue, three-helix bundle protein, while the EH domain is a 59-residue, three-helix bundle protein (*16, 17*). All four proteins are small fast-folding helix proteins and each has a glycine at a helix C-capping position (*18-23*). Ribbon diagrams of the structures are shown in figure 2-1. The location of the glycines is given in table 2-1.

2.2 Materials and Methods

2.2.1 Peptide synthesis and purification

The UBA, GA, PSBD and EH domains and their D-Ala or D-Val variants were all prepared by solid phase synthesis using a ABI 433 A peptide synthesizer and standard Fmoc chemistry. The first residue linked to the resin, all the residues with a β-branch sidechain and the following residues were double-coupled. All peptides have a free N-terminus and amidated C-terminus. Peptides were cleaved from the PAL-PEG-PS resin with a 91% trifluoroacetic acid (TFA), 3%anisole, 3%thioanisole, 3%ethanedithiol

mixture. Purification was achieved by HPLC using a water/acetonitrile gradient containing 0.1% (v/v) trifloroacetic acid on a C18 column (Vydac) with a flow rate of 3 to 5 ml/min. The retention times for the preparative HPLC are listed in table 2-2. Peptide identity was confirmed by mass spectrum and all peptides were greater than 95% pure.

2.2.2 Circular dichroism spectroscopy (CD)

All CD experiments were performed on Aviv model 62A DS and 202SF circular dichroism spectrophotometers equipped with Peltier temperature control units. A 50mM phosphate buffer with 1mM DTT at pH 6.5 was used for UBA wildtype and its D-Ala mutant. For the GA wildtype and its D-Ala mutant, a 50mM phosphate buffer with pH 7.0 was used. For PSBD wildtype and its D-Ala and D-Val mutants, all CD experiments were performed using a buffer containing 2mM sodium phosphate, 2mM sodium borate and 50mM sodium chloride at pH 8.0. For EH wildtype and its D-Ala mutant, a 50mM sodium acetate buffer with pH 5.7 was used. All the wavelength scans were performed in a 1mm quartz cuvette from 195 nm to 260 nm with 5 repeats and an averaging time of 1 second at each wavelength. All of the thermal denatuaration and urea/guanidine denaturation experiments were monitored at 222 nm in a 1cm cuvette. The temperature range for thermal denaturation was from 2 ^{o}C to 98 ^{o}C. Urea / guanidine denaturation experiments were carried out with a titrator unit interfaced to the CD spectrometer. The peptide concentrations were determined by UV spectroscopy and denaturant concentrations by measuring the index of refraction. The peptide concentrations in all samples are around 20 μM.

2.2.3 Denaturation curve fitting

A two-state model was used to fit the urea/guanidine denaturation curves to determine the stability of peptides. The measured CD signals can be fit to the following equation to determine thermodynamic parameters:

$$y([den]) = \frac{a_n + b_n[den] + (a_d + b_d[den])e^{-\Delta G^0_{N-D}([den])/RT}}{1 + e^{-\Delta G^0_{N-D}([den])/RT}} \qquad (2\text{-}1)$$

where,

$$\Delta G^0([den]) = \Delta G^0(H_2O) - m[den] \qquad (2\text{-}2)$$

y is the measured ellipticity. a_n, b_n, a_d, b_d are the parameters that define the signals of the native state (N) and denatured state (D). $\Delta G^0([den])$ is the free energy change upon unfolding with denaturant and $\Delta G^0(H_2O)$ is free energy change in the absence of denaturant.

Thermal denaturation curves were fitted to similar equations but with $\Delta G^0([den])$ changed to $\Delta G^0(T)$. The Gibbs-Helmholtz equation gives the temperature dependence of the Gibbs free energy of unfolding:

$$\Delta G^0_{N-D}(T) = \Delta H^0(T_m) - T\Delta S^0(T_m) + \Delta C^0_p(T - T_m - T\ln\frac{T}{T_m}) \quad (2\text{-}3)$$

where T_m is the midpoint temperature. $\Delta H^0(T_m)$ is the enthalpy change for unfolding. $\Delta C_p{}^0$ is the change in heat capacity, which is generally considered to be proportional to the difference of the non-polar surface area between the native and denaturated states. T_m can be determined by numerical differentiation of the denaturation curve.

2.2.4 Nuclear magnetic resonance (NMR)

All one-dimensional NMR experiments were conducted on either Varian Inova 500 MHz or 600 MHz NMR spectrometers. The UBA wildtype and its D-Ala variant were prepared in D_2O with 50mM phosphate and 1mM deuterated DTT at pD 6.5, using dioxane as a chemical shift standard. pD refers to the uncorrected meter reading. The GA wildtype and its D-Ala variant were dissolved in 50mM phosphate D_2O buffer at pH 7.0 with 3,3,3-trimethylsilyl-proprionate (TSP) as a chemical shift standard. The PSBD samples were dissolved in 10% D_2O with 20mM Tris-d_{11} and 50mM sodium chloride at pH 8.0 with TSP as an internal standard. The EH samples were prepared in D_2O containing 50mM deuterium sodium acetate and 100mM sodium chloride at pD 5.7 with TSP. The peptide concentrations in all samples are around 1 mM. The temperature of the NMR probe was calibrated by methanol when below 35°C and ethylene glycol when above 35°C.

Total correlated spectroscopy (TOCSY) and nuclear overhauser effect spectroscopy (NOESY) of the UBA and PSBD D-Ala variants were recorded on Varian Inova 600MHz NMR spectrometer. The UBA sample was prepared in 10% D_2O/90% H_2O with the same buffer used for 1-D NMR. The mixing time was 50ms for TOCSY and 300ms for NOESY. For PSBD, the sample was dissolved in 10% D_2O/90% H_2O and pH was adjusted to 5.0. The mixing time was 80 ms for TOCSY and 300 ms for NOESY. The matrix size for both TOCSY and NOESY spectra are 4096 x 256. The spectral width of both dimensions is 8000 Hz. The peptide concentrations in all samples are around 1 mM.

2.2.5 NMR lineshape analysis

NMR lineshape analysis can be used to determine the exchange rate between the folded and unfolded state when the protein undergoes two-state folding and the exchange rate is within a certain range. When the difference of the NMR resonance frequencies between the folded and unfolded states is large compared to the exchange rate, two sharp peaks corresponding to each state will be observed, while when the difference in frequency is small compared to the exchange rate, only one sharp peak will be observed, located between the two resonances. However, when the difference of the NMR resonance frequencies of the folded and unfolded states is comparable to the exchange rate, one peak will be observed with significantly broadened line. The location of the peak is determined by the population and chemical shift of each state. In this case, the

exchange rate can be determined by fitting the broad peak to the following equations (*24*):

$$I(\nu) = -C_0 \frac{P[1+\tau(\frac{p_D}{T_{2N}}+\frac{p_N}{T_{2D}})]+QR}{P^2+R^2} \tag{2-4}$$

where $I(\nu)$ represents the intensity of the NMR peak; C_0 is a normalization constant, p_D and p_N are the population of the unfolded and folded states determined by equilibrium experiments; T_{2N} and T_{2D} are the linewidth of the folded and unfolded state resonances; P, Q, and R are given by:

$$P = \tau[(\frac{1}{T_{2N}T_{2D}}) - 4\pi^2\Delta\nu^2 + \pi^2(\delta\nu)^2] + \frac{p_N}{T_{2N}} + \frac{p_D}{T_{2D}} \tag{2-5}$$

$$Q = \tau[2\pi\Delta\nu - \pi\delta\nu(p_N - p_D)] \tag{2-6}$$

$$R = 2\pi\Delta\nu[1+\tau(\frac{1}{T_{2N}}+\frac{1}{T_{2D}})] + \pi\delta\nu\tau[\frac{1}{T_{2N}} - \frac{1}{T_{2D}} + \pi\delta\nu(p_N - p_D)] \tag{2-7}$$

where δν, Δν and τ are given by:

$$\delta\nu = \nu_N - \nu_D \tag{2-8}$$

$$\Delta\nu = \frac{\nu_N + \nu_D}{2} - \nu \tag{2-9}$$

$$\tau = \frac{1}{k_u + k_f} \tag{2-10}$$

ν_N and ν_D are the frequencies of the folded and unfolded state resonances, k_f and k_u are the folding and unfolding rates and the exchange rate is the sum of k_f and k_u.

In this work, the γ-protons from Val16 and Val21 in PSBD and the δ proton from Leu16 in EH were followed as a function of temperature. The folding and unfolding rates of these two proteins were determined by fitting the broadened resonances with the above equations.

2.3 Results

All of the proteins have a glycine located at a helix C-capping position. They are G331 (Residues are numbered according to the full-length protein. The first residue of the UBA domain is residue Q319 and the C-terminal residue is E363.) in UBA with φ = 122° (121° to 130°), ψ = 6.5° (-2° to 12°) from 18 NMR structures; G16 in GA with φ, ψ values of 108° and –22° from 1 NMR structure; G15 in PSBD with φ, ψ values of 94° (56° to 94°)

and 81° (1° to 81°) from 35 NMR structures; G39 in EH with φ, ψ values of 79° and 24° from the X-ray structure. The first number is the value for φ, ψ reported for the best representative structure and the numbers in brackets are the range of values observed in the different members of the NMR ensemble. An important consideration is whether or not introduction of a new methyl group leads to steric clashes. Modeling indicates that a D-Ala can be accommodated at each of these sites and shows that the new methyl group is exposed to solvent. All of the glycines are largely exposed to the solvent and the solvent exposure of the backbone and sidechain at these sites are listed in table 2-3. Thus the difference in solvation between the folded and unfolded states is likely to be modest, and unfolded state entropy effects should be the main factor in any change in stability. Model proteins with L and D-amino acids substitutions were built to check the change in the exposed surface area for the backbone and sidechain at the substitution site. L and D-amino acids replacements experience similar changes in exposed surface area.

2.3.1 D-Ala substitutions do not perturb the structures

Far-UV CD and NMR experiments were conducted in order to compare the structure of the D-Ala variants to the respective wildtypes. The CD spectra all show the classic double minimum at 222 and 208 nm, indicative of helical secondary structure. Based on the CD data, wildtype and respective mutants have similar secondary structure content (figure 2-2). NMR studies confirm the CD results. The one-dimensional ^{1}H spectra of the GA and EH domains contain characteristic sets of ring current shifted methyl resonances which provide a convenient probe of tertiary structure (*16, 23*). The same peaks are observed in the respective D-Ala variants (figure 2-3). TOCSY and NOESY spectra were recorded for the UBA and PSBD D-Ala mutants. $C_\alpha H$ chemical shifts are a sensitive indicator of secondary structure. The difference in $C_\alpha H$ chemical shifts is plotted in figure 2-4. For UBA, 90% of the residues have been assigned and most $C_\alpha H$ chemical shifts are very close to the wildtype values with differences within $\pm$0.1ppm. For PSBD, 80% of the residues have been assigned and most of the $C_\alpha H$ shift differences are within $\pm$0.1 ppm of wildtype except for some residues located at the N and C terminus, which have differences around 0.2 ppm. The complete lists of backbone assignments are given in table 2-4 and table 2-5.

2.3.2 Substitution with D-Ala increases protein stability

Thermal denaturation of the wildtype proteins and D-Ala variants and the D-Val variant of PSBD were monitored by far-UV CD at 222 nm. The curves were fit assuming a two-state model and are showed in figure 2-5. There is some controversy concerning the cooperativity of the folding of PSBD. Some have argued that its folding process is two-state while others have argued that it is a downhill folding (*25, 26*). The available evidence suggests that the other domains fold via a two state transition. The thermodynamic parameters derived from the unfolding curves are listed in table 2-6. The T_m of PSBD G15D-Ala, UBA G331D-Ala and EH G39D-Ala mutants are 9°C, 8°C and

5°C higher than the respective wildtype values. The T_m of the GA module is too high to be measured reliably. The free energy of unfolding of all of the domains and their D-Ala variants was determined by urea or guanidine denaturation. Guanidine denaturation was used for PSBD and GA because the urea transitions are too broad and the c_M values are too high to allow the curves to be fit accurately. The urea denaturations of EH wildtype and its D-Ala variant were done at 5°C because the pre-transition can be better defined at this temperature. The unfolding curves are shown in figure 2-6. The m values of all of the D-Ala mutants are close to the respective wildtype values, suggesting that there is not a significant change in the amount of surface area exposed upon unfolding. The free energy of unfolding of the D-Ala variants is 0.6 to 1.2 kcal/mol higher than the respective wildtype proteins. These values are close to some estimates for the free energy change resulting from changes in backbone conformational entropy caused by a Gly to Ala substitution. For example, Freire and coworkers have estimated that replacement of glycine by alanine will reduce the conformational entropy of the peptide backbone by 2.46 ± 0.2 cal/K·mol. This will result in a change in free energy of 0.7 kcal/mol at 298K (*4, 27*).

2.3.3 Analysis of the folding and unfolding rates of the PSBD and EH variants is consistent with denatured state effects

The folding and unfolding rates of PSBD can be estimated by NMR lineshape analysis. The resonances of the γ-protons of Val16 and Val21 were followed. V16 is near the start of the loop which connects the two helices and V21 is located at the end of 3_{10}-helix. At 25°C, their chemical shifts are 0.42 and 0.23 ppm respectively. As the temperature is increased, these two resonances shift downfield significantly and become broader because of rapid conformation exchange between the native and unfolded structures. These two resonances appear in a relatively uncrowded region of the spectrum and do not overlap with other peaks. Thus NMR lineshape analysis can be used to define the exchange rates. To obtain the folding and unfolding rates, p_n (population of the native state) and p_d (population of the unfolded state) were determined at each temperature by separate equilibrium unfolding experiments monitored by CD in D_2O. D_2O was used since the NMR studies were done in D_2O. Analysis of the Val16 and Val21 lineshapes gives similar exchange rates. Figure 2-7 shows a plot of $\ln k_f$ and $\ln k_u$ versus 1/T. Compared to the wildtype, at each temperature the folding rate of D-alanine variant increases significantly while the unfolding rate is almost unchanged. The simplest interpretation of these results is that the unfolded state free energy levels change while the transition state and native state levels are not significantly altered (figure 2-8).

Analysis of the folding and unfolding rates argues that denatured state effects are more important than folded state effects. Consider the schematic reaction coordinate diagram shown in figure 2-8,

$$\Delta G^{TS}_{f,mutant} - \Delta G^{TS}_{f,wildtype} = -RT\ln k_{f,m} + RT\ln k_{f,wt} = -RT\ln(k_{f,m}/k_{f,wt})$$
$$= (a+b) - (b+c) = a - c \qquad (2\text{-}11)$$

$$\Delta G^{TS}_{u,mutant} - \Delta G^{TS}_{u,wildtype} = -RTlnk_{u,m} + RTlnk_{u,wt} = -RTln(k_{u,m}/k_{u,wt})$$
$$= (a'+b') - (b'+c') = a'-c' \qquad (2\text{-}12)$$

where $k_{f,m}$ is the folding rate of the D-Ala variant; $k_{f,wt}$ is the folding rate of wildtype; $k_{u,m}$ and $k_{u,wt}$ represent the unfolding rates of the D-Ala variant and wildtype, respectively. a, b, c, a', b', c' are defined in figure 2-8. a-c is the change in the folding activation free energy while a'-c' represents the change in the activation free energy for unfolding. From the lineshape analysis result, we know that $k_{f,m}/k_{f,wt}$ is larger than $k_{u,m}/k_{u,wt}$, so (a-c) will be smaller than (a'-c'). Since a'=a, we can conclude c is larger than c', which means that the change in denatured state is more significant than native state. This is consistent with our assumption that the stabilization of D-alanine variant results from the destabilization of the unfolded state. The folding and unfolding rates of EH wildtype and D-alanine mutant have also been determined by NMR lineshape analysis. The δ-proton from Leu16 was followed at different temperatures. The plots of lnk_f and lnk_u versus 1/T are shown in Figure 2-7. Although the unfolding rate of the D-alanine variant does increase relative to wildtype, the change is not as large as the change in the folding rate. Again this is consistent with the assumption that the stabilization by D-Ala substitution mainly comes from the destabilization of the unfolded state. Unfortunately the folding and unfolding rates of UBA and GA could not be determined by lineshape analysis since there are no resonances that can be followed that shift and broaden significantly with increasing temperature.

2.3.4 The effects of the replacement of a C-capping Gly by a β-branched amino acid

Backbone desolvation can also affect protein stability (*8*). The C-capping glycines studied here are largely exposed, thus substitution of glycine by other amino acids leads to desolvation of the peptide backbone. However, if the glycine is located on the surface, the desolvation of the peptide backbone in the folded and unfolded states caused by the substitution with other amino acids should be similar. Thus incorporating D-Val, with β-branched sidechain, into the helical C-capping position should have similar effect to the protein stability as D-Ala. PSBD G15D-Val was synthesized and thermal and guanidine denaturations were monitored by far-UV circular dichroism (figure 2-9). The far UV CD spectrum indicates the significant helix secondary structure which is similar to that observed for wildtype and D-Ala variant, suggesting the substitution does not perturb the protein structure. The melting temperature of D-Val variant is 59.7^{o}C, which is 7^{o}C higher than the wildtype, but 2^{o}C lower than the D-Ala variant. The free energy of unfolding for the D-Val variant is 3.39 kcal/mol, which is about 0.7 kcal/mol higher than the wildtype but 0.6 kcal/mol lower than the D-Ala variant (table 2-6). The m-value defined from the denaturation curve is 0.68, which is close to both the wildtype and D-Ala, suggesting that the mutation does not significantly affect the change of solvent exposed area upon unfolding. However, the destabilization of the protein by incorporating D-Val into PSBD compared to D-Ala gives us a hint that the desolvation effect introduced into the folded and unfolded states is not canceled out in this system. If

purely entropic effects were involved we would expect the D-Val variant to be more stable than the D-Ala variant since replacement of Ala by Val was estimated by Freire et. al. to reduce the conformational entropy by around 1.92 cal/K·mol at 25^{o}C (*4, 27*). This difference from the experimental value result from the partial burial of the C-Capping site in the wildtype. The backbone exposure is 51 Å^2 for the PSBD site while it ranges from 59 to 72 Å^2 in the other proteins. In the D-Val variant, the desolvation effect on the peptide backbone is larger than the D-Ala variant, which results in the lower stability of the protein. Thus desolvation of the peptide backbone may play an unfavorable role in the protein stability, but the dominant factor is still the conformational entropy since the D-Val variant is still more stable than the wildtype even though there is an unfavorable desolvation effect. Steric clashes introduced by the large D-Val sidechain could contribute to the smaller $\Delta\Delta G^0$ for the Gly to D-Val mutant. Modeling of the D-Val assuming the same φ, ψ angles as observed for wildtype and a χ angle equals 180^{o} indicates that the side chain of D-Val is pointing to the outside of the folded protein, making no contacts with other residues. Thus the contribution from steric clash or extra interactions by the substitution should be less important.

2.4 Discussion

All of the small helical peptides we studied have a glycine located at the helix C-capping position with positive φ-angles. This conformation is strongly disfavored for L-amino acids but allowed for D-amino acids. In this chapter, we reported that the substitution of these glycines by D-Ala stabilizes the proteins by 0.6 - 1.2 kcal/mol. This value is close to some estimates for the free energy change resulting from changes in backbone conformational entropy caused by a Gly to Ala substitution (*4*). From the fact that the folding rate changes much more than the unfolding rate upon mutation, we deduce that unfolded state effects are more important and the decrease of the unfolded state entropy is likely the major factor contributing to the stabilization of the protein. Of course, other factors besides changes in conformational entropy can affect protein stability. These include differential desolvation of the peptide backbone in the folded and unfolded states, conformational strain in the native structure, new sidechain interactions introduced into the folded state and specific interactions in the unfolded states (*28-30*). The small difference in the α-proton chemical shifts indicates that the substitution does not perturb the native structure and modeling shows that the sidechain of the D-Ala does not make any new native contacts. Our previous studies which compared L-Ala and D-Ala variants showed that substitution by D-amino acids does not introduce backbone conformational strain into the native structure.

Substitution of glycine by other amino acids will affect backbone desolvation, which can also play a role in the protein stability. Since the C-capping glycines are largely exposed on surface, the extent of changes in the solvation of the sidechain in the folded and unfolded states should be similar. Thus any changes of sidechain desolvation upon mutation should have little effect on the free energy difference between these two states.

However, replacement of glycine by other amino acids will protect the backbone from the solvent and valine will have a bigger effect than alanine because of its larger sidechain. In the study of D-Val variant at the C-capping position in PSBD, we observed a net stabilization compared to wildtype, but the increase was not as large as for the D-Ala variant. This suggests that desolvation effects might play a role in PSBD assuming that no new interactions are induced by the bulky D-Val sidechain. The total stabilization effect suggests that the contribution from conformational entropy is larger than the desolvation effect and still plays a dominant role in the changing of the protein stability upon substitution. In some cases where the glycine is buried or partially buried, the desolvation effect can be the major factor which can result in a destabilization.

Previous studies showed a destabilization with the substitution of glycine by L-amino acids. The conformational energy penalty of incorporating a L-amino acid is the main source of the destabilization because it introduces strain into the protein at any sites with positive φ angles. Our substitution with D-amino acids successfully avoids this unfavorable effect, resulting in stabilization of the protein. In our study, the C-capping glycines are largely exposed, which makes the desolvation effect modest, conformational entropy being the major factor. Glycines with positive φ angles are frequently found at the helix C-capping position, thus substitution by D-amino acids at this site should be a general strategy to enhance protein stability.

Table 2-1 φ and ψ angles of the C-capping glycines. The dihedral angles were measured using the indicated PDB files. The structure of PSBD and EH are NMR structures and the numbers in brackets are the range of the angle for different members of the NMR ensemble. The GA structure is also a NMR structure but only a 'best representative structure' was deposited in the PDB. The EH structure was solved by X-ray crystallography.

	Φ (^{o})	Ψ (^{o})
PSBD G15	94 (56 to 94)	81 (1 to 81)
GA G16	108	-22
UBA G331	122 (121 to 130)	6.5 (-2 to 12)
EH G39	79	24

Table 2-2 Methods for protein purification

	Column	Buffer	Gradient	Retention time	Flow rate
UBA	Semi-prepartive C4	A: 100% H_2O, 0.1% TFA B: 90% isopropyl alcohol, 10% H_2O, 0.1% TFA	25%-65%B in 80 min	34 min	3ml/min
GA	Semi-prepartive C4	A: 100% H_2O, 0.1% TFA B: 90% isopropyl alcohol, 10% H_2O, 0.1% TFA	25%-65%B in 40 min	18 min	3ml/min
PSBD	Semi-prepartive C18	A: 100% H_2O, 0.1% TFA B: 90% acetonitrile, 10% H_2O, 0.1% TFA	25%-75%B in 100 min	30 min	5ml/min
EH	Semi-prepartive C18	A: 100% H_2O, 0.1% TFA B: 90% acetonitrile, 10% H_2O, 0.1% TFA	25%-65%B in 80 min	22 min	5ml/min

Table 2-3 Calculated solvent exposure of backbone and sidechain at the C-capping site for wildtype, L-Ala and D-Ala variants. The PSBD structure comes from the average of 35 NMR conformations; the UBA structure comes from 18 NMR conformations; the EH structure was solved by X-ray Crystallography and the GA structure was defined by one NMR conformation. The backbone exposure refers to the amide proton and nitrogen, α-carbon, carbonyl carbon and oxygen. For comparison the values for L-Ala, D-Ala and Gly are given for the central residue in a L-Ala-X-L-Ala tripeptide which adopts an extended conformation (φ, ψ=180^{o}). The program GETAREA was used to calculate exposed area.

	Backbone exposed area (Å^2)	Sidechain exposed area (Å^2)
PSBD G15	50.8	NA
PSBD G15L-Ala	25.1	56.8
PSBD G15D-Ala	30.8	38.2
UBA G331	70.8	NA
UBA G331L-Ala	45.1	49.8
UBA G331D-Ala	42.3	50.6
EH G39	72.2	NA
EH G39L-Ala	40.6	57.1
EH G39D-Ala	43.0	48.5
GA G16	58.6	NA
GA G16L-Ala	31.8	55.2
GA G16D-Ala	31.1	47.5
Ala-Gly*-Ala	80.9	NA
Ala-Ala*-Ala	39.6	72.3

*A tri-peptide was constructed by PyMol. The surface area of the middle residue was calculated.

Table 2-4 ^{1}H chemical shifts of the backbone of UBA G331D-Ala variant refer to TSP at pH 6.5, 25^{o}C.

residue	δ(NH) ppm	δ(C_{α}H) ppm	residue	δ(NH) ppm	δ(C_{α}H) ppm
E1			F24	8.15	4.04
E2	8.62	3.95	A25	8.28	4.24
K3	7.97	4.02	C26	7.37	4.49
E4	8.23	3.96	E27	7.63	3.72
A5	7.73	3.91	K28	8.00	1.59
I6	7.68	3.25	N29	6.92	4.76
E7	7.89	3.99	E30	9.02	3.85
R8			N31	8.30	4.52
L9	8.11	4.16	L32	7.89	4.16
K10	8.56	4.37	A33	8.61	4.36
A11	7.70	4.24	A34	8.38	3.83
L12	7.54	4.24	N35	7.59	4.37
D-Ala13	8.13	4.17	F36	8.38	4.12
F14			L37	8.38	3.51
E15			L38	8.14	4.24
E16	8.75	4.21	S39	7.63	4.36
S17	8.97	4.22	Q40	7.42	4.02
L18	7.05	4.25	N41	8.03	4.69
V19	8.02	3.51	F42	8.23	4.62
I20	8.59	3.55	D43	8.29	4.60
Q21	7.42	4.09	D44	8.19	4.55
A22	8.25	3.74	E45		
Y23	9.06	3.72			

Table 2-5 ^{1}H chemical shifts of the backbone of PSBD G15D-Ala variant refer to TSP at pH 5.0, 25^{o}C.

residue	δ(NH) ppm	δ(C_{α}H) ppm	residue	δ(NH) ppm	δ(C_{α}H) ppm
A3			T24		4.51
M4			G25	8.59	4.43
P5			K26	8.61	4.09
S6	8.35	4.28	N27	9.45	4.40
V7	7.60	3.72	G28	8.02	3.90
R8	8.02	4.03	R29	6.92	4.06
K9	7.95	4.06	V30	8.60	4.10
Y10	7.95	4.37	L31	9.53	4.68
A11	8.54	3.75	K32	8.13	3.75
R12	7.68	4.15	E33	9.62	4.18
E13	8.28	4.03	D34		4.60
K14	7.63	4.21	I35	7.61	3.72
D-Ala15			D36	8.00	4.52
V16	8.20	4.00	A37	8.24	4.28
D17	8.32	4.50	F38	8.04	4.21
I18	8.30	3.59	L39	7.90	4.12
R19	8.02	3.93	A40	7.76	4.31
L20	7.77	4.22	G41	8.10	3.96
V21	6.99	3.68	G42	7.76	3.96
Q22	8.76	4.31	A43	8.03	3.90
G23	8.76	3.63			

Table 2-6 Thermodynamic parameters for the unfolding of PSBD, EH, UBA and GA and their Gly to D-Ala and D-Val variants. The numbers after the ± symbol are the standard errors to the fit. Urea was used as the denaturant for PSBD, EH and UBA. Guanidine was used for the GA module. Urea and guanidine denaturations were performed at 25°C except for the EH domain which was studied at 5°C.

	ΔG^o (kcal/mol)	m (kcal/mol mol^{-1})	Tm (°C)	ΔH^o (kcal/mol)
PSBD	2.75 ± 0.07	0.67 ± 0.01	52.5 ± 0.1	29.6 ± 0.5
PSBD G15D-Ala	4.00 ± 0.34	0.73 ± 0.07	61.3 ± 0.2	31.9 ± 0.8
PSBD G15D-Val	3.39 ± 0.03	0.68 ± 0.01	59.7 ± 0.1	32.0 ± 0.5
EH	1.91± 0.03	0.61 ± 0.01	55.6 ± 0.2	32.5 ± 0.7
EH G39D-Ala	2.55 ± 0.13	0.66 ± 0.03	60.7 ± 0.4	33.1 ± 1.4
UBA	1.50 ± 0.04	0.51 ± 0.01	65.2 ± 0.5	20.6 ± 0.3
UBA G331D-Ala	2.27 ± 0.07	0.56 ± 0.02	73.7 ± 1.1	22.4 ± 0.4
GA	4.71 ± 0.16	1.00 ± 0.03		
GA G16D-Ala	5.52 ± 0.19	1.02 ± 0.04		

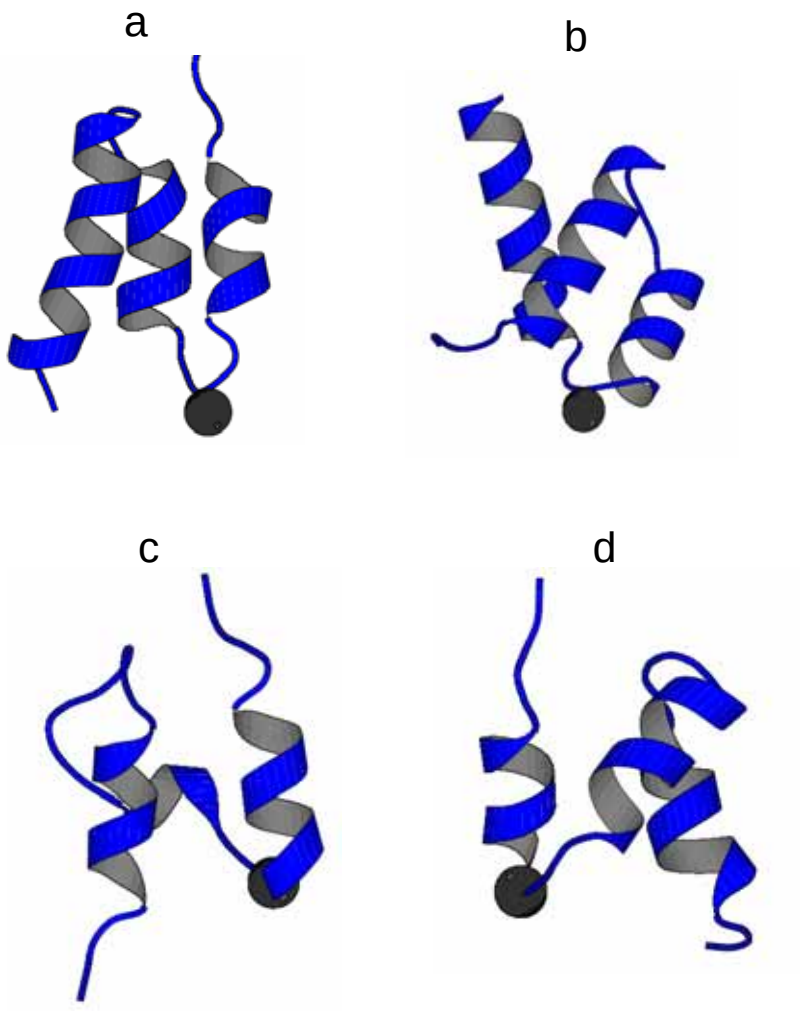


Figure 2-1 Ribbon diagram of (a) GA, (b) EH, (c) PSBD and (d) UBA. The Glycine at the helix C-capping position is shown as a sphere. The figure was generated using the program molscript and the PDB files 2PDD, 1ENH, 1PRB and 1UBA respectively.

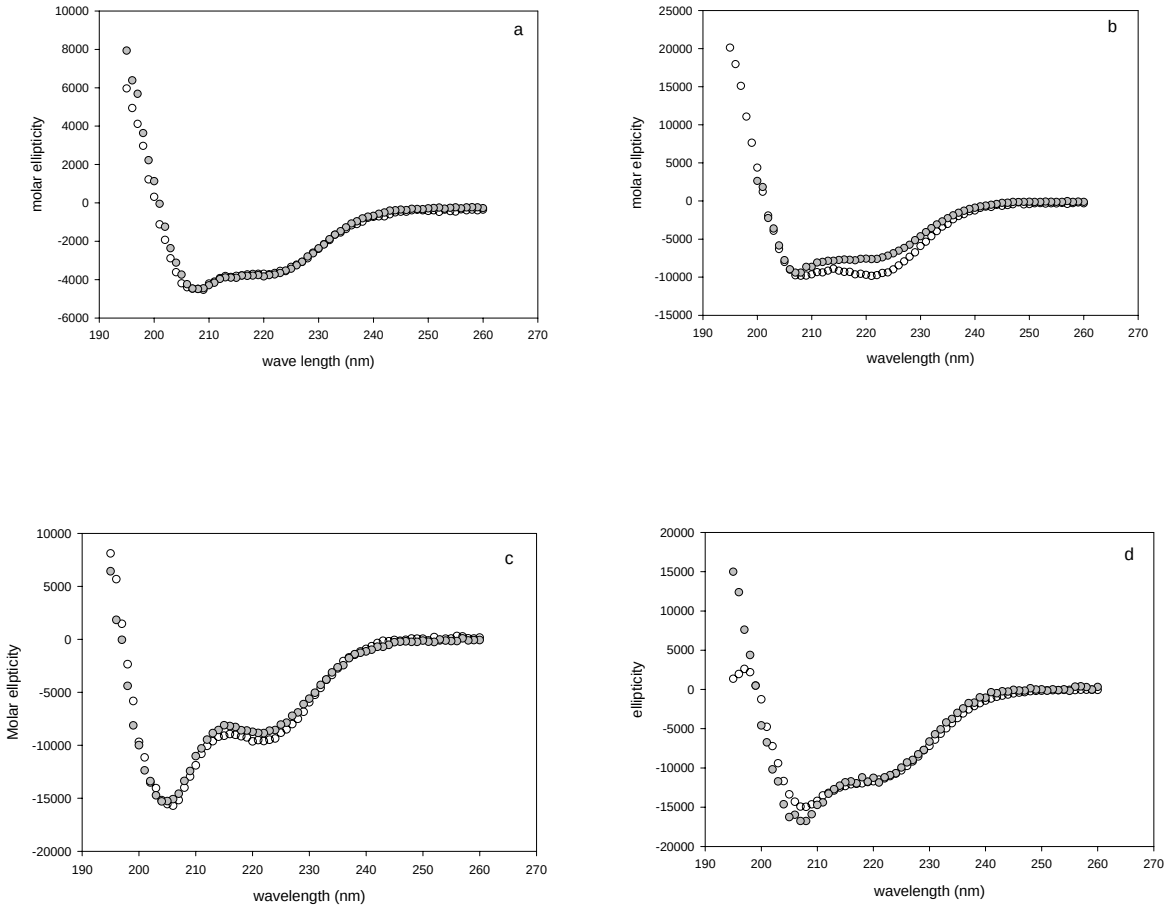


Figure 2-2 Far-UV CD wave scan of (a) UBA; (b) GA; (c) PSBD and (d) EH from 260 to 195 nm. The open circles represent the wildtype and the gray circles represent respective D-Ala variants. The relative low intensity of the spectra in figure (a) could be due to uncertainties in concentration determination.

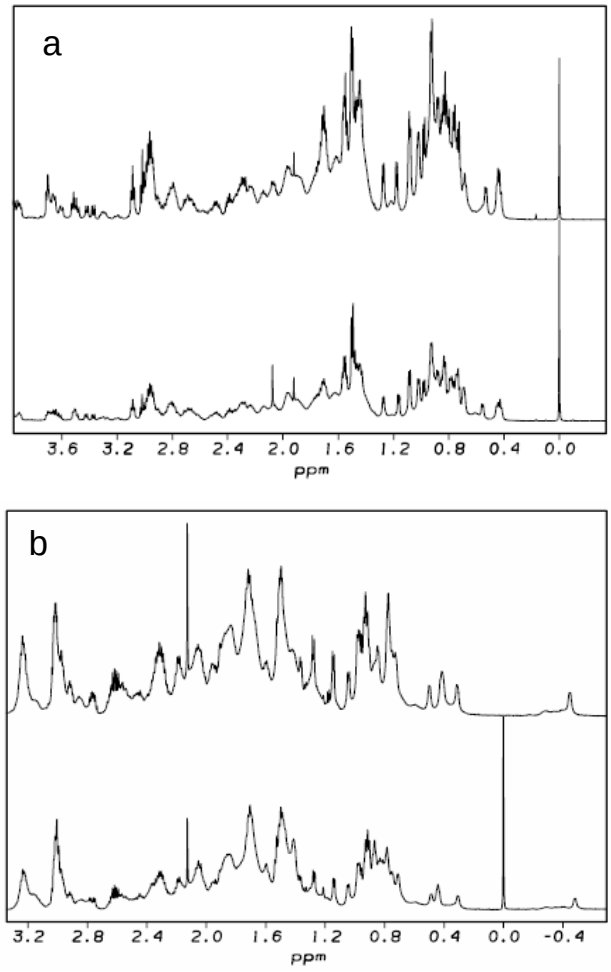


Figure 2-3 Downfield region of one-dimensional ^{1}H NMR spectrum of (a) GA wildtype and D-Ala mutant and (b) EH wildtype and D-Ala mutant. The upper spectrum is from wildtype while the bottom spectrum is from the D-Ala variant. The sharp peak at 0 ppm is the internal standard. The spectra of UBA were collected in 99% D_2O with 50mM phosphate and 1mM deuterated DTT at pD 6.5, 25^{o}C. The spectra of GA were collected in 99% D_2O with 50mM phosphate at pD 7.0, 25^{o}C. The spectra of PSBD were collected in 10% D_2O with 20mM Tris and 50mM sodium chloride at pH 8.0, 25^{o}C. The spectra of EH were collected in 99% D_2O with 50mM deuterated sodium acetate and 100mM sodium chloride at pD 5.7, 25^{o}C.

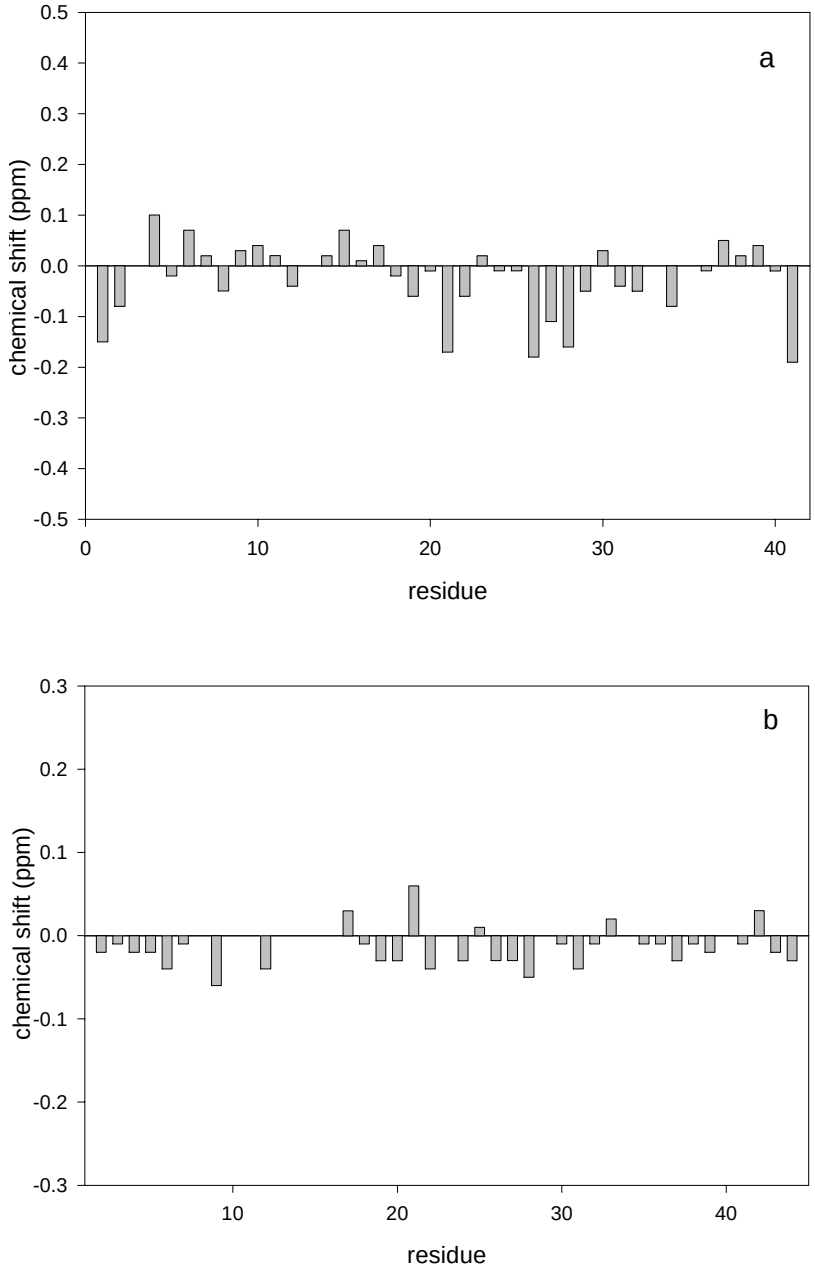


Figure 2-4 Plot of α-proton chemical shift differences between wildtype and D-Ala variant for (a) PSBD and (b) UBA. The spectra of PSBD were collected in 10% D_2O with 20mM Tris and 50mM sodium chloride at pH 5.0, 25^{o}C. The spectra of UBA were collected in 10% D_2O with 50mM phosphate and 1mM DTT at pH 6.5, 25^{o}C.

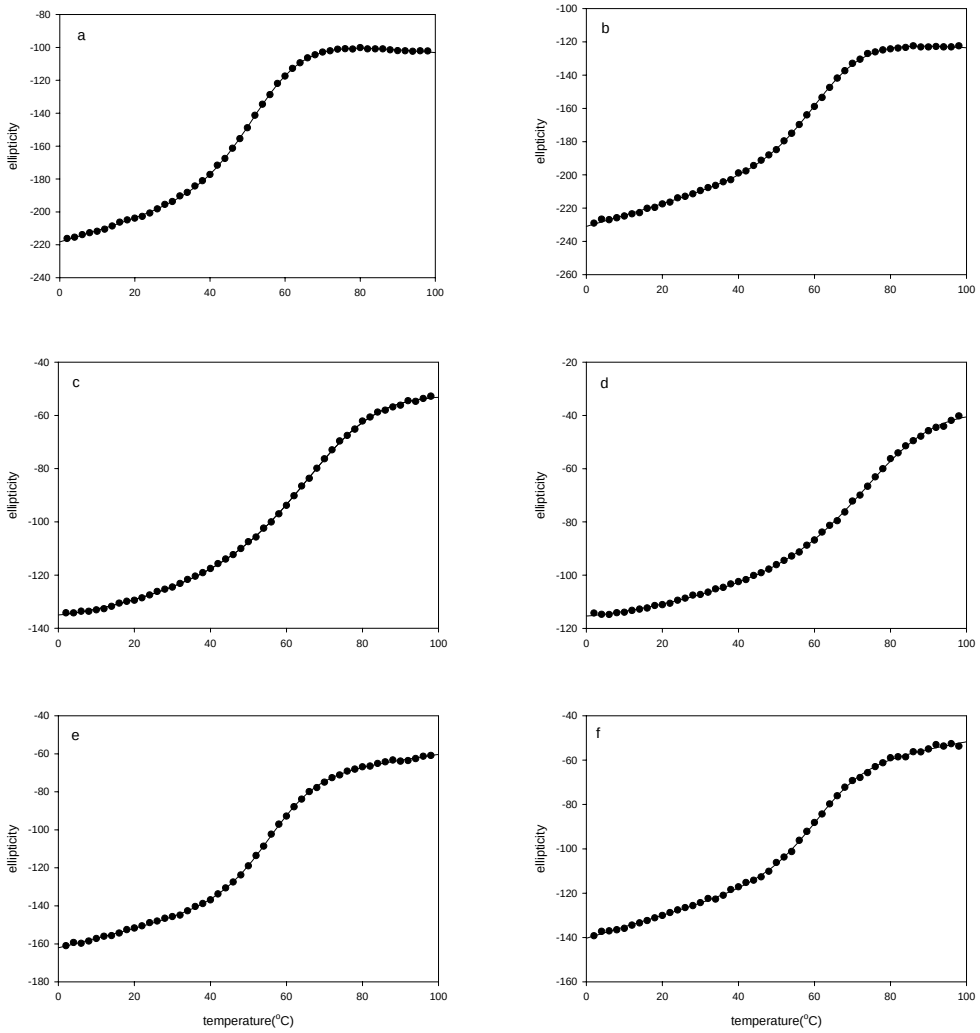
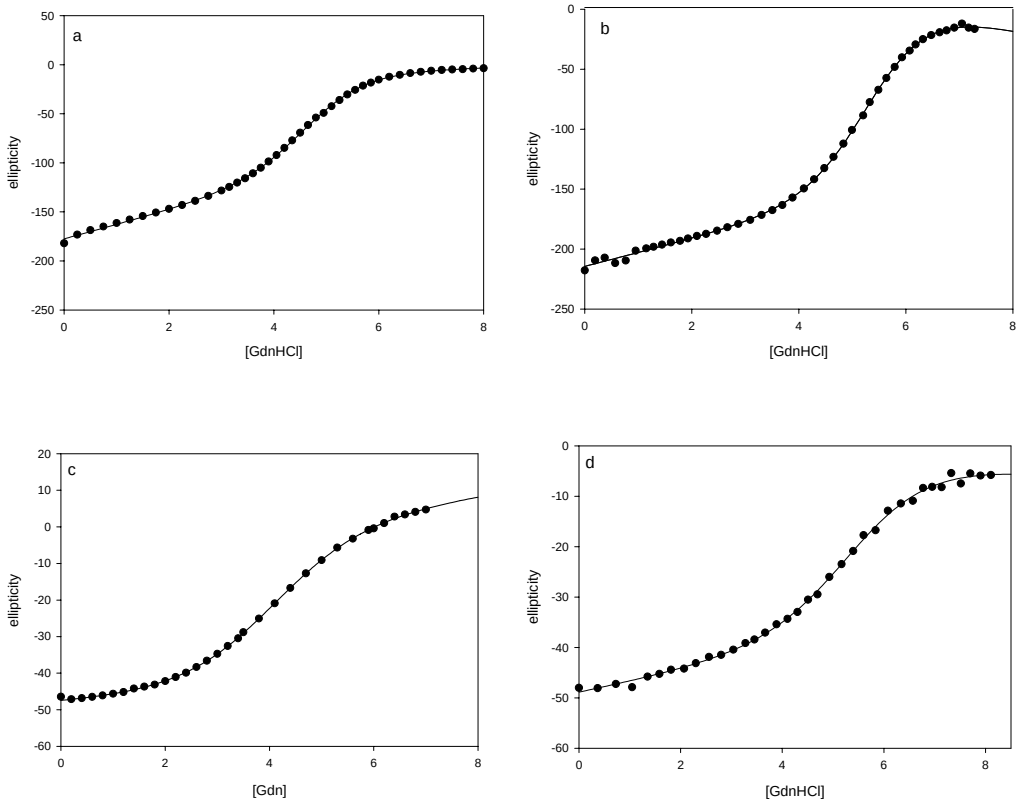
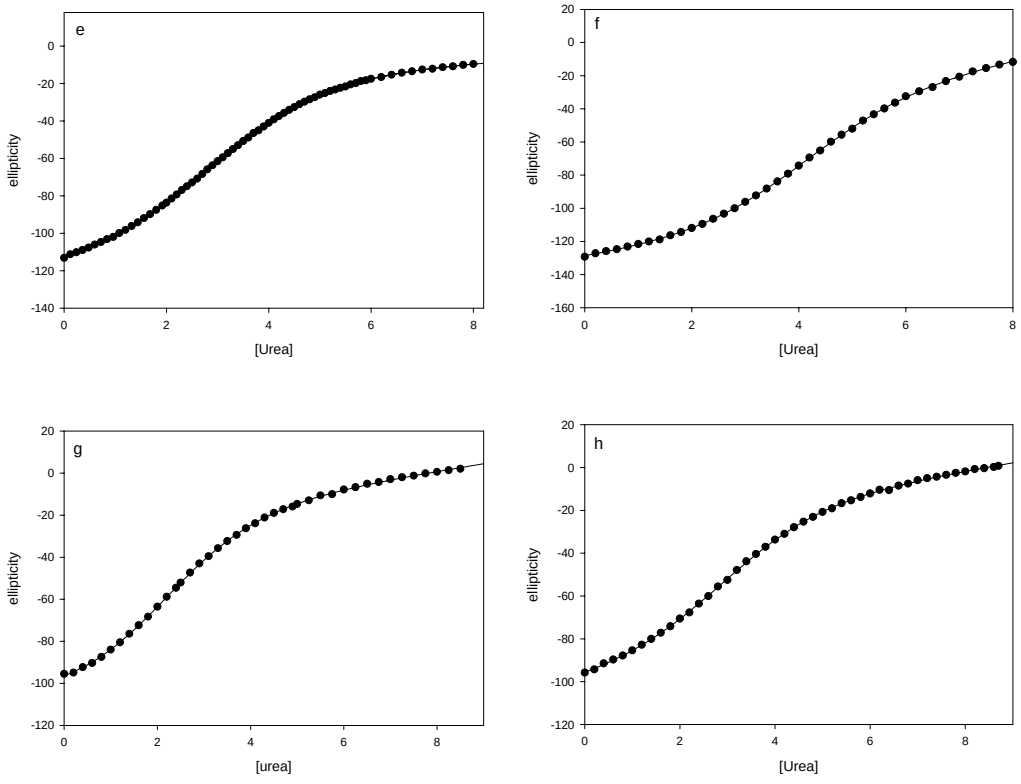


Figure 2-5 Thermal unfolding curves of (a) PSBD wildtype, (b) PSBD D-Ala mutant, (c) UBA wildtype, (d) UBA D-Ala mutant, (e) EH wildtype, (f) EH D-Ala mutant. A 50mM phosphate buffer with 1mM DTT at pH 6.5 was used for UBA wildtype and variants. For the GA wildtype variants, a 50mM phosphate buffer at pH 7.0 was used. For PSBD wildtype and variants, a buffer containing 2mM sodium phosphate, 2mM sodium borate and 50mM sodium chloride at pH 8.0 was used. For EH wildtype and variants, a 50mM sodium acetate buffer with pH 5.7 was used.

Figure 2-6 Urea or guanidine unfolding curves for wildtype and D-Ala variants of (a) GA wildtype, (b) GA D-Ala mutant, (c) PSBD wildtype, (d) PSBD D-Ala mutant, (e) UBA wildtype, (f) UBA D-Ala mutant, (g) EH wildtype, (h) EH D-Ala mutant monitored by CD at 222nm. Experiments with the GA, PSBD and UBA domains were performed at 25^{o}C. The denaturation of EH was conducted at 5^{o}C. A 50mM phosphate buffer with 1mM DTT at pH 6.5 was used for UBA wildtype and variants. For the GA wildtype variants, a 50mM phosphate buffer at pH 7.0 was used. For PSBD wildtype and variants, a buffer containing 2mM sodium phosphate, 2mM sodium borate and 50mM sodium chloride at pH 8.0 was used. For EH wildtype and variants, a 50mM sodium acetate buffer with pH 5.7 was used.




e
f
g
h
ellipticity
[Urea]
[urea]

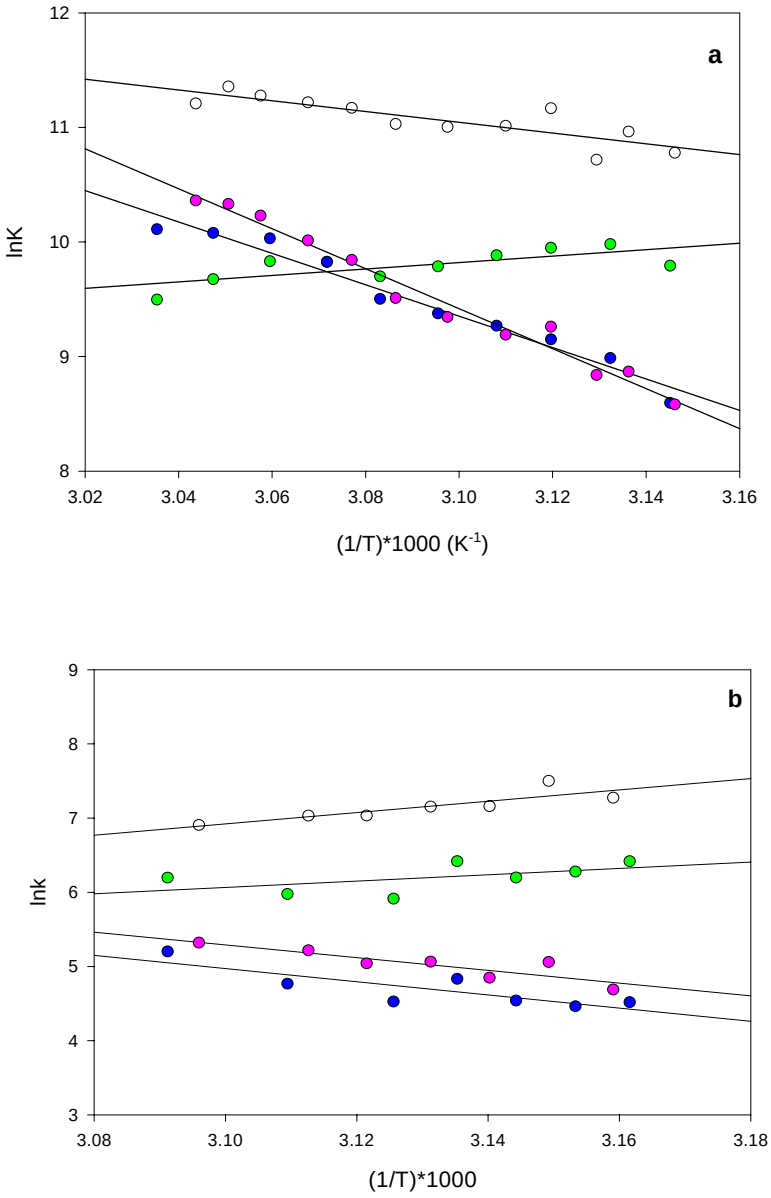


Figure 2-7 Plots of lnk versus (1/T)*1000 for the wildtype and D-Ala variant of (a) PSBD and (b) EH. Open circles represent the folding rate of the D-Ala variant; red circles represent the unfolding rate of the D-Ala variant; green circles represent the folding rate of wildtype; blue circles represent the unfolding rate of the wildtype. The change in the slope of the plots for wildtype in panel A and panel B are different. This may be real or it could reflect the difficulty in fitting the data. Note a plot of lnk_f vs 1/T is expected to have downward curvature but the curvature should be small for a small protein.

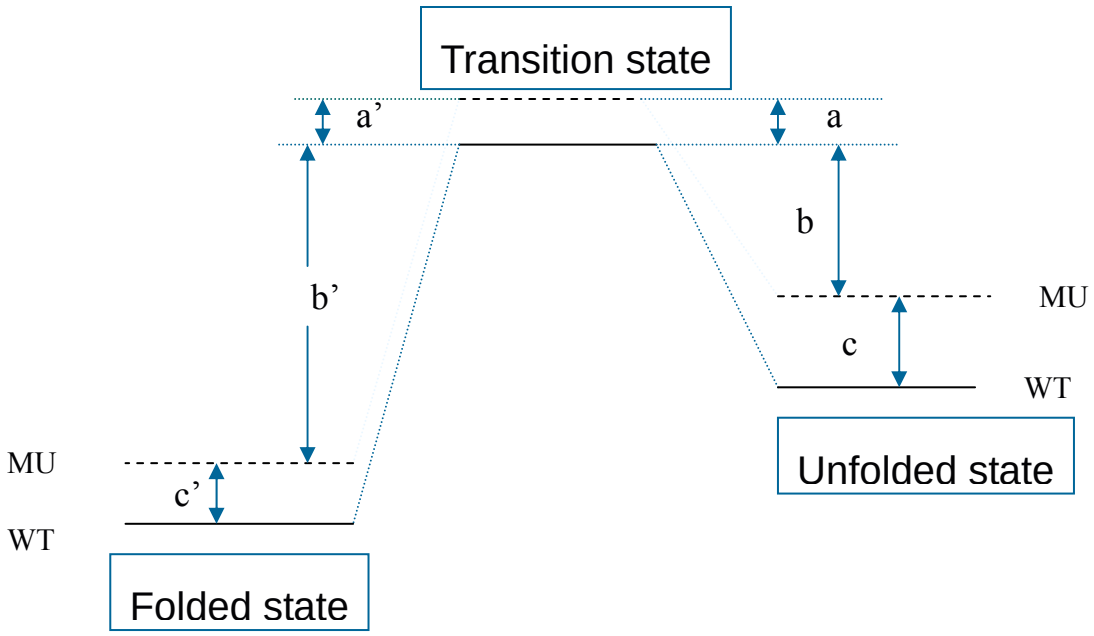


Figure 2-8 Diagram of the free energy levels of the folded state, transition state and unfolded state of wildtype and a D-Ala variant. The solid line represents the energy levels of wildtype; the dashed line represents the energy levels of the D-Ala variant.

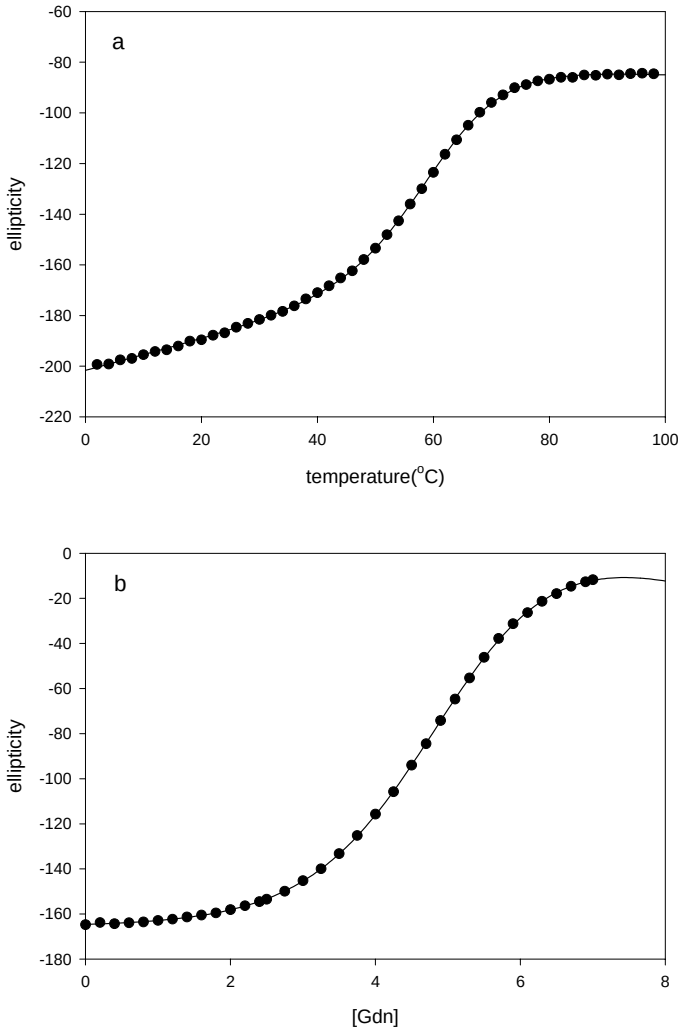


Figure 2-9 (a) Thermal unfolding curve of PSBD G15D-Val variant monitored by CD at 222nm; (b) Guanidine unfolding curve of PSBD G15D-Val variant performed at 25°C. All the experiments were done at 2mM sodium borate and 50mM sodium chloride at pH 8.0.

2.5 References

1. Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987) Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding, *Proc. Natl. Acad. Sci. U. S. A. 84*, 6663-6667.
2. Anil, B., Song, B., Tang, Y., and Raleigh, D. P. (2004) Exploiting the right side of the Ramachandran plot: substitution of glycines by D-alanine can significantly increase protein stability, *J. Am. Chem. Soc. 126*, 13194-13195.
3. Ganter, C., and Pluckthun, A. (1990) Glycine to alanine substitutions in helices of glyceraldehyde-3-phosphate dehydrogenase: effects on stability, *Biochemistry 29*, 9395-9402.
4. Stites, W. E., and Pranata, J. (1995) Empirical evaluation of the influence of side chains on the conformational entropy of the polypeptide backbone, *Proteins 22*, 132-140.
5. Hovmoller, S., Zhou, T., and Ohlson, T. (2002) Conformations of amino acids in proteins, *Acta. Crystallogr. D. Biol. Crystallogr. 58*, 768-776.
6. Nicholson, H., Soderlind, E., Tronrud, D. E., and Matthews, B. W. (1989) Contributions of left-handed helical residues to the structure and stability of bacteriophage T4 lysozyme, *J. Mol. Biol. 210*, 181-193.
7. Karplus, P. A. (1996) Experimentally observed conformation-dependent geometry and hidden strain in proteins, *Protein Sci. 5*, 1406-1420.
8. Ermolenko, D. N., Thomas, S. T., Aurora, R., Gronenborn, A. M., and Makhatadze, G. I. (2002) Hydrophobic interactions at the Ccap position of the C-capping motif of alpha-helices, *J. Mol. Biol. 322*, 123-135.
9. Schnolzer, M., Alewood, P., Jones, A., Alewood, D., and Kent, S. B. (1992) In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences, *Int. J. Pept. Protein Res. 40*, 180-193.
10. Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) Synthesis of proteins by native chemical ligation, *Science 266*, 776-779.
11. Muir, T. W. (2003) Semisynthesis of proteins by expressed protein ligation, *Annu. Rev. Biochem. 72*, 249-289.
12. Aurora, R., Srinivasan, R., and Rose, G. D. (1994) Rules for alpha-helix termination by glycine, *Science 264*, 1126-1130.
13. Richardson, J. S., and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of alpha helices, *Science 240*, 1648-1652.
14. Kalia, Y. N., Brocklehurst, S. M., Hipps, D. S., Appella, E., Sakaguchi, K., and Perham, R. N. (1993) The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of Bacillus stearothermophilus, *J. Mol. Biol. 230*, 323-341.
15. Mueller, T. D., and Feigon, J. (2002) Solution structures of UBA domains reveal a

conserved hydrophobic surface for protein-protein interactions, *J. Mol. Biol. 319*, 1243-1255.

16. Johansson, M. U., de Chateau, M., Wikstrom, M., Forsen, S., Drakenberg, T., and Bjorck, L. (1997) Solution structure of the albumin-binding GA module: a versatile bacterial protein domain, *J. Mol. Biol. 266*, 859-865.
17. Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: A framework for understanding homeodomain-DNA interactions, *Cell 63*, 579-590.
18. Spector, S., and Raleigh, D. P. (1999) Submillisecond folding of the peripheral subunit-binding domain, *J. Mol. Biol. 293*, 763-768.
19. Spector, S., Young, P., and Raleigh, D. P. (1999) Nativelike structure and stability in a truncation mutant of a protein minidomain: the peripheral subunit-binding domain, *Biochemistry 38*, 4128-4136.
20. Spector, S., Rosconi, M., and Raleigh, D. P. (1999) Conformational analysis of peptide fragments derived from the peripheral subunit-binding domain from the pyruvate dehydrogenase multienzyme complex of Bacillus stearothermophilus: evidence for nonrandom structure in the unfolded state, *Biopolymers 49*, 29-40.
21. Spector, S., Kuhlman, B., Fairman, R., Wong, E., Boice, J. A., and Raleigh, D. P. (1998) Cooperative folding of a protein mini domain: The peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex, *J. Mol. Biol. 276*, 479-489.
22. Religa, T. L., Johnson, C. M., Vu, D. M., Brewer, S. H., Dyer, R. B., and Fersht, A. R. (2007) The helix-turn-helix motif as an ultrafast independently folding domain: the pathway of folding of Engrailed homeodomain, *Proc. Natl. Acad. Sci. U. S. A. 104*, 9272-9277.
23. Mayor, U., Grossmann, J. G., Foster, N. W., Freund, S. M., and Fersht, A. R. (2003) The denatured state of Engrailed Homeodomain under denaturing and native conditions, *J. Mol. Biol. 333*, 977-991.
24. Burton, R. E., Busby, R. S., and Oas, T. G. (1998) ALASKA: A Mathematica package for two-state kinetic analysis of protein, *J. Biomol. NMR 11*, 355-359.
25. Garcia-Mira, M. M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M., and Munoz, V. (2002) Experimental identification of downhill protein folding, *Science 298*, 2191-2195.
26. Ferguson, N., Sharpe, T. D., Schartau, P. J., Sato, S., Allen, M. D., Johnson, C. M., Rutherford, T. J., and Fersht, A. R. (2005) Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family, *J. Mol. Biol. 353*, 427-446.
27. D'Aquino, J. A., Gomez, J., Hilser, V. J., Lee, K. H., Amzel, L. M., and Freire, E. (1996) The magnitude of the backbone conformational entropy change in protein folding, *Proteins 25*, 143-156.
28. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pKa values and

the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions, *Biochemistry 38*, 4896-4903.

29. Luisi, D. L., Wu, W. J., and Raleigh, D. P. (1999) Conformational analysis of a set of peptides corresponding to the entire primary sequence of the N-terminal domain of the ribosomal protein L9: Evidence for stable native-like secondary structure in the unfolded state, *J. Mol. Biol. 287*, 395-407.
30. Cho, J. H., Sato, S., and Raleigh, D. P. (2004) Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state, *J. Mol. Biol. 338*, 827-837.

3. The Important Role of A Buried Salt Bridge In The Villin Headpiece Domain

Abstract

The villin headpiece, HP67, is a small globular protein that consists of two subdomains; the N-terminal subdomain and the C-terminal subdomain, which form an extended hydrophobic core. The C-terminal subdomain, which is denoted HP35, can fold independently in isolation while the N-terminal subdomain can not. A salt bridge between E39 and K70 links the two subdomains and is buried in the hydrophobic core. The contribution of this salt bridge to the structure and stability of HP67 was studied by characterizing a series of variants (HP67E39Q, HP67E39L, HP67K70M and HP67E39QK70M). One-dimensional and ^{1}H-^{15}N HSQC NMR were used to characterize the structure of the variants and the N-terminal subdomain was found to be partially unfolded in all cases. Previous studies have shown that a H41Y mutation can stabilize the protein thus a HP67E39QH41Y double mutant was designed. The results show that HP67E39QH41Y is more stable than HP67E39Q with more native structure, although the N-terminal subdomian remains partially unfolded. Thus the buried salt bridge is indispensable to ensure a correct fold for the N-terminal subdomain. Another variant, HP67K38M, was analyzed. The mutation also results in an unfolded N-terminal subdomain even though K38 does not directly participate in the salt bridge. The K38M mutant likely exerts its effects by perturbing the conformation of E39.

3.1 Introduction

The villin headpiece domain is an F-actin-binding protein which consists of the C-terminal 76 residues of villin (*1-4*). The first nine residues can be removed without affecting its structure or function (*4, 5*). This construct is denoted as HP67 and is often used as a model system for protein folding studies (figure 3.1, 3.2) (*4-7*). HP67 consists of two subdomains, the N- and C-terminal subdomains. The two subdomains pack against each other tightly, forming an extended hydrophobic core comprising residues from both subdomains. The N-terminal subdomain has very little secondary structure while the C-terminal subdomain is rich in α-helix containing most of the residues responsible for F-actin binding (*4*). The isolated C-terminal domain is denoted as HP35 and it can fold independently to a compact globular protein (*5, 6*). Many experimental and computational studies of protein folding have been performed using HP35 as a model protein since it is one of the smallest proteins which can fold independently and it folds extremely rapidly (*8-14*). In contrast, the isolated N-terminal subdomain does not fold in

the absence of the C-terminal subdomain, which suggests that subdomain interface interactions play an important role in HP67 folding. Consistent with this hypothesis, previous studies found that the N-terminal subdomain stabilizes the C-terminal subdomain. HP35 can fold to a tertiary structure but it loses the ability to tightly bind actin without the N-terminal subdomain (*4, 7*).

HP67 undergoes multistate unfolding, which is unusual for such a small globular protein (*7*). The N-terminal subdomain is less stable than the C-terminal subdomain, which is found to unfold before the C-terminal subdomain (*7*). However, the unfolding of the two subdomains is not independent since the N-terminal subdomain can stabilize the C-terminal subdomain. An intermediate is suggested by both experimental and computational data in which the N-terminal subdomain is unfolded or partially unfolded but the C-terminal subdomain is natively folded (*7, 15*). This makes HP67 an interesting system for the study of cooperativity and protein folding. Both X-ray and NMR structures of HP67 have been reported (*4, 16*). Recently computational studies have suggested that the structure resolved by NMR is closer to an intermediate structure (*15*). Whether this is related to its biological function of HP67 is still unknown.

A histidine is often found at position 41 in the headpiece sequence and the only naturally occurring substitution is a tyrosine. His 41 acts as a pH sensitive folding switch in vitro (*4*). When pH is lower than its pKa, the histidine will be protonated and the introduction of a positive charge to the imidazole ring destabilizes the hydrophobic core and results in the unfolding of the N-terminal subdomain. However, the unfolding of the N-terminal subdomain does not require protonation of His-41. NMR relaxation dispersion studies have shown that the N-terminal subdomain can transiently and partially unfolded (*17*). Protonation of His-41 is easier in this partially unfolded state since the pKa of His-41 is elevated relative to the folded state. Continuum electrostatic calculations suggest that the pKa in the folded state is 1.1 lower (*17*). The His-41 to Tyr substitution has been shown to stabilize the protein and the effect is propagated throughout the structure even though the sidechain of residue 41 makes no contact with C-terminal subdomain (*7*). This result suggested that His-41 could contribute to the segmental unfolding process. However, HP67H41Y still undergoes multistate unfolding pathway, showing that His-41 is not the sole reason for the deviation from a two-state mechanism.

A buried salt bridge between E39 and K70 was suggested. Based on the NMR structure both residues are located in the hydrophobic core (*4, 18, 19*). In the NMR structure (PDB ID: 1QQV), the distance between the carboxyl oxygen atoms of E39 and the sidechain amino nitrogen atom of K70 is about 4.5 to 5.4 Å (*4*). In the crystal structure (PDB ID: 1YU5), the minimal distance between the carboxyl oxygen atoms of E39 and the sidechain nitrogen atom of K70 is only 2.8 Å (*16*). This salt bridge could stabilize the two subdomains and contribute to the cooperativity of the structure. Thus we designed a series of variants to break the salt bridge in order to examine the effect on the protein. We prepared HP67E39Q, HP67E39L, HP67K70M and HP67E39QK70M

variants since Q and M are conservative mutants for E and K respectively. They should allow disruption of the salt bridge without introducing any conformational strain.

3.2 Materials and Methods

3.2.1 Protein expression and purification

HP67 wildtype and variants were prepared by overexpression in Escherichia coli strain BL21(DE3). Cells were grown in Luria broth medium containing 100mg/L kanamycin at 37^{o}C. When the OD at 600 nm reached 0.6-0.8, 120mg/L IPTG was added to the LB medium for induction for 4 hours. Cells were harvested by centrifugation at 5000 rpm for 10 minutes then stored in a -20^{o}C freezer. Cell pellets were resuspended in 50 mM sodium phosphate buffer at pH 7.0 and lysed by sonication. Cell debris was removed by centrifugation at 33,000 rpm for 1 hour and the supernatant was loaded onto an S-Sepharose column and eluted with a gradient of 0 to 2M NaCl. Proteins were further purified by HPLC with a C8 preparative column. An A-B gradient system was used where buffer A consists of 99.9% (v/v) H_2O and 0.1% TFA, buffer B consists of 90% (v/v) acetonitrile, 9.9% (v/v) H_2O and 0.1% TFA. A gradient program was used in which the percent of buffer B increased from 10% to 90% in 80 minutes. The protein eluted at around 43% B. The retention time for the mutants is the same as the wildtype. ^{15}N-labeled HP67 variants were expressed and purified by the same method except that the cells were grown in M9 medium with $^{15}NH_4Cl$ as the nitrogen source instead of LB medium. The purified proteins were identified by MALDI-TOF mass spectrometry. The expected and observed molecular weights are as follows: HP67K70M, expected 7603.7, observed 7604.5; HP67E39Q, expected 7599.6, observed 7600.2; HP67E39QK70M, expected 7602.7, observed 7604.3; HP67E39QH41Y, expected 7625.7, observed 7626.2. Typical yields of purified protein from normal media were around 60 mg/L. The yields from M9 media were around 20 mg/L.

3.2.2 Circular dichroism (CD) spectroscopy

All CD experiments were performed using Aviv model 62A DS and 202SF circular dichroism spectrophotometers. All far-UV wavelength scans were performed in a 1mm quartz cuvette and data was collected from 195 nm to 260 nm with 5 repeats and an averaging time of 1 second at each wavelength. All of the thermal denatuaration and urea/guanidine denaturation experiments were monitored at 222 nm in a 1cm quartz cuvette. The protein concentrations were around 20 μM. Samples were dissolved in 10mM sodium phosphate buffer at pH 6.5 except for the guanidine denaturation experiments which were performed at pH 7.0. 1.5 minutes was allowed after each change in temperature or change in guanidine concentration. The temperature range for thermal denaturation was from 2^{o}C to 98^{o}C. All urea and guanidine denaturation curves were fit to the following equation:

$$y([den]) = \frac{a_n + b_n[den] + (a_d + b_d[den])e^{-\Delta G^0_{N-D}([den])/RT}}{1 + e^{-\Delta G^0_{N-D}([den])/RT}} \quad (3\text{-}1)$$

where,

$\Delta G^0([den]) = \Delta G^0(H_2O)$- m[den] (3-2)

y is the measured ellipticity. a_n, b_n, a_d, b_d are the parameters that define the signals of the native state (N) and denatured state (D). ΔG^0([den]) is the free energy change upon unfolding with denaturant and $\Delta G^0(H_2O)$ is the free energy change in the absence of denaturant. m is the so called m-value which is thought to be related to the change in solvent exposed surface area between the folded and unfolded states (*20*). Thermal denaturation curves were fitted to similar equations but with ΔG^0([den]) changed to ΔG^0(T). The Gibbs-Helmholtz equation gives the temperature dependence of the Gibbs free energy of unfolding:

$$\Delta G^0_{N-D}(T) = \Delta H^0(T_m) - T\Delta S^0(T_m) + \Delta C^0_p(T - T_m - T\ln\frac{T}{T_m}) \quad (3\text{-}3)$$

where T_m is the midpoint for thermal unfolding. $\Delta H^0(T_m)$ is the enthalpy change for unfolding. $\Delta C_p{}^0$ is the change in heat capacity, which is generally considered to be proportional to the difference of the non-polar surface area between the native and denaturated states. $\Delta C_p{}^0$ is assumed to be constant over the temperature range of interest. T_m can be determined by numerical differentiation of the denaturation curve.

3.2.3 Nuclear magnetic resonance (NMR) spectroscopy

All NMR experiments were conducted on either Varian Inova 500 MHz or 600 MHz NMR spectrometers. All samples for one-dimensional experiments were prepared in 99.9% D_2O containing 10 mM sodium phosphate at pD 6.5. pD refers to the uncorrected meter reading. Protein concentrations were around 500 μM. Each spectrum was referenced to TSP. All samples for ^1H-^{15}N HSQC experiments were prepared in 10% D_2O/90% H_2O with 10mM sodium phosphate at pH 7.0. The spectral width was 8000 Hz in the ^1H dimension and 2200 Hz in the ^{15}N dimension. 128 t_1 increments were collected and the data size in t_2 was 2048. Assignments of the ^1H-^{15}N HSQC spectra of HP67 variants were obtained using the published assignments (*4*). The protein concentrations were around 1 mM and all spectra were taken at 25°C except for HP67E39QK70M, for which a spectrum at 5°C was also taken.

3.3 Results

3.3.1 Disruption of the N-terminal subdomain upon disruption of the E39-K70 salt bridge

The far-UV CD spectra shows significant helix secondary structure for all three variants (HP67E39Q, HP67K70M and HP67E39QK70M) similar to that observed for the wildtype. However, far-UV CD wavelength spectra can not give information about individual residues, nor do they provide information about tertiary structure. The CD

spectrum of HP67 is expected to be dominated by the C-terminal subdomain thus the wavelength scans do not provide information about the N-terminal subdomain. Thus one dimensional NMR spectra were recorded on each variant to obtain more detailed structural information (figure 3.3). One of the methyl resonances from V33 is found at 0.22 ppm in the wildtype protein and this provides a useful marker of the state of the N-terminal subdomain. This peak disappeared in the spectra of the variants which indicates the N-terminal subdomain may lose its tertiary structure. We also observed changes in some of the other resonances (unassigned) from the ring current shifted methyl groups in upfield region which gives further proof that the structure of N-terminal subdomain is altered in the variants. However, the resonance from the methyl group of V50 remains at around -0.3 ppm in the spectra of the variants showing the C-terminal subdomian retains structure upon substitution (*4*).

^{1}H-^{15}N HSQC spectra were recorded of each variant. The resonances were assigned by comparision to the published assignments of HP67 wildtype (figure 3.4-3.6) (*4*). Most resonances from the C-terminal subdomain have chemical shifts which are similar to the wildtype values while the resonances from the N-terminal subdomain experience a significant change. The resonances from the N-terminal subdomain which are unassigned fall in a restricted region of the spectrum typically of unfolded protein. This suggests that the N-terminal subdomain is partially unfolded since a dispersed HSQC spectrum is expected for a folded structure while a poorly dispersed spectrum corresponds to an unfolded conformation. The spectra were recorded at 25^{o}C. Since some proteins can refold to the native conformation at lower temperature, we recorded the HSQC spectrum of HP67E39QK70M at 5^{o}C (figure 3.7). However, the spectrum recorded at 5^{o}C shows no significant difference from the one recorded at 25^{o}C. We also tested the effects of adding 1M trimethylamine-*N*-oxide (TMAO) to the HP67E39QK70M sample. TMAO stabilizes proteins but the N-terminal subdomain remains unfolded as judged by the HSQC spectrum (figure 3.8) (*21, 22*). All of these results show that the salt bridge between E39 and K70 is crucial for stabilizing the native fold of the N-terminal subdomain. Although most resonances from C-terminal subdomain remain at chemical shifts that are similar o wildtype, some residues display changes indicating that they may experience local conformational changes. Thus the N-terminal subdomain can also affect the structure of the C-terminal subdomain.

3.3.2 Breaking the salt bridge destabilized the protein

The apparent stabilities of the HP67E39Q, HP67K70M and HP67E39QK70M variants were measured by CD monitored urea denaturations at 222 nm (figure 3.9). The thermodynamic parameters of the variants are shown in table 3.1. Again it is important to stress that these experiments are monitoring the apparent stability of the C-terminal subdomain. Substitution of Glu39 by Gln significantly destabilized the C-terminal subdomain by about 2 kcal/mol. The HP67K70M and HP67E39QK70M variants are also about 0.3 kcal/mol destabilized comparing to the wildtype, while they are more stable

than the HP67E39Q mutant, likely because the Lys70 to Met mutation can locally stabilize the C-terminal subdomain. Studies of the isolated C-terminal subdomain have shown that a K70M mutation stabilizes the subdomain in isolation (Yuan Bi personal communication). m-values of the HP67 variants are much smaller than the wildtype, which means the exposed surface area change upon unfolding of the variants is smaller than the wildtype (*20, 23*). The most likely reason is that the surface area of the variants under native conditions is larger than the wildtype, due to partially unfolding of the N-terminal subdomain in the variants. Comparing the midpoint of the unfolding, c_m is also a useful way to estimate the effect of mutation. The urea c_m-value of HP67E39Q is 1.2 mol lower than the wildtype while the c_m of HP67K70M and HP67E39QK70M variants are 1.8 and 1.5 mol higher than the wildtype respectively. Guanidine denaturations were also performed for the three variants and the results consistently showed that substitution of Glu39 by Gln destabilizes the protein (figure 3.10). The stabilities derived from guanidine denaturation are smaller than from urea denaturation. This might be because guanidine is a salt thus it can screen electrostatic interactions (*24-27*).

CD was also used to monitor the thermal stability by following the signal at 222 nm as a function of temperature. The melting temperature of HP67E39Q is 4^oC lower than the wildtype while the T_m of HP67K70M and HP67E39QK70M variants are 7 to 8^oC higher than the wildtype (figure 3.11). This is because of the local stabilization of the C-terminal subdomain by the Lys70 to Met substitution. What needs to be pointed out is that the equation we used to fit the denaturation curves are for a two-state folding mechanism. Thus is strictly inaccurate for HP67 because it proved undergoes multistate unfolding. However, the CD signal at 222 nm primarily arises from helical structure which is mainly from the C-terminal subdomian. Thus the CD experiments report primarily on the unfolding of the C-terminal subdomain which is thought to undergo two state unfolding.

The effects of replacement of Glu39 by Leu were also studied and compared to the HP67E39Q variant. The ^1H-^{15}N HSQC spectrum is similar to the HP67E39Q variant, also indicating a partially unfolded N-terminal subdomain (figure 3.12). The denaturation curves show that the melting point of HP67E39L is 4^oC higher than HP67E39Q but the same as wildtype, and the apparent free energy of unfolding (from urea induced unfolding) is about 1.3 kcal/mol higher than HP67E39Q but 0.7 kcal/mol less stable than wildtype (figure 3.13). This means the C-terminal subdomain in HP67E39L is more stable than in HP67E39Q. This stabilization may arise from the contribution of leucine to the hydrophobic core, forming hydrophobic interactions with other residues. The apparent m-value of HP67E39L is 0.46 kcal/mol M^{-1}, which is also larger than the value for HP67E39Q, 0.37 kcal/mol M^{-1} but smaller than wildtype. To summarize, we observed a decreased free energy of unfolding and a smaller m-value compared to the wildtype but the effects are smaller than for the HP67E39Q mutant. The thermodynamic data for all variants is summarized in table 3.1.

3.3.3 The effect of a H41Y substitution on the HP67E39Q variant

His41 has been shown to act as a pH sensitive switch for N-terminal subdomain folding in HP67 (*4*). Previous work showed a H41Y substitution can stabilize the protein (*7*). Thus we explored the effect of a H41Y substitution on the HP67E39Q variant. We prepared the HP67E39QH41Y double mutant and the structure was compared to the wildtype by NMR. In the one-dimensional ^{1}H NMR spectrum, there is a broad peak at around 0.2 ppm which is the chemical shift for the methyl group of V33 (figure 3.14). The broadening of the resonance may arise from rapid exchange between the folded and unfolded conformations. And the spectrum recorded at lower temperature shows no significant change from the one recorded at room temperature. The ^{1}H-^{15}N HSQC spectrum of HP67E39QH41Y is more resolved than the spectrum of HP67E39Q and some resonances from the N-terminal residues have chemical shifts similar to the wildtype values (figure 3.15). However, it is still less resolved than the spectrum of the wildtype indicating partial unfolding of the N-terminal subdomain. The results show that H41Y substitution can stabilize the N-terminal subdomain and help maintain more native structure compared to the HP67E39Q variant. But the N-terminal subdomain is still partially unfolded without the salt bridge.

The stability of HP67E39QH41Y was measured by CD monitored urea denaturation (figure 3.16). The apparent free energy of unfolding is 3.08 kcal/mol, which is 1.2 kcal/mol lower than the wildtype value but 0.7 kcal/mol higher than HP67E39Q. The value for HP67H41Y is 5.4 kcal/mol thus the E39Q mutation leads to an apparent destabilization of 2.0 kcal/mol in the wildtype background and to a 2.3 kcal/mol destabilization in the H41Y background. The midpoint, c_m, of the urea denaturation of HP67E39QH41Y is the same as the wildtype and 1.2 mol higher than HP67E39Q. This is expected since the H41Y substitution stabilizes the C-terminal subdomain and the CD signal at 222 nm primarily arises from the helical structure in the C-terminal subdomain. The melting temperature, T_m, of HP67E39QH41Y is 6^{o}C and 2^{o}C lower than the wildtype and HP67E39Q variant respectively. Compared to HP67H41Y, the E39Q mutation decreases the melting temperature by 16 ^{o}C in the H41Y background while in the wildtype background it leads to a decrease of 4 ^{o}C. According to the urea denaturation data, the melting point of HP67E39QH41Y should be higher than HP67E39Q while in fact we observed a lower value. However changes in Tm do not have to correlate with changes in $\Delta\Delta G^{0}_{m}$ at a fixed temperature.

The m-value reports on the surface area change upon unfolding. For HP67E39QH41Y variant, the apparent m-value is 0.43 kcal/mol M^{-1}, which is higher than HP67E39Q, 0.37 kcal/mol M^{-1}, but lower than the wildtype, 0.60 kcal/mol M^{-1} and lower than the value for HP67H41Y which is the same as wildtype . This is consistent with the picture of partial unfolding of the N-terminal subdomain from NMR results. The buried salt bridge between E39 and K70 is essential to keep the N-terminal subdomain folded even in the H41Y background.

3.3.4 The effect of a K38M substitution on HP67

Another variant was designed which keeps the residues involved in the salt bridge but mutates the residue K38. A HP67K38M variant was prepared since K38 is next to E39 thus a K38M substitution might perturb electrostatic interactions made by E39 and indirectly affect the salt bridge. The structure of this variant was explored by recording the ^{1}H-^{15}N HSQC spectrum. The spectrum is similar to the spectra of the other variants with a region of poorly resolved N-terminal domain resonances indicating the unfolding of the N-terminal subdomain (figure 3.17). Although the residues involved in the salt bridge remain, the K38M substitution likely affects the residue next to it, E39, changing its local conformation and thus disrupting the salt bridge. The stability of HP67K38M was measured by CD monitored thermal and urea denaturation (figure 3.18). The melting temperature of HP67K38M is $3^{o}C$ lower than the wildtype and the free energy of unfolding is 1.0 kcal/mol lower. The m-value of HP67K38M is 0.43 kcal/mol M^{-1}, which is 0.06 kcal/mol M^{-1} higher than HP67E39Q and 0.17 kcal/mol M^{-1} lower than the wildtype indicating a smaller surface area change upon unfolding. This is consistent with the HSQC result suggesting the partial unfolding of the N-terminal subdomain. The sidechain of K38 points out to the solvent in both the NMR and X-ray structures. There is less possibility that K38 can form interactions with C-terminal subdomain. Thus the destabilization caused by the K38M mutation is likely arising from the disruption of the salt bridge between E39 and K70 by perturbing the conformation of E39.

3.4 Discussion

E39 and K70 are buried in the hydrophobic core of HP67. Burial of charged residues is a very unstable arrangement unless a salt bridge formed. The close distance between these two residues in the protein structure indicates the existence of a salt bridge (*59, 101*). A series of variants were designed to break this salt bridge. All of them resulted in destabilization of the N-terminal subdomain and its partially unfolding. A H41Y substitution stabilizes the protein but it still does not induce the N-terminal subdomain to fold without the salt bridge. Thus the salt bridge between E39 and K70 is indispensable for the N-terminal subdomain to fold to the native structure. The extended hydrophobic core is essential to ensure the correct N-terminal fold. Since the salt bridge is buried in the hydrophobic core, the E39Q and K70M substitutions may disrupt the hydrophobic core resulting in the unfolding of N-terminal subdomain. K38 is located next to E39 and since they are oppositely charged, it could contribute to stabilizing the conformation of E39. Thus it is not surprising that the K38M substitution also results in a partially unfolded N-terminal subdomain. The interactions between the two subdomains are not only crucial for the correct fold of N-terminal subdomain, they are also likely to be essential for the F-actin binding function. Most of the residues responsible for binding are located in the C-terminal subdomain, but HP35 in isolation can not bind to F-actin tightly. Compared to the structure of HP35, the conformations of some residues in the C-terminal

subdomain of HP67 are rearranged (*16*). The interactions between these two subdomains may account for the rearrangements. Thus a correctly folded N-terminal subdomain is likely essential for orienting and positioning the C-terminal residues to function properly.

HP67 is a small globular protein, however it surprisingly undergoes multistate unfolding, but the unfolding of the two subdomains are not completely independent. An intermediate with a folded C-terminal subdomain and a partially unfolded N-terminal subdomain was suggested to be on the unfolding pathway. Computational studies lead to the interesting suggestion that formation of the intermediate is induced by disruption of a hydrogen bond network involving His-41 (*17*). In the variants without the salt bridge we observed structures with a partially unfolded N-terminal subdomain. This demonstrates that partial unfolding can also be induced by disruption of the salt bridge. The variants we designed may mimic the intermediate structure proposed in the computational studies and detected by our previous NMR relaxation studies. However, further experiments will need to be done to verify this.

Table 3.1 Thermodynamic parameters for the unfolding of WT HP67 and variants. All experiments monitored the CD signal at 222 nm and thus are primarily reporting on the apparent stability of the C-terminal subdomain. The standard errors to the fit are given after the ± symbol. The values for wildtype HP67 and HP67H41Y were taken from reference (*95*). Experiments were conducted in 10mM sodium phosphate buffer at pH 6.5 except the guanidine denaturation experiments which were conducted at pH 7.0. Urea and guanidine induced unfolding was used to estimate ΔG^0 at 25°C. T_m values were determined from fitting the thermal unfolding curves and c_m values were determined by taking derivatives of the urea unfolding curves. c_m is the midpoint concentration for denaturant unfolding.

			ΔG^0 (kcal/mol)		m (kcal/mol M^{-1})		
	T_m (°C)	$\Delta H^0(T_m)$ (kcal/mol)	urea	Gdn	urea	Gdn	c_m(Urea) (mol)
HP67	76±0.7	61±5.0	4.3±0.07	2.75±0.09	0.58±0.01	0.87±0.02	7.2
HP67E39Q	72±0.7	36±1.5	2.35±0.11	2.63±0.10	0.37±0.02	0.83±0.03	6.0
HP67K70M	83±1.0	32±0.5	4.06±0.52	3.08±0.09	0.40±0.01	0.75±0.02	9.0
HP67E39Q K70M	84±0.9	33±0.4	3.99±0.30	3.19±0.09	0.46±0.01	0.75±0.02	8.7
HP67H41Y	86±0.5	71±4.7	5.38±0.21	4.10±0.06	0.60±0.01	1.12±0.01	9.0
HP67E39Q H41Y	70±0.7	44±2.9	3.08 ± 0.04		0.43±0.01		7.2
HP67E39L	76±0.7	33±0.8	3.67±0.05		0.49±0.01		7.8
HP67K38M	73±0.8	38±1.6	3.36±0.05		0.45±0.01		7.7

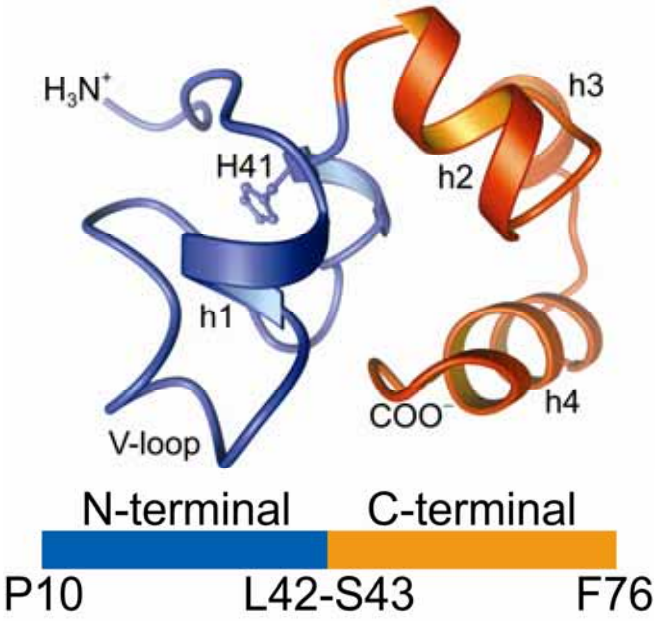
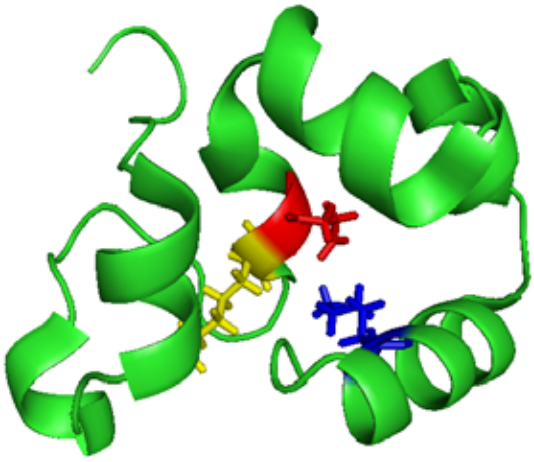


Figure 3.1 Ribbon diagram of villin headpiece HP67. The NMR structure was used to make the figure (PDB code: 1QQV). The C-terminal subdomain is highlighted in orange and the N-terminal subdomain is highlighted in blue. The position of His 41 is shown in ball-and-stick format. The N and C-termini are labeled and the four helices are labeled h1 through h4. The figure was created by PyMol.

a



b

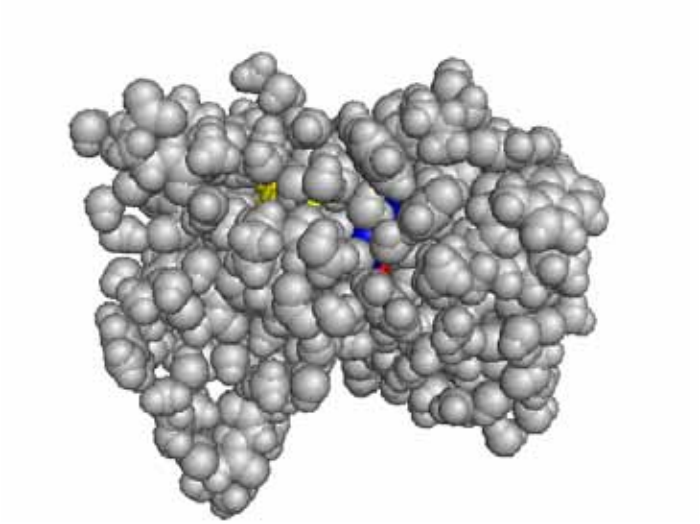


Figure 3.2 (a) Ribbon diagram and (b) space filling graph of villin headpiece 67. The X-ray structure was used to make the figures (PDB code: 1YU5). Residue E39 is shown in stick format and colored red, residue K70 is colored blue and K38 is colored yellow.

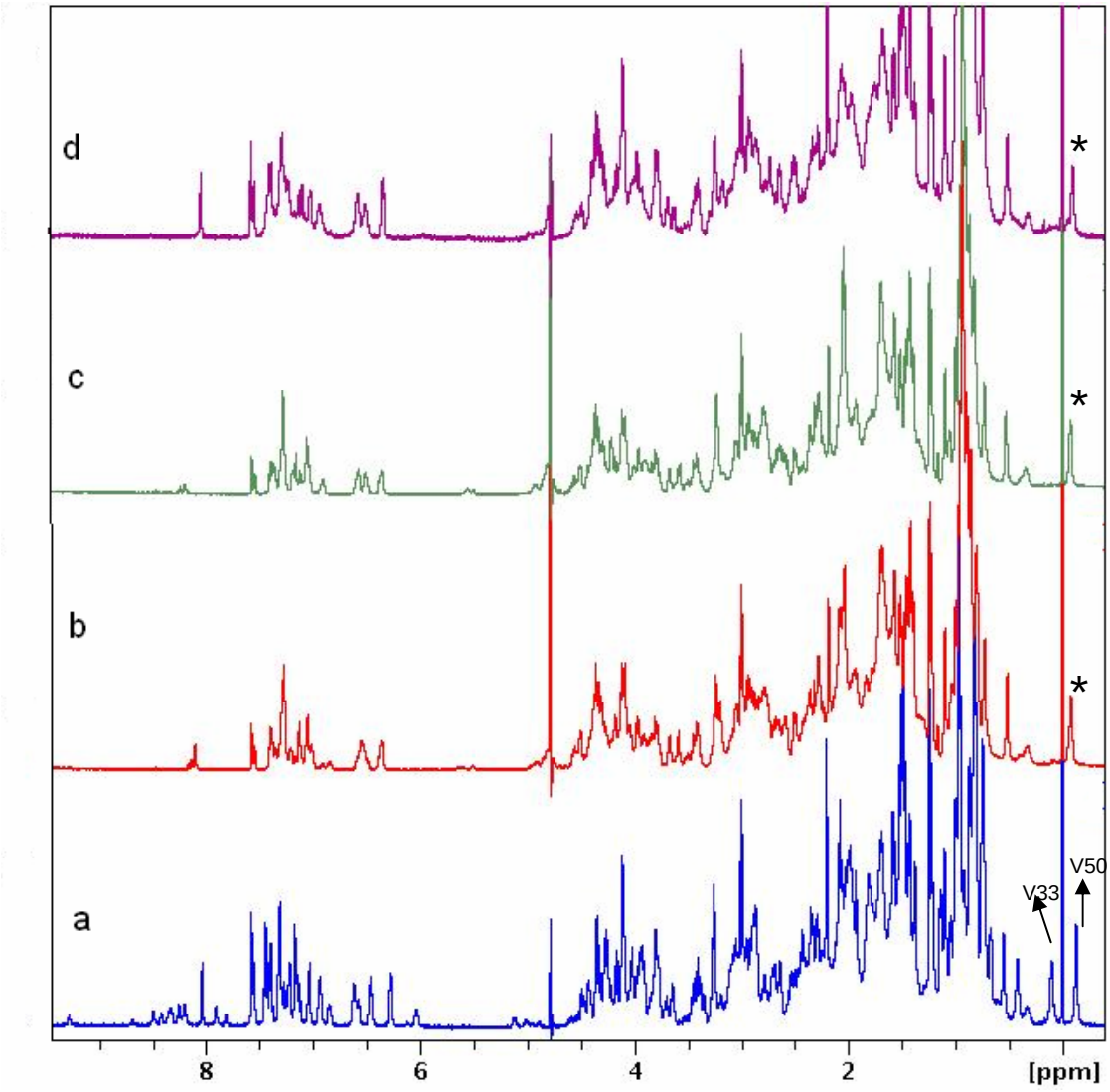


Figure 3.3 ^{1}H spectra of (a) WT HP67; (b) HP67E39QK70M; (c) HP67K70M; (d) HP67E39Q. The spectra were recorded in 99.9% D_2O containing 10 mM sodium phosphate at pD 6.5, 25^{o}C. pD refers to the uncorrected meter reading in D_2O. The methyl resonances of V33 and V50 are shown in (a) by arrows and the V50 resonances in (b), (c) and (d) are labeled by a star. The resonances in the downfield region with chemical shifts higher than 8ppm are from exchangeable amide protons. The difference in the intensity of these resonances may due to sample preparation.

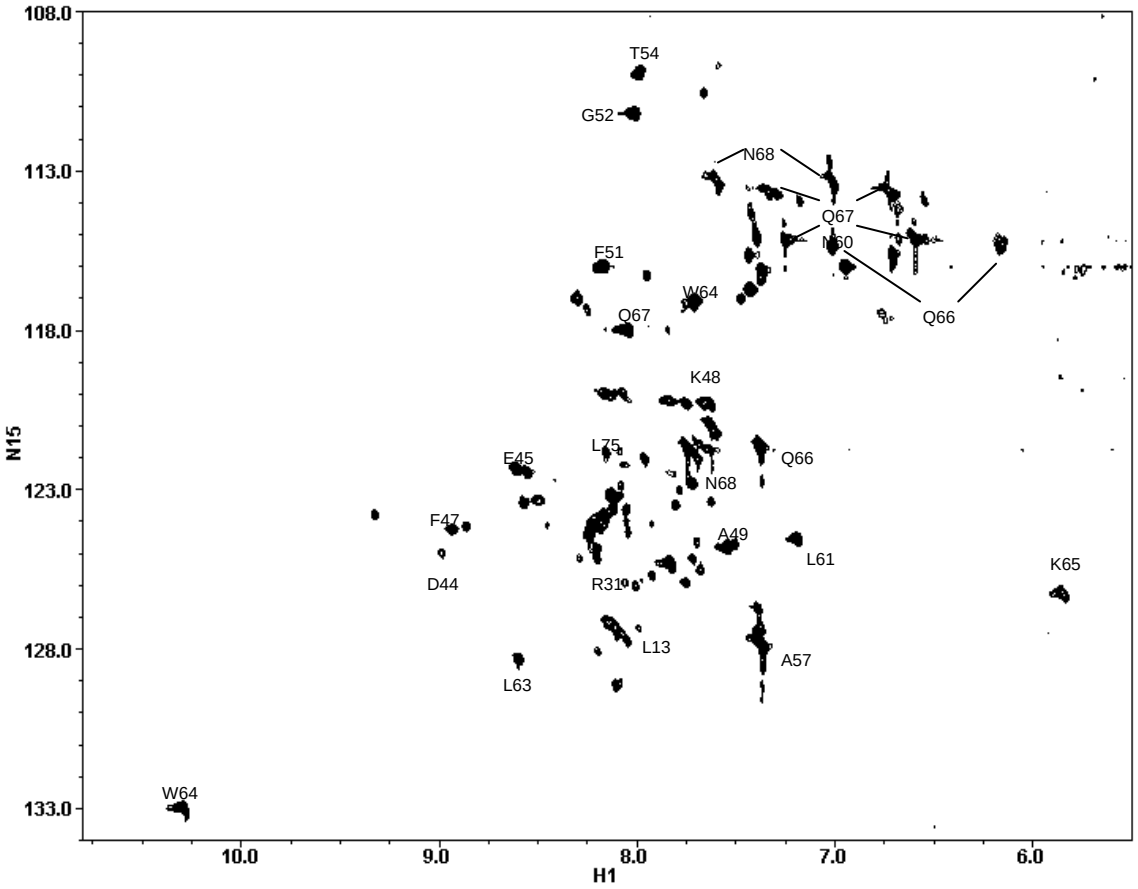


Figure 3.4 ^{1}H-^{15}N HSQC spectrum of HP67E39QK70M. The spectrum was recorded in 10% D_2O/90% H_2O with 10mM sodium phosphate at pH 7.0, 25^{o}C.

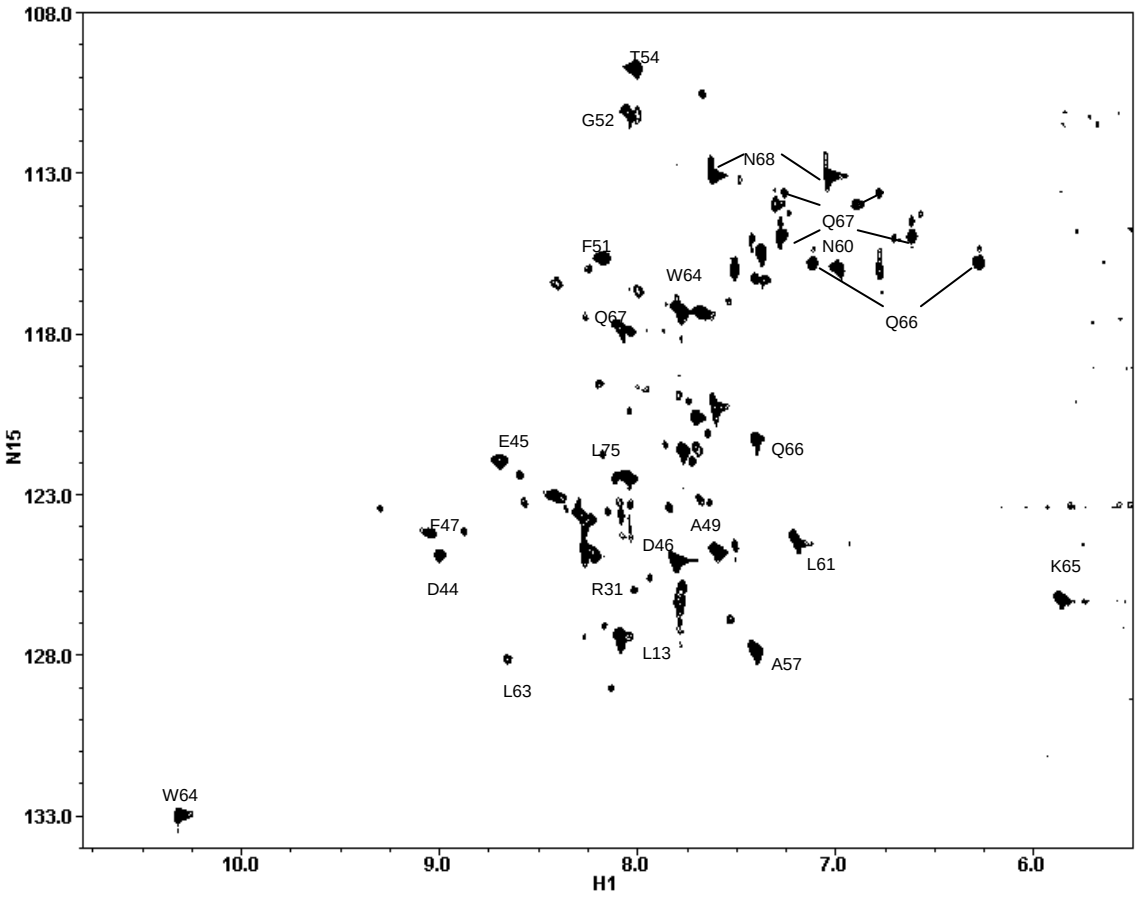


Figure 3.5 ^{1}H-^{15}N HSQC spectrum of HP67E39Q. The spectrum was recorded in 10% D_2O/90% H_2O with 10mM sodium phosphate at pH 7.0, 25^{o}C.

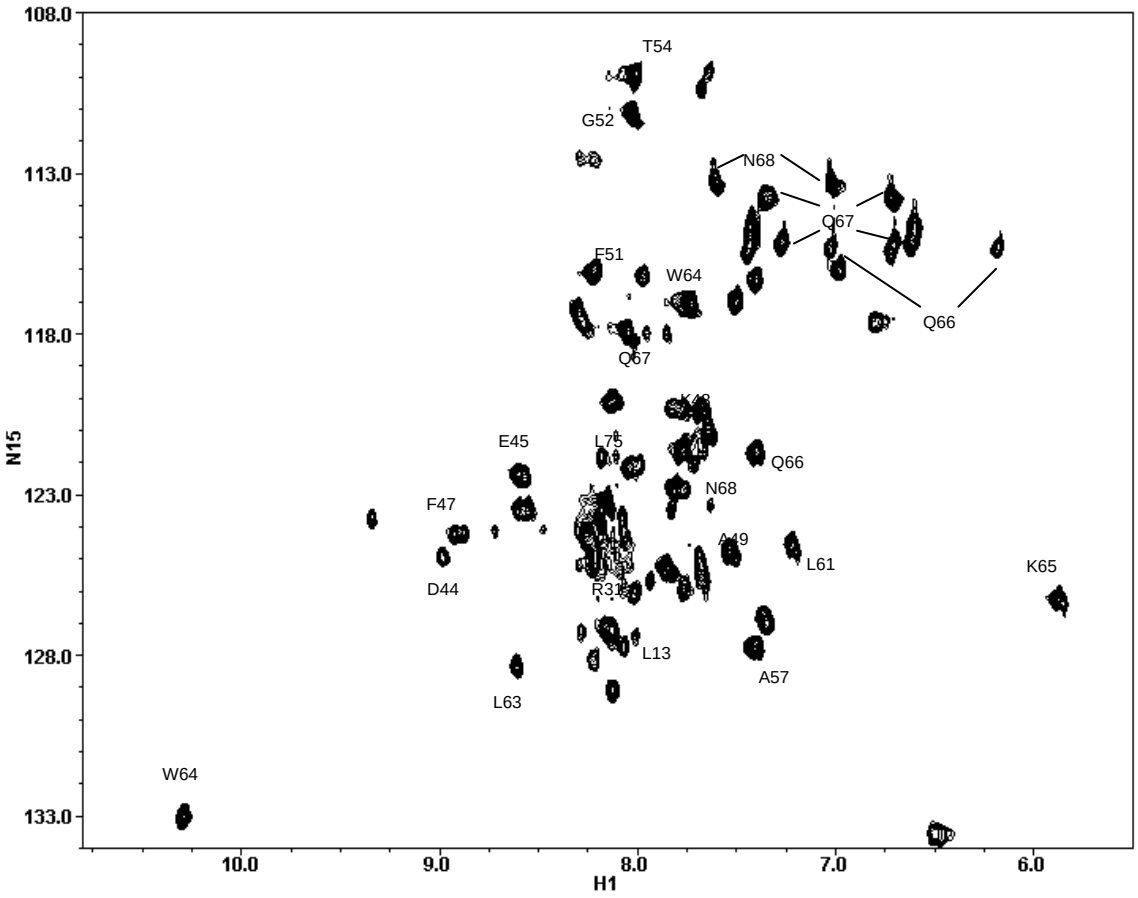


Figure 3.6 ^{1}H-^{15}N HSQC spectrum of HP67K70M. The spectrum was recorded in 10% D_2O/90% H_2O with 10mM sodium phosphate at pH 7.0, 25^{o}C.

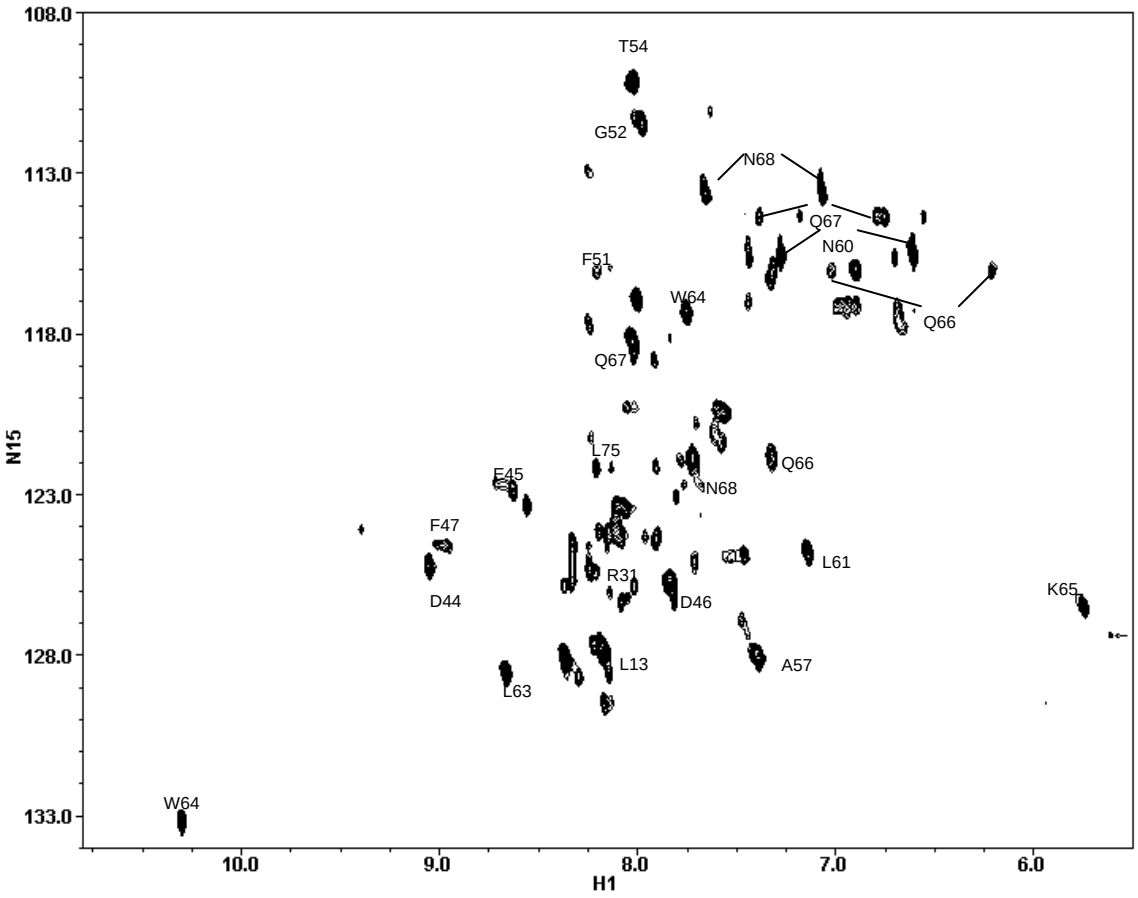


Figure 3.7 ^{1}H-^{15}N HSQC spectrum of HP67E39QK70M. The spectrum was recorded in 10% D_2O/90% H_2O with 10mM sodium phosphate at pH 7.0, 5^{o}C.

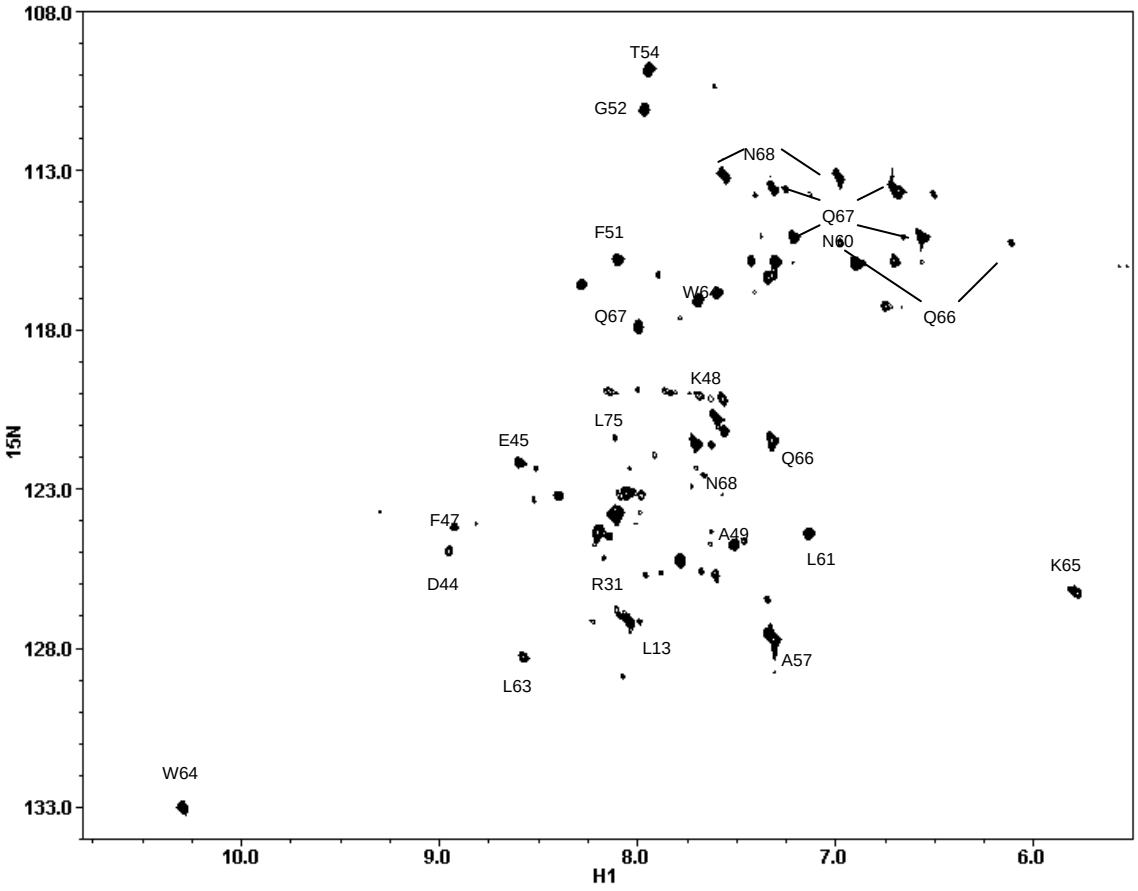


Figure 3.8 ^{1}H-^{15}N HSQC spectrum of HP67E39QK70M in the presence of 1M TMAO. The spectrum was recorded in 10% D_2O/90% H_2O with 10mM sodium phosphate at pH 7.0, 25^{o}C.

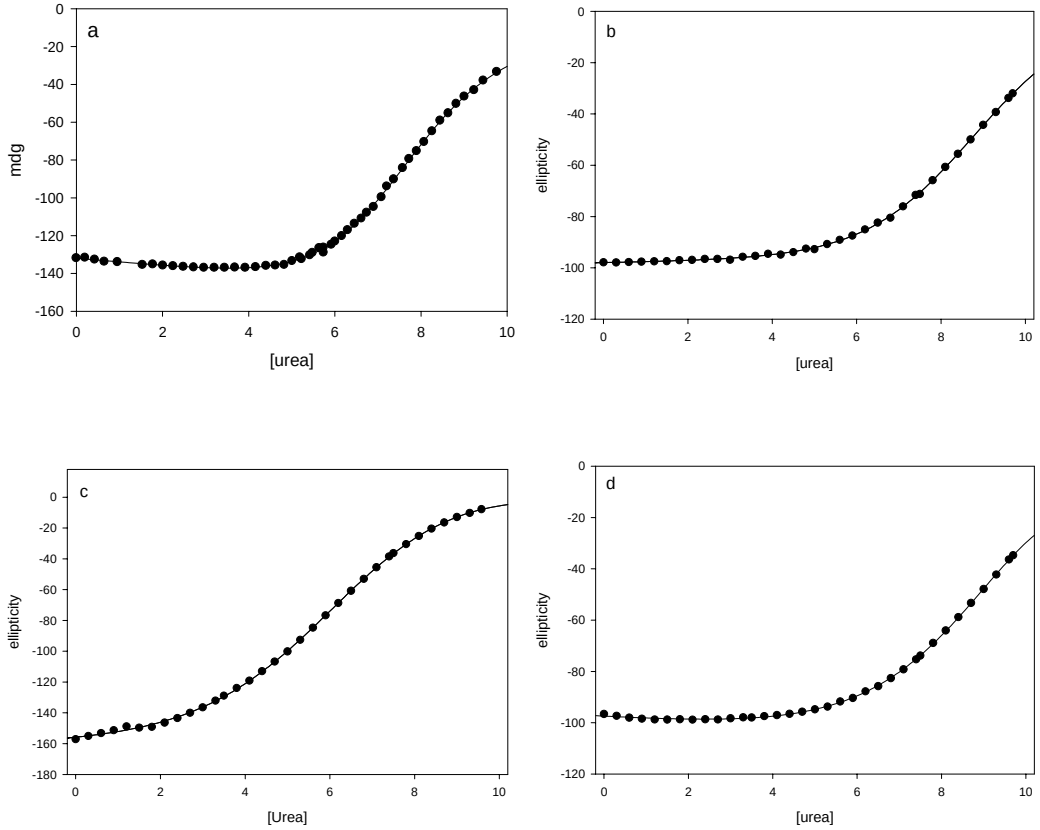


Figure 3.9 Plots of the measured elipticity vs. [urea] showing the fits. (a) HP67; (b) HP67E39QK70M; (c) HP67E39Q; (d) HP67K70M. All experiments were done in 10 mM sodium phosphate buffer at pH 6.5, 25^{o}C. The solid line is the best fit to equation 3-1.

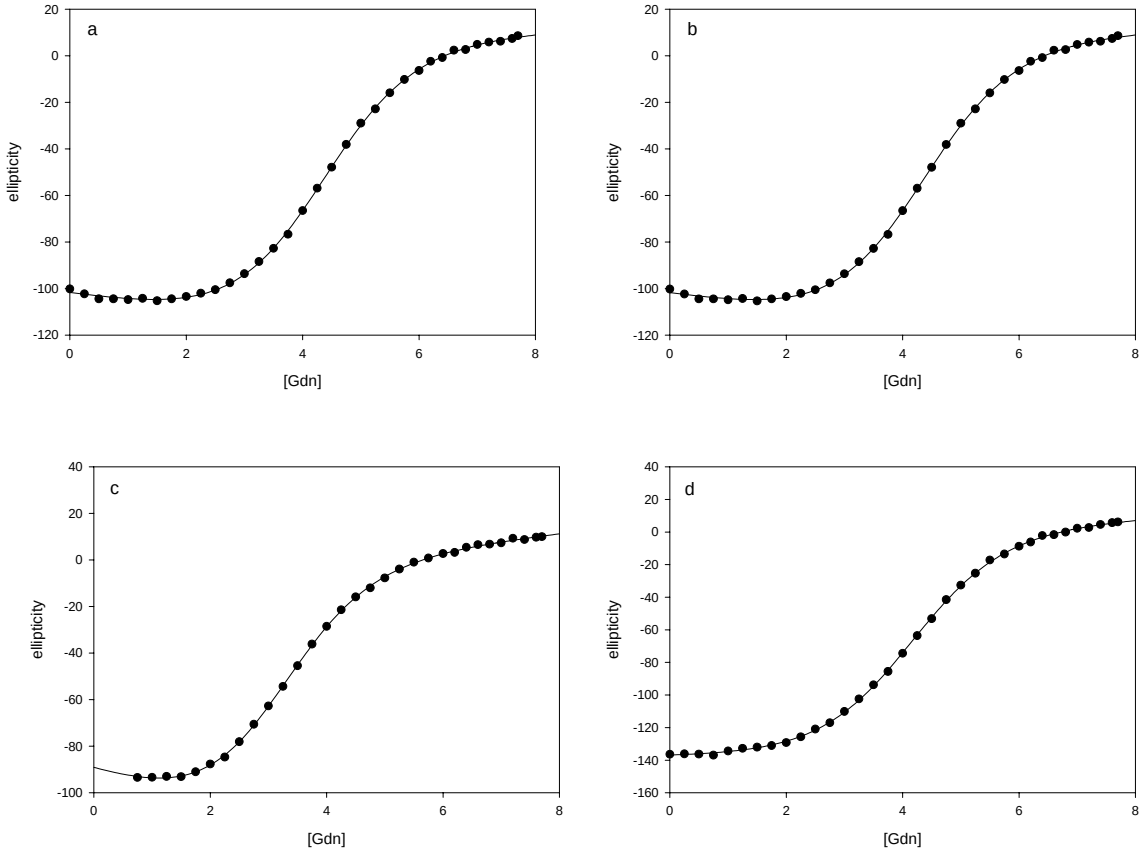


Figure 3.10 Plots of the measured elipticity vs. guanidine hydrochloride concentration showing the fits. (a) HP67; (b) HP67E39QK70M; (c) HP67E39Q; (d) HP67K70M. All experiments were done in 10 mM sodium phosphate buffer at pH 7.0, 25^{o}C. The solid line is the best fit to equation 3-1.

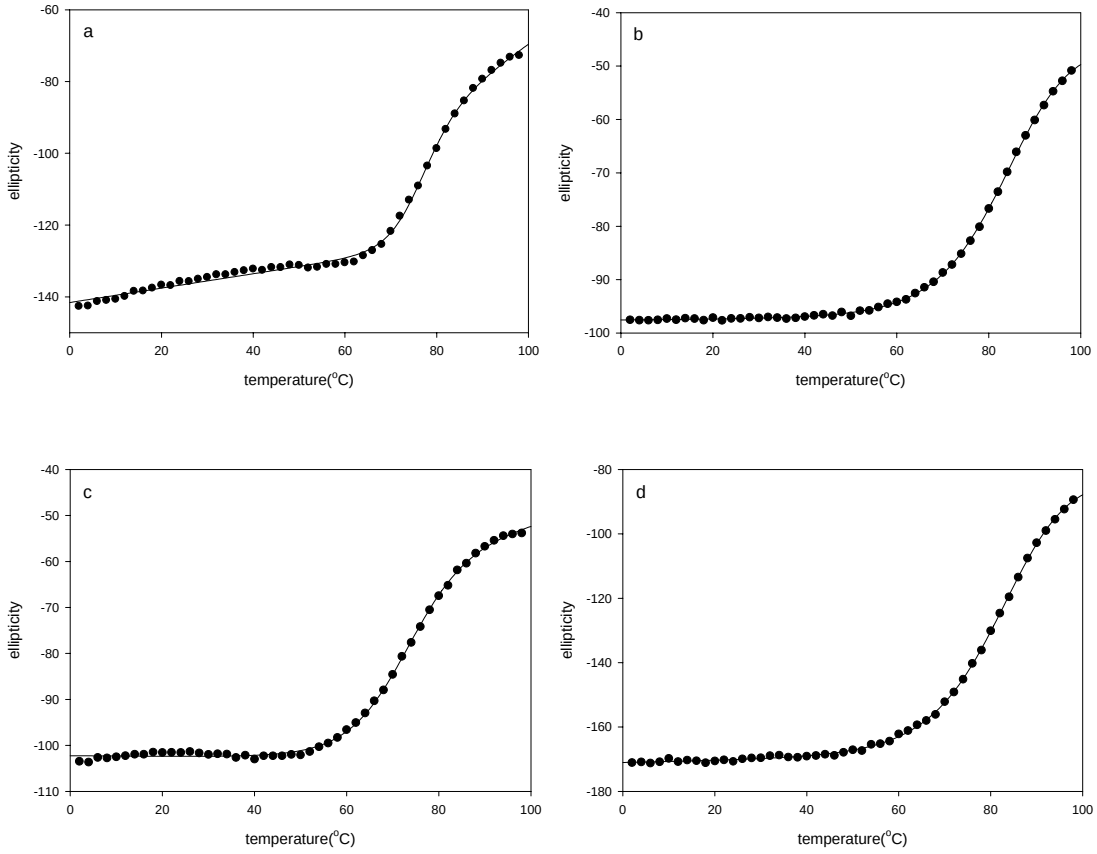


Figure 3.11 Plots of the measured elipticity vs. temperature showing the fits. (a) HP67; (b) HP67E39QK70M; (c) HP67E39Q; (d) HP67K70M. All experiments were done in 10 mM sodium phosphate buffer at pH 6.5. The solid line is the best fit to equation 3-1.

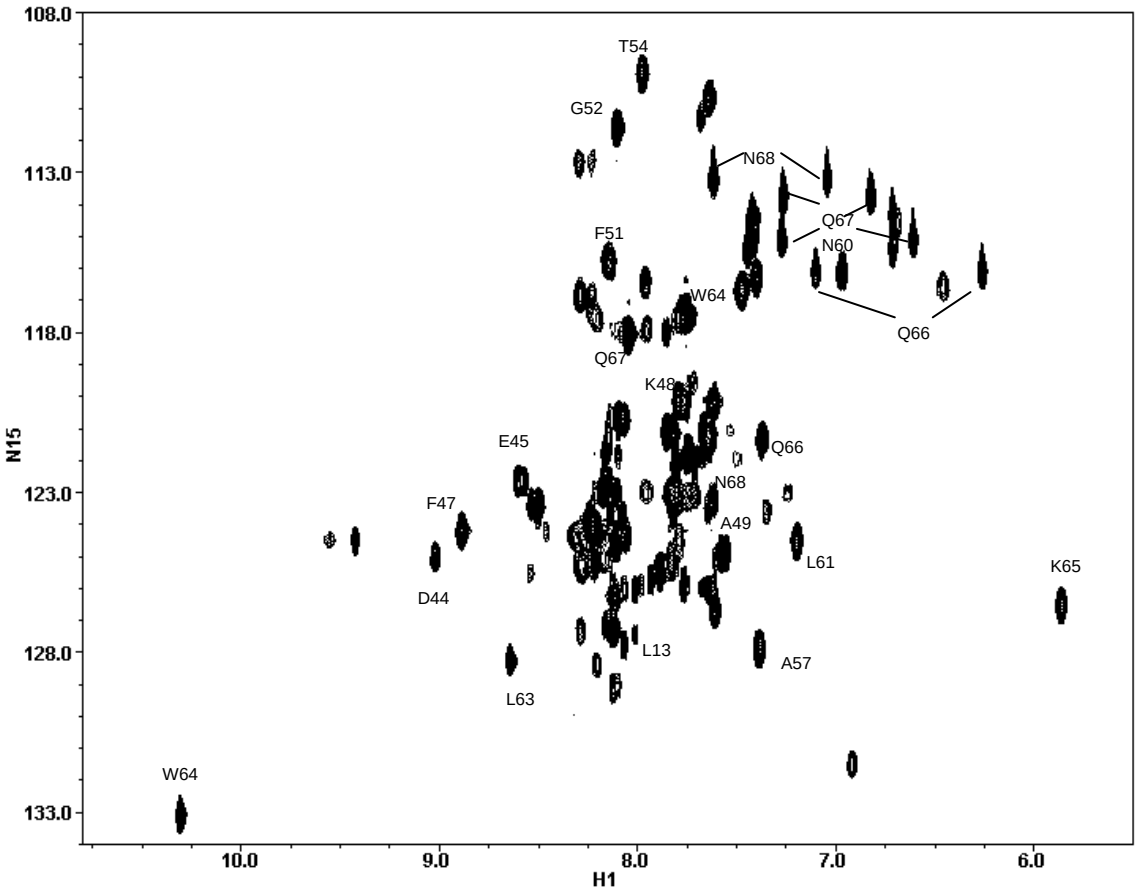


Figure 3.12 ^{1}H-^{15}N HSQC spectrum of HP67E39L. The spectrum was recorded in 10% D_2O/90% H_2O with 10mM sodium phosphate at pH 7.0, 25^{o}C.

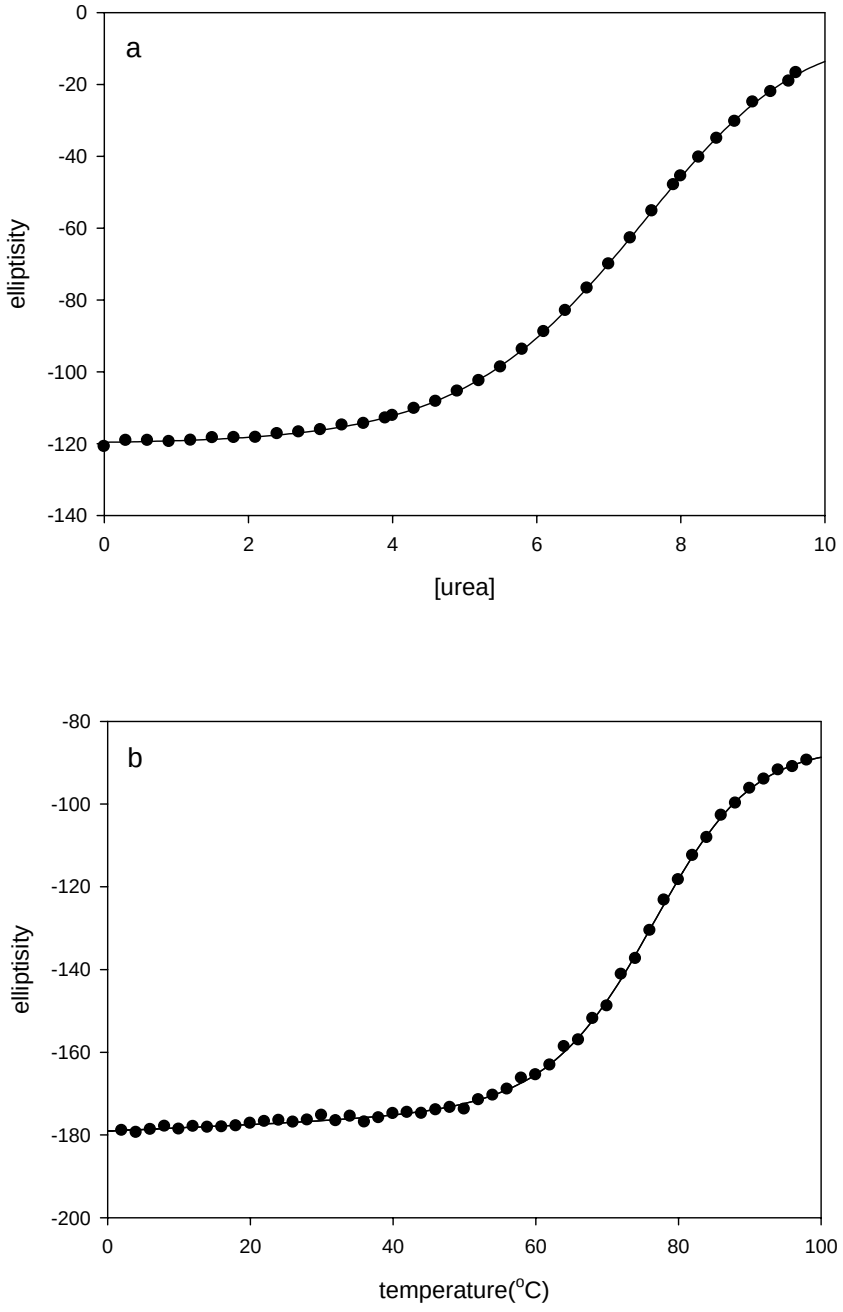


Figure 3.13 Plots of the measured elipticity vs. (a) [urea] and (b) temperature showing the fits of HP67E39L. All experiments were done in 10 mM sodium phosphate buffer at pH 6.5 and the urea denaturation curve was monitored at 25°C.

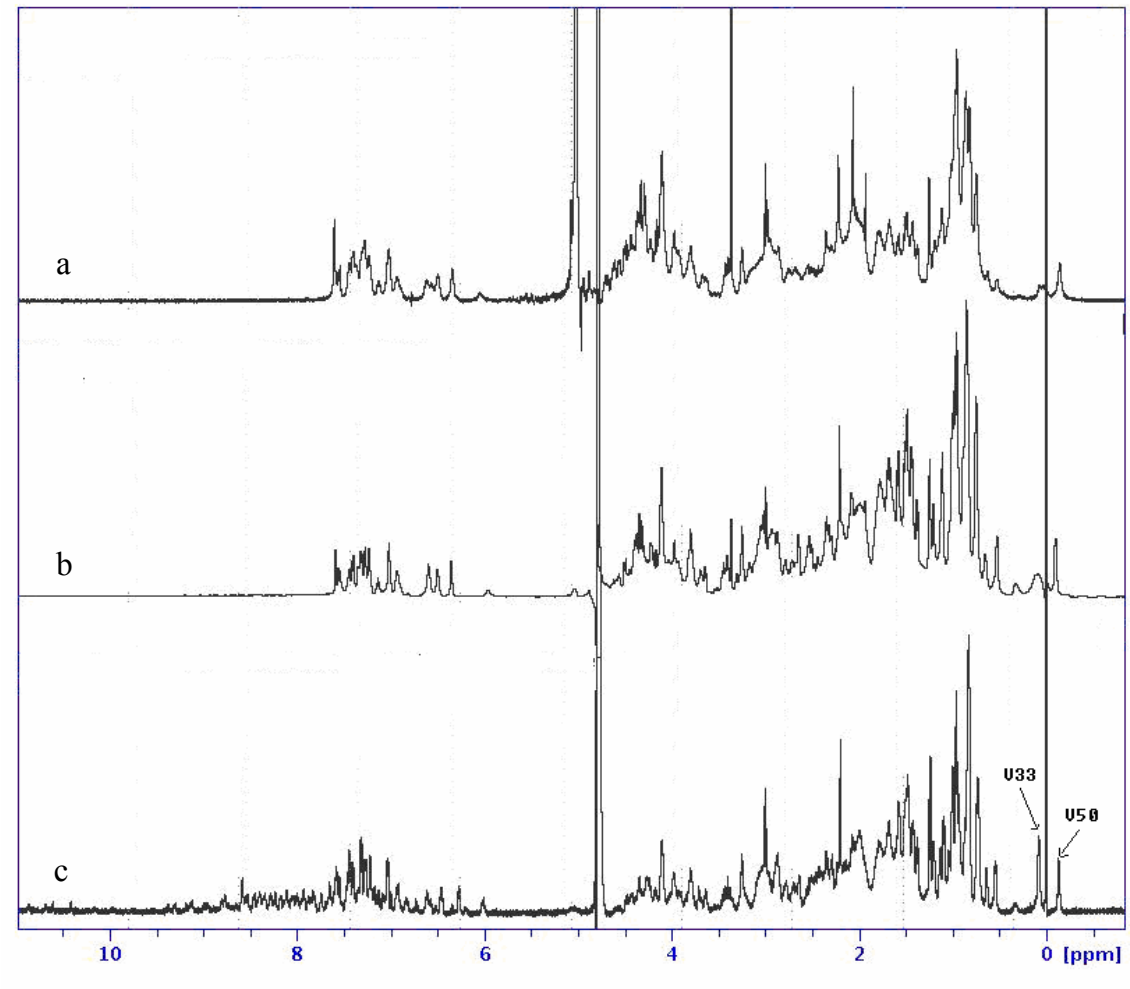


Figure 3.14 ^{1}H spectra of (a) HP67E39QH41Y at 5°C; (b) HP67E39QH41Y at 25°C; (c) HP67H41Y at 25°C. The spectra were recorded in 99.9% D_2O containing 10 mM sodium phosphate at pD 6.5. pD refers to the uncorrected meter reading in D_2O. The methyl resonances of V33 and V50 are shown in (c) by arrows.

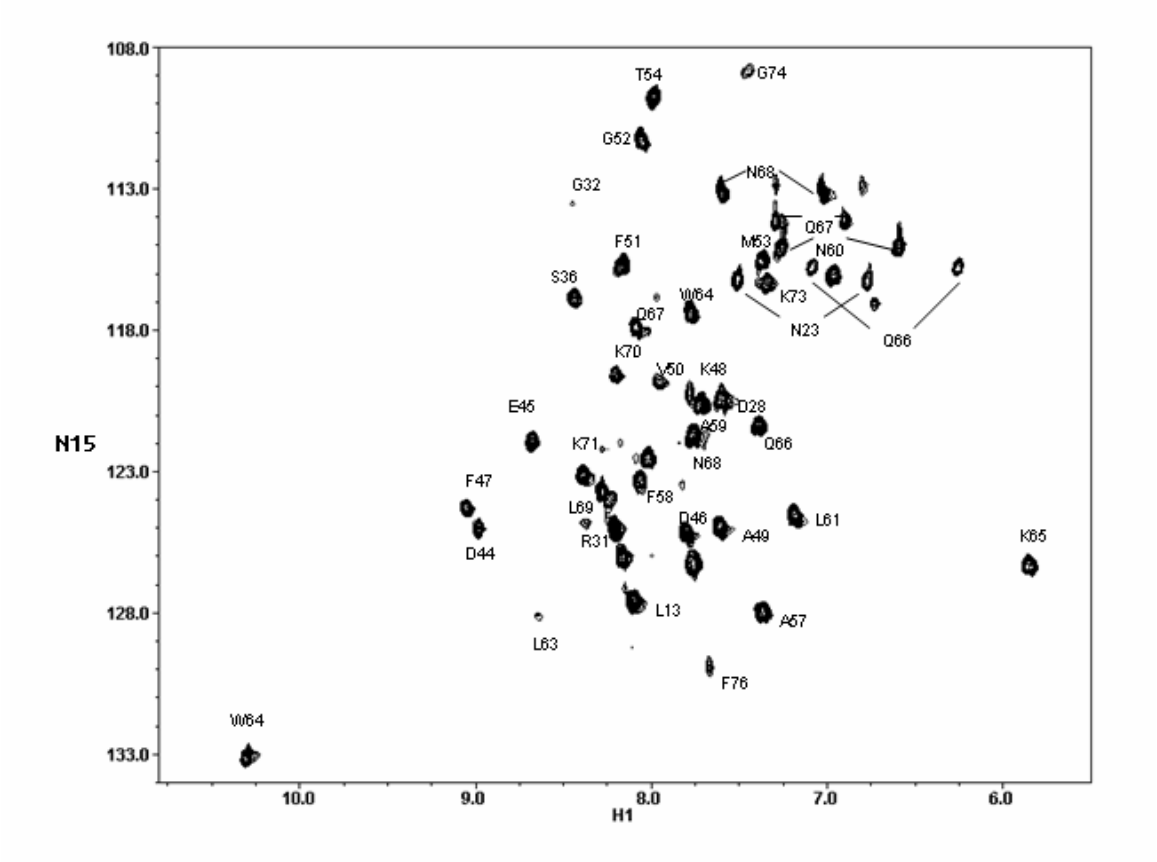


Figure 3.15 ^{1}H-^{15}N HSQC spectrum of HP67E39QH41Y. The spectrum was recorded in 10% D_2O/90% H_2O with 10mM sodium phosphate at pH 7.0, 25^{o}C.

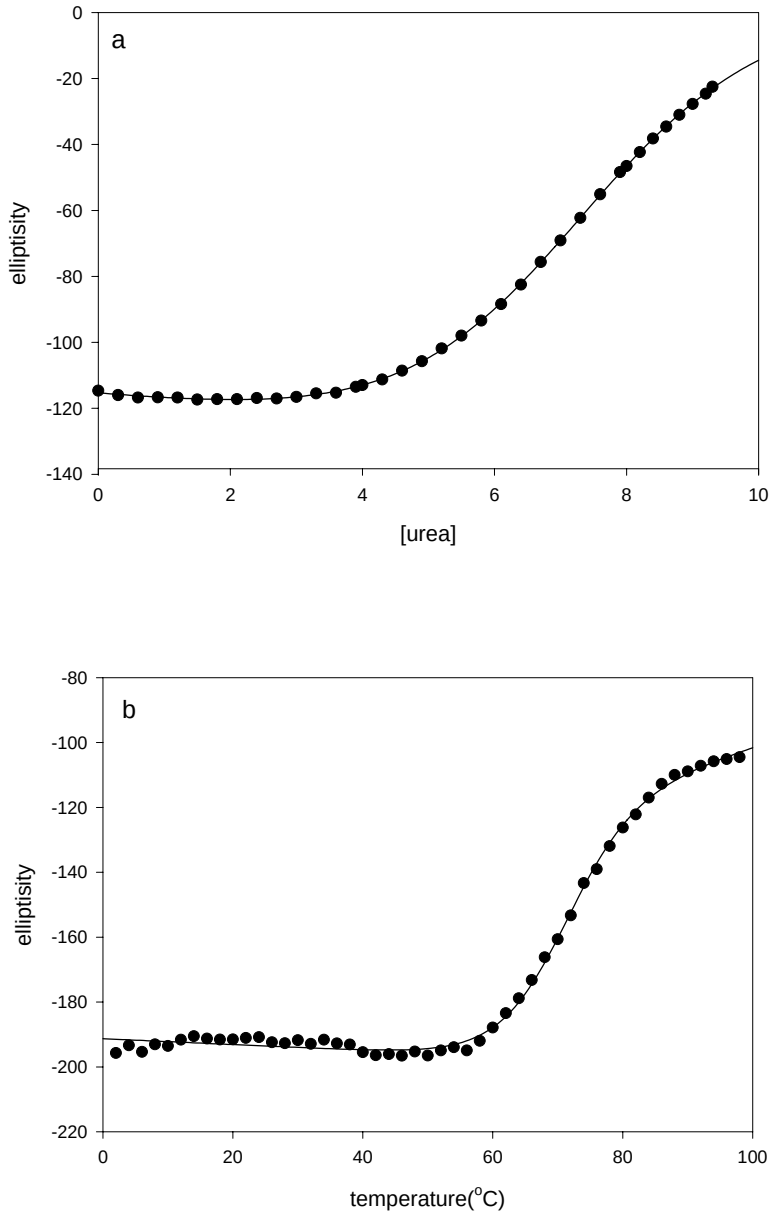


Figure 3.16 Plots of the measured elipticity vs. (a) temperature and (b) [urea] for HP67E39QH41Y. All experiments were done in 10 mM sodium phosphate buffer at pH 6.5 and the urea denaturation curve was monitored at 25^{o}C.

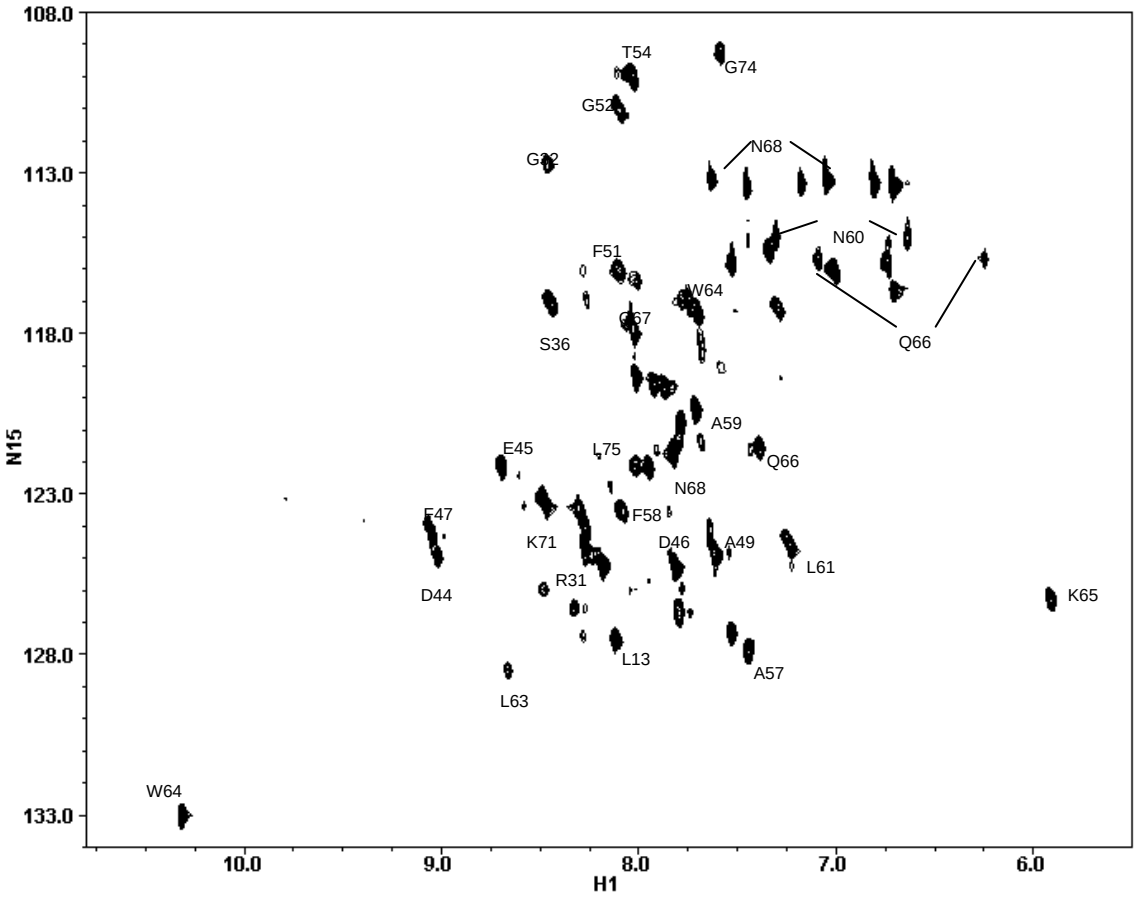


Figure 3.17 ^{1}H-^{15}N HSQC spectrum of HP67K38M. The spectrum was recorded in 10% D_2O/90% H_2O with 10mM sodium phosphate at pH 7.0, 25^{o}C.

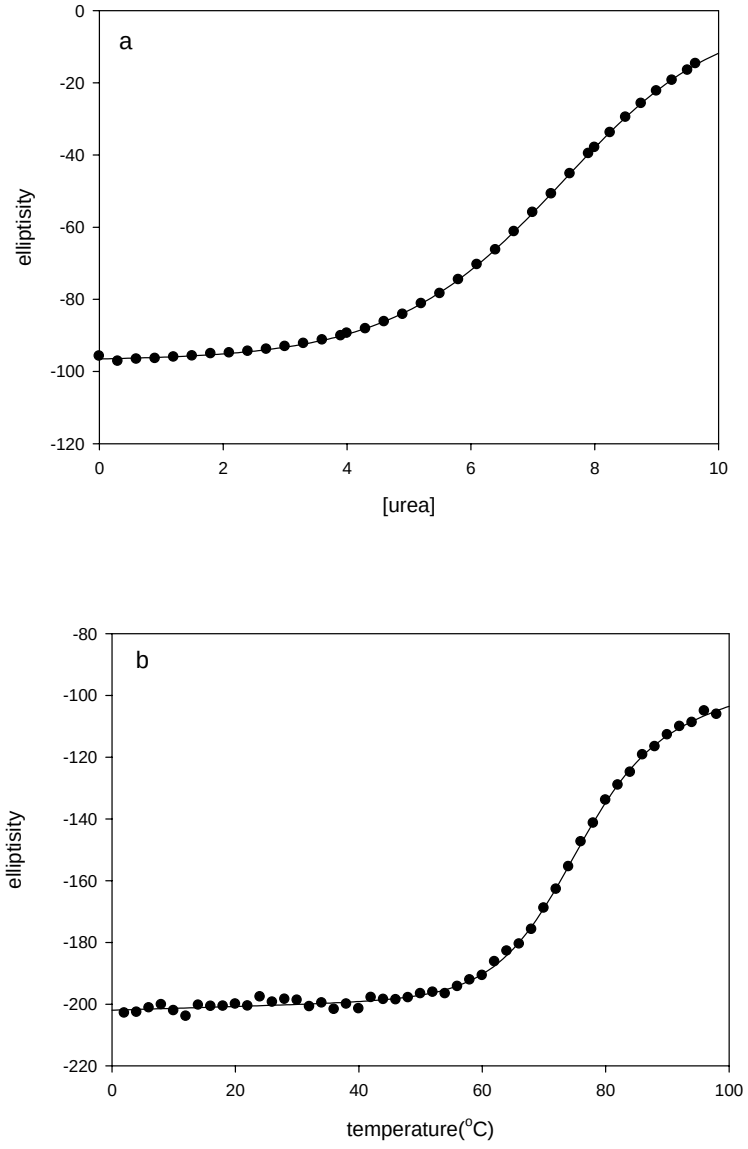


Figure 3.18 Plots of the measured elipticity vs. (a) [urea] and (b) temperature showing the fits for HP67K38M. All experiments were done in 10 mM sodium phosphate buffer at pH 6.5 and the urea denaturation curve was monitored at 25^{o}C.

3.5 References

1. Azim, A. C., Knoll, J. H., Beggs, A. H., and Chishti, A. H. (1995) Isoform cloning, actin binding, and chromosomal localization of human erythroid dematin, a member of the villin superfamily, *J. Mol. Biol. 270*, 17407-17413.
2. Glenney, J. R., Jr., and Weber, K. (1981) Calcium control of microfilaments: uncoupling of the F-actin-severing and -bundling activity of villin by limited proteolysis in vitro, *Proc. Natl. Acad. Sci. U. S. A. 78*, 2810-2814.
3. Pope, B., Way, M., Matsudaira, P. T., and Weeds, A. (1994) Characterisation of the F-actin binding domains of villin: classification of F-actin binding proteins into two groups according to their binding sites on actin, *Febbs. lett. 338*, 58-62.
4. Vardar, D., Buckley, D. A., Frank, B. S., and McKnight, C. J. (1999) NMR structure of an F-actin-binding "headpiece" motif from villin, *J. Mol. Biol. 294*, 1299-1310.
5. McKnight, C. J., Doering, D. S., Matsudaira, P. T., and Kim, P. S. (1996) A thermostable 35-residue subdomain within villin headpiece, *J. Mol. Biol. 260*, 126-134.
6. McKnight, C. J., Matsudaira, P. T., and Kim, P. S. (1997) NMR structure of the 35-residue villin headpiece subdomain, *Nat. Struct. Biol. 4*, 180-184.
7. Tang, Y., Grey, M. J., McKnight, J., Palmer, A. G., 3rd, and Raleigh, D. P. (2006) Multistate folding of the villin headpiece domain, *J. Mol. Biol. 355*, 1066-1077.
8. Kubelka, J., Eaton, W. A., and Hofrichter, J. (2003) Experimental tests of villin subdomain folding simulations, *J. Mol. Biol. 329*, 625-630.
9. Wang, M., Tang, Y., Sato, S., Vugmeyster, L., McKnight, C. J., and Raleigh, D. P. (2003) Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale, *J. Am. Chem. Soc. 125*, 6032-6033.
10. Duan, Y., Wang, L., and Kollman, P. A. (1998) The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation, *Proc. Natl. Acad. Sci. U. S. A. 95*, 9897-9902.
11. Fernandez, A., Shen, M. Y., Colubri, A., Sosnick, T. R., Berry, R. S., and Freed, K. F. (2003) Large-scale context in protein folding: villin headpiece, *Biochemistry 42*, 664-671.
12. Zagrovic, B., Snow, C. D., Khaliq, S., Shirts, M. R., and Pande, V. S. (2002) Native-like mean structure in the unfolded ensemble of small proteins, *J. Mol. Biol. 323*, 153-164.
13. Zagrovic, B., Snow, C. D., Shirts, M. R., and Pande, V. S. (2002) Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing, *J. Mol. Biol. 323*, 927-937.
14. Duan, Y., and Kollman, P. A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science 282*,

740-744.
15. Khandogin, J., Raleigh, D. P., and Brooks, C. L., 3rd. (2007) Folding intermediate in the villin headpiece domain arises from disruption of a N-terminal hydrogen-bonded network, *J. Am. Chem. Soc. 129*, 3056-3057.
16. Meng, J., Vardar, D., Wang, Y., Guo, H. C., Head, J. F., and McKnight, C. J. (2005) High-resolution crystal structures of villin headpiece and mutants with reduced F-actin binding activity, *Biochemistry 44*, 11963-11973.
17. Grey, M. J., Tang, Y., Alexov, E., McKnight, C. J., Raleigh, D. P., and Palmer, A. G., 3rd. (2006) Characterizing a partially folded intermediate of the villin headpiece domain under non-denaturing conditions: Contribution of His41 to the pH-dependent stability of the N-terminal subdomain, *J. Mol. Biol. 355*, 1078-1094.
18. Friederich, E., Vancompernolle, K., Huet, C., Goethals, M., Finidori, J., Vandekerckhove, J., and Louvard, D. (1992) An actin-binding site containing a conserved motif of charged amino acid residues is essential for the morphogenic effect of villin, *Cell 70*, 81-92.
19. Doering, D. S., and Matsudaira, P. (1996) Cysteine scanning mutagenesis at 40 of 76 positions in villin headpiece maps the F-actin binding site and structural features of the domain, *Biochemistry 35*, 12677-12685.
20. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138-2148.
21. Mello, C. C., and Barrick, D. (2003) Measuring the stability of partly folded proteins using TMAO, *Protein Sci. 12*, 1522-1529.
22. Auton, M., and Bolen, D. W. (2005) Predicting the energetics of osmolyte-induced protein folding/unfolding, *Proc. Natl. Acad. Sci. U. S. A. 102*, 15065-15068.
23. Tanford, C. (1964) Isothermal unfolding of globular proteins in aqueous urea solutions, *J. Am. Chem. Soc. 86*, 2050-2059.
24. Pace, C. N., and Shaw, K. L. (2000) Linear extrapolation method of analyzing solvent denaturation curves, *Proteins Suppl. 4*, 1-7.
25. Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I., and Sanchez-Ruiz, J. M. (1999) Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability, *Biochemistry 38*, 8138-8149.
26. Monera, O. D., Kay, C. M., and Hodges, R. S. (1994) Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions, *Protein Sci. 3*, 1984-1991.
27. Santoro, M. M., and Bolen, D. W. (1992) A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range,

Biochemistry 31, 4901-4907.

28. Lieberthal, W. L., J. S. (1996) Mechanisms of apoptosis and its potential role in enal tubular epithelial cell injury, *Am. J. Phys. 271*, F477-F488.
29. Potten, C. S. (1997) Epithelial cell growth and differentiation. II. Intestinal apoptosis, *Am. J. Phys. 273*, G253-G257.

4. Ionic Strength Dependent Effects in Protein Folding: Analysis of Rate Equilibrium Free Energy Relationships and Their Interpretation

Abstract

The traditional approach to studying protein folding involves applying a perturbation, usually denaturant or mutation, and determining the effect upon the free energy of folding, ΔG^0, and the activation free energy, $\Delta G^{\neq}$. Data collected as a function of the perturbation can be used to construct rate equilibrium free energy relationships which report on the development of interactions in the transition state for folding. We examine the use of ionic strength dependent rate equilibrium free energy relationship in protein folding using the N-terminal domain of L9, a small α-β protein, as a model system. Folding is two state for the range of ionic strength examined, 0.045 to 1.52 M. The plot of $\Delta G^{\neq}$ vs. ΔG^0 is linear (r^2= 0.918) with slope equal to 0.45. The relatively low value of the slope indicates that the ionic strength dependent interactions are modestly developed in the transition state. The slope is, however greater than that of a plot of $\Delta G^{\neq}$ vs. ΔG^0 constructed by varying pH, thus demonstrating directly that ionic strength dependent studies probe more than simple electrostatic interactions. Potential transition movement was probed by analysis of the denaturant, ionic strength cross-interaction parameters. The values are small but non-zero and positive suggesting a small shift of the transition state towards the native state as the protein is destabilized, i.e. Hammond behavior. The complications that arise in the interpretation of ionic strength dependent rate equilibrium free energy relationships are discussed and it is concluded that ionic strength do not provide a reliable indicator of the role of electrostatic interactions. These include incomplete screening of electrostatic interactions, specific ion binding, Hofmeister effects and the potential presence of electrostatic interactions in the denatured state ensemble. Parts of this chapter are taken from the published paper by B. Song & D. P. Raleigh.

4.1 Introduction

The characterization of the structure and energetics of the transition state ensemble (TSE) is a central issue in biophysical studies of protein folding, particularly for single domain structures which fold in a two-state fashion (*1-5*). Experimental investigations often make use of so called rate equilibrium free energy relationships (REFERs) (*6-9*). In the simplest case a linear relationship is observed between the change in the logarithm of the folding rate, $\ln k_f$, with a perturbation and the change in the logarithm of the equilibrium constant, lnK, with the same perturbation. The ratio:

$$\alpha_x = \frac{\partial \ln k_f / \partial x}{\partial \ln K / \partial x} \qquad 4\text{-}1$$

defines the interaction parameter α, where x denotes the perturbation. The observation of constant α_x values over a broad range of ΔG^0 argues that the barrier is narrow and well defined along the reaction coordinate being probed (*6, 9*). Non-linear relationships between $\Delta G^{\neq}$ and ΔG^0 can also be observed. Nonlinearities can result from a variety of factors: (i) the movement of the transition state along broad barrier; (ii) a change in the rate-limiting step of a sequential pathway; (iii) changes in parallel pathways; (iv) a change of one or both of the ground states induced by the perturbation (*6*).

If x is the concentration of denaturant, the relationship is simply the well known Tanford β parameter (β_T or θ_m) which defines the position of the transition state ensemble (TSE) relative to a dimensionless order parameter that reports upon the relative compactness of the transition state (*10*). Other examples albeit less commonly employed ones, include temperature, pressure and pH. Temperature dependent studies define α_T which is equal to $\Delta S^{\neq}/\Delta S^0$ where $\Delta S^{\neq}$ is the activation entropy and ΔS^0 is the equilibrium change in entropy (*11-13*). Pressure dependent studies yield $\alpha_p = \Delta V^{\neq}/\Delta V^0$ and probe the change in volume of the system at the TSE, $\Delta V^{\neq}$, relative to the total volume change associated with folding, ΔV^0 (*14*). pH dependent investigations provide a global view of the development of interactions involving titratable groups (*15-17*). α_{pH} is formerly equal to $\Delta Q^{\neq}/\Delta Q^0$ where $\Delta Q^{\neq}$ is the difference in the number of protons bound to the TSE and denatured state ensemble (DSE) while ΔQ^0 is the difference between the number of protons bound to the folded state and the protons bound to the DSE. The concept can be extended to include mutations as a perturbation (*2, 3, 6, 17*). Within this context the commonly employed φ-values, $\varphi = \Delta\Delta G^{\neq}/\Delta\Delta G^0$, can be viewed as a linear REFER defined by two points, wildtype and mutant.

φ-values are widely used in protein folding to probe the development of interactions at the level of individual residues and can, if denatured state effects are not important, be used to define the fractional development of sidechain interactions (*1-3, 6, 17, 18*). The φ-value methodology has been applied to a large number of proteins and β_T has been measured for even more. There have been surprisingly few investigations of the development of electrostatic interactions during protein folding despite the fact that they can be important for stability and can play a key role in specifying a specific structure (*15, 16, 19-32*). A number of investigations have relied on the pH dependence of protein folding to probe the development of electrostatic interactions (*15, 16, 19, 20, 23*).Complimentary information might, in principle, be obtained from ionic strength dependent studies however these have been relatively rare (*33-37*). Furthermore, their analysis can be complicated because variation in ionic strength can affect more than charge charge interactions; specific binding may occur and Hofmeister effects might contribute (*26, 36-41*). There is another important but often unappreciated additional complication with the analysis of ionic strength dependent studies, namely that not all

charge-charge interactions will be equally screened by salt even if the interacting groups lie on the surface of the protein. Interactions involving well separated charges are screened more effectively than those involving charges which are in close proximity and there are even examples of charge charge interactions which are not screened by salt. This effect has been observed independently by several groups in both computational and experimental studies (*41-43*). In a qualitative sense this can arise if the groups in question lie close to each other and the ion cloud does not interpenetrate.

Application of the REFER approach to ionic strength dependent studies offers the advantage that it does not depend upon the assumption of a particular physical model, such as, for example, a Debye-Hükel type approach, or the validity of a particular theory of ionic solutions (*33*). The REFER analysis has not yet been applied to ionic strength dependent studies of protein folding. In the present work we analyze the ionic strength dependence of the folding of a small α-β protein, the N-terminal domain of the ribosomal protein L9 (NTL9) and show that ionic strength can be used to define a REFER. NTL9 has been the subject of extensive folding studies and the development of electrostatic interactions in the TSE have been characterized by pH dependent studies (*13, 19, 32, 43-45*). Thus NTL9 offers an ideal test case to asses if ionic strength dependent REFERs can be interpreted in terms of the development of just electrostatic interactions. The folding of NTL9 has been well studied by more traditional approaches and the folding of the core split β-α-β motif is two state over a wide range of conditions, making it a useful model system for biophysical studies.

4.2 Materials and Methods

4.2.1 Protein expression and purification

NTL9 was overexpressed in the BL21(DE3) strain of *E. Coli.* and purified by ion-exchange chromatography and reverse phase HPLC as previously described with the exception that TFA was avoided during the HPLC purification stage (*30*). TFA is difficult to remove and we wanted to avoid variations in ionic strength due to the associated TFA counter ion. HPLC purification made use of C18 preparative column and a A-B gradient of 15% B to 65% B in 50 minutes. Buffer A was 99.9% water with 0.1% TFA; buffer B was 80% acetonitrile and 19.9% water with 0.1% TFA. NTL9 eluted at 25 minute. The identity of the peptide was confirmed by MALDI-TOF mass spectrometry (expected: 6219.3; observed: 6218.9).

4.2.2 Protein stability measurements

Protein stability was determined by chemical denaturation monitored by circular dichroism at 222nm. Measurements were made using an Aviv Instrument model 202SF spectrometer. Denaturation experiments were performed in 20mM sodium acetate (NaAc) buffer with different salt concentrations at pH 5.5, 25^{o}C. The protein concentration was about 20 μM. Urea was chosen as the denaturant because it is non-ionic. The data from chemical denaturation experiments were fit as plots of ellipticity versus denaturant

concentration to the following equation:

$$\theta = \frac{(a_n + b_n[denaturant]) + (a_d + b_d[denaturant]e^{-(\Delta G^0[denaturant])/RT})}{1 + e^{-(\Delta G^0[denaturant])/RT}} \quad 4\text{-}2$$

Where θ is the ellipticity, T is the temperature, R is the gas constant, ΔG^0 is the free energy of unfolding, a_n and b_n are used to define the ellipticity of the native state; a_d and b_d are used to define the ellipticity of the denatured state. The free energy of unfolding is assumed to be a linear function of denaturant concentration (*46, 47*):

$$\Delta G^0 = \Delta G^0(H_2O) - m_{eq}[denaturant] \quad 4\text{-}3$$

The linear relationship has been shown to be valid for NTL9 (*44*). At the highest salt concentrations the curves did not have good post transition baselines, making the fits difficult. In these cases the stability was estimated by determining the concentration of urea at the midpoint, c_M, by numerically differentiating the curve. $\Delta G^0(H_2O)$ was calculated via

$$\Delta G^0(H_2O) = c_M * m_{avg} \quad 4\text{-}4$$

Where m_{avg} is the average m-value for data collected between 0 and 1.0M added salt. The ionic strength of the solutions was calculated using the standard expression:

$$I = \frac{1}{2}\sum_k m_k z_k^2 \quad 4\text{-}5$$

where m_k is the molar concentration and z_k is the valency of ionic species k.

4.2.3 Kinetic folding studies

Protein folding and unfolding rates were determined by fluorescence stopped-flow experiments using an Applied Photophysics Model SX.18MV instrument. The fluorescence signal of the single tyrosine at position 25 of NTL9 was followed. The excitation wavelength was 280nm and the fluorescence signal above 305nm was recorded using a cutoff filter. All experiments were done in a 20mM sodium acetate buffer at pH 5.5 with different salt concentrations at 25°C. Final protein concentrations were about 50 μM. For refolding studies, the peptide was dissolved in 9-10M urea and folding was initiated by mixing with 10 volumes of buffer with low concentration of denaturant. For unfolding studies, the peptide was dissolved in native buffer and unfolding was initiated by mixing with solutions of high denaturant concentration. Four or five fluorescence traces were typically averaged at each concentration of denaturant. The resulting trace was fit to a single exponential to determine the observed rate constant (k_{obs}). Plots of $\ln k_{obs}$ versus denaturant concentration, so-called chevron plots, were fit to the following equation to obtain the folding and unfolding rates in the absence of denaturant:

$$\ln(k_{obs}) = \ln(k_f^{H_2O} e^{-(m_f[denaturant])} + k_u^{H_2O} e^{(m_u[denaturant])}) \quad 4\text{-}6$$

Where $k_u^{H_2O}$ and $k_f^{H_2O}$ are the unfolding and folding rates in the absence of denaturant, m_f and m_u describe how $\ln(k_f)$ and $\ln(k_u)$ dependent on the concentration of denaturant. m_f is widely thought to report on the change in solvent accessible surface area between the DSE and TSE while m_u reports on the change in surface area between the native state and the TSE (*10, 47*). Only the folding branch was collected at higher concentrations of salt because the unfolding branch is so small that the fitted parameters k_u and m_u are not reliable.

4.3 Results

4.3.1 NTL9 as a model system

NTL9 is a basic protein with 11 Lys, 1 Arg and 6 acidic residues. The domain forms one of the simplest examples of the split β-α-β motif (Figure 4-1). The fold consists of a three stranded anti-parallel β-sheet with a short α-helix connecting strands two and three. The ordered loop connecting the first and second β-strand is rich in Lys. The second helix forms the long interdomain connector between the N and C terminal domains. The construct utilized here consists of residues 1 to 56 of the intact protein. The final few residues of the C-terminal helix are frayed in solution (*45*). The acidic residues in NTL9 are well distributed across the surface and only one, D23, appears to be involved in a well defined pairwise interaction, namely a salt bridge with the N-terminus (Figure 4-2) (*43*). Analysis of the native state pKa's shows that D8, E17 and D23 have pKa's below model compound values, indicating that they are involved in favorable electrostatic interactions in the native state (*27*). A surface representation coded by charge is displayed in figure 2.

4.3.2 Stability studies

The stability of NTL9 was measured by CD monitored urea denaturation as a function of salt concentration over a range of 0 to 1.5M. A total of 21 different measurements were made at 10 different values of the ionic strength. Urea was used since guanidine HCl is a salt and will change the ionic strength. All experiments were performed at pH 5.5 because previous folding studies were conducted at this pH. NTL9 is a small protein which leads to a relatively broad urea induced unfolding transition and this in turn causes problems with the post-transition baselines at the highest salt concentration. The stability was determined by directly fitting the curve to equation 4-2 for salt concentrations below 1.0 molar. The midpoint of the urea induced unfolding transition, c_m, can be estimated even with a poor post-transition baseline. This allows the stability at high salt concentration to be determined using equation 4-4 and the average value of m_{eq}. The equilibrium m-value, m_{eq}, is independent of ionic strength to within the experimental precision. The m_{eq}-value is related to the change in solvent accessible surface area between the native state and the DSE. The observation that the m_{eq}-value is independent of ionic strength argues that there is no significant change in the compactness of the DSE at high salt concentration. At first glance this may seem

surprising since the DSE of NTL9 is expected to have a net charge near +6 at pH 5.5 and thus reduction of electrostatic interactions via screening might be expected to lead to compaction. However there are also specific electrostatic interactions that favor a compact DSE and these too should be screened by high salt concentration (*27, 32*). Thus the two competing effects conspire to reduce the dependence of m_{eq} on ionic strength. The stability increased from 4.16 to 5.53 kcal/mol over the range of salt concentrations studied, indicating that ionic strength dependent interactions make a net unfavorable contribution to ΔG^0 under these conditions (table 4-1). Plots of the apparent fraction of unfolded vs. [urea] at different salt concentrations are shown in figure 4-3. The plots of ellipticity vs. [urea] together with the associated fits are shown in figure 4-4.

Salts can modulate protein stability by changes in ionic strength, by specific binding and via Hofmeister effects (*26, 28, 36-40*). In theory, although often not in practice, the functional form of the dependence of ΔG^0 upon the ionic strength, I, yields information about the mechanism of ionic strength effects. Specific binding leads to a linear dependence of ΔG^0 upon I when ligand concentration is above k_d. Evidence against specific binding is provided by studies which examined the effect of varying the anion and the cation. It is easy to vary the cation by using KCl. Variation of the anion is more problematic since the absorbance of Br^- or I^- causes problems and sulfate introduces complications because it non-specifically stabilizes proteins via preferential exclusion. We conducted additional measurements at a total ionic strength of 120mM using 20mM sodium acetate with 100mM KCl, 100mM NaBr, 100mM NaF and 100mM NaAc and 33mM Na_2SO_4. The measured stability was in excellent agreement with that determined using NaCl to adjust the ionic strength (table 4-1). Additional measurements were made at 770mM ionic strength using 750mM KCl and 250mM Na_2SO_4. The agreement with the NaCl data is good.

The ionic strength dependence of electrostatic screening is more complicated. Some workers have used a simple Debye-Hükel limiting law $\sqrt{I}$ dependence expected for the activity coefficient of an ion in very dilute salt concentration (*33*). It is not at all obvious why such a relationship should be expected to describe the variation of ΔG^0 with ionic strength over the range of salt concentrations used in protein stability studies. It is also worth bearing in mind that it can be difficult to distinguish between an $\sqrt{I}$ dependence and a linear dependence upon I such as that expected for classic Hofmeister effects (*39*). Electrostatic screening has also been modeled using a Columbic potential with a Debye screening term derived from model dependent solution of the linearized Poisson-Boltzman equation (*42*):

$$U_c(r_{ij})e^{-\kappa r_{ij}} \qquad 4\text{-}7$$

where U_c is the columbic potential for two charges separated by a distance r_{ij} in a medium of dielectric constant ε. κ is the Debye parameter which is related to the Debye screening

length. κ is proportional to $\sqrt{\frac{I}{\varepsilon T}}$ where I is the ionic strength and T is the absolute temperature. The relationship predicts that the screening of pairwise charge charge interactions should have a $Ae^{-\sqrt{I}}$ dependence on ionic strength ignoring the small contribution from salt effect on ε. This simple relationship has been successfully used to rationalize salt dependent pKa shifts even for salt concentrations well above the expected range of validity of the Debye-Hükel model. Note that equation 6 does not predict that ΔG^0 will decrease with increasing ionic strength since both favorable and unfavorable electrostatic interactions will be screened. If the electrostatic free energy can be written as the sum of a set of pairwise terms then the effect of increasing ionic strength should follow a $Ae^{-\sqrt{I}}$ dependence where A is independent of ionic strength. Figure 4-6 compares plots of ΔG^0 vs. I, $\sqrt{I}$ and $Ae^{-\sqrt{I}}$. In all three cases the data appears to be well fit. The plot of ΔG^0 vs. $\sqrt{I}$ appears linear with a r^2 value of 0.904. Plotting ΔG^0 vs. I also yields a linear plot with $r^2 = 0.908$. Plotting ΔG^0 vs. $Ae^{-\sqrt{I}}$ also yields a linear plot with $r^2 = 0.873$. Comparison of the plots highlights the practical difficulty in distinguishing between an I or other dependence of ΔG^0 on ionic strength. This is a general difficulty with ionic strength dependent studies and is not specific for NTL9. This is formally irrelevant to the REFER analysis.

4.3.3 The dependence of the folding rate upon ionic strength

The folding rate of NTL9 was determined for each of the samples used for the stability studies (table 4-1). The rate increased with ionic strength although the effect upon $\ln k_f$ was less than the effect upon $\frac{\Delta G^0}{-RT}$ (i.e. lnK). The effects of varying the composition of the salt were tested in exactly the same way as used for the stability studies. The effect on the rate constant varied with ionic strength and not with the choice of salt for the points tested. Plots of the dependence of the natural logarithm of the observed first order rate constant, $\ln k_{obs}$, vs. [urea], "chevron plots", collected at 0.1 and 1M [NaCl] are shown in figure 4-5. The plots have the typical V-shape expected for two state folding with no observable rollover at low denaturant concentration. The stability calculated from k_f and k_u is in good agreement with the values determined by equilibrium measurements; 4.12 kcal/mol from the kinetic measurements and 4.34 kcal/mol for the equilibrium measurements at 100mM added salt and 5.02 kcal/mol vs. 5.05 kcal/mol for the kinetic and equilibrium values at 1M NaCl. The m_{eq}-values measured in the

equilibrium studies are also in very good agreement with the values calculated from the kinetic parameters m_f and m_u.

A plot of $\ln k_f$ vs. $\sqrt{I}$ is shown in figure 4-7. The plot is well fit to a straight line with $r^2 = 0.900$. The data can also be fit to an $Ae^{-\sqrt{I}}$ dependence with $r^2 = 0.845$. A linear dependence on I also nicely fits the data with $r^2 = 0.918$, making it difficult to distinguish between classic Debye-Hükel behavior and an alternate salt dependence. Again this highlights the difficulties in analyzing ionic strength dependent data, but, once again, this is formally irrelevant for REFER analysis.

4.3.4 REFER analysis of the effects of ionic strength

The data presented in the proceeding two subsections highlights the difficulties which can arise in trying to deduce if salt dependent folding and or stability data follows the Debye-Hükel relationship. The REFER analysis is not affected by these considerations since it simply seeks to deduce how $\ln k_f$ varies with lnK. A Leffler plot of lnkf vs. lnK is shown in figure 4-8 (*6, 48, 49*). The parameter

$$\alpha_I = \left(\frac{\partial \Delta G^{\neq}}{\partial I}\right) / \left(\frac{\partial \Delta G^0}{\partial I}\right) \qquad 4\text{-}8$$

defines the ionic strength α-value. There is an excellent linear relationship with slope, $\alpha_I = 0.45 \pm 0.030$, $r^2 = 0.918$. The modest value of α_I indicates that salt dependent interactions are only partially developed in the transition state for folding. Interestingly α_I is noticeable smaller than β_T (0.65) and $\Delta C_p^{\neq}/\Delta C_p$ (0.60), parameters which report on the burial of surface area (*13*). However, the slope of the Leffler plot, α_I is larger than α_{pH} =0.33, determined from pH dependent studies (*19*).

There is, within the precision of the data, no hint of deviation from linearity in a plot of α_I vs. ΔG_I^0. A linear plot is expected for narrow transition state barriers, however it can be difficult to detect deviations from linearity (*6*). Cross-interaction parameters can be more sensitive probes of transition state movement (*6, 9*). The cross-interaction parameter α_{xy} is defined by

$$\frac{\partial \alpha_x}{\partial \Delta G_y^0} = \frac{\partial^2 \Delta G^{\neq}}{\partial \Delta G_x^0 \partial \Delta G_y^0} \qquad 4\text{-}9$$

where the subscripts x and y refer to different perturbations (*6*). The data presented here can be used to calculate the cross-interaction parameter relating ionic strength and denaturant. Plots of α_I vs. ΔG_{urea}^0 and of β_T vs. ΔG_I^0 are shown in figure 4-9. The subscript urea indicates that the free energy is varied by changing the concentration of denaturant and the subscript I indicates that ionic strength is used to change ΔG^0. The

slope of the α_I vs. ΔG^0_{urea} plot is the cross-interaction parameter $\frac{\partial^2 \Delta G^{\neq}}{\partial \Delta G^0_{urea} \partial \Delta G^0_I}$ while the slope of the plot of β_T vs. ΔG^0_I is $\frac{\partial^2 \Delta G^{\neq}}{\partial \Delta G^0_I \partial \Delta G^0_{urea}}$. Both plots are linear with positive slope although the scattering in the β_T vs. ΔG^0_I plot coupled with the, necessarily, limited range of the data leads to a weak correlation (r^2 = 0.404). There is a strong correlation between α_I and ΔG^0_{urea} (r^2 = 0.9997). The slopes of the two plots are essentially equal as expected for cross-interaction parameters and the magnitude of the slope is comparable to cross-interaction parameters determined in other systems (*6*). The positive cross-interaction parameters are consistent with Hammond effect i.e. movement of the transition state towards the native state upon destabilization.

4.4 Discussion

The REFER analysis described here indicates that salt dependent interactions are moderately developed in the transition state compared to the burial of hydrophobic surface area. This seems reasonable since the charged residues in NTL9 all lie on the surface while the hydrophobic groups are buried in the structure. Previous work involving modification of the hydrophobic core residues of NTL9 showed that there was an excellent correlation between the change in hydrophobicity caused by a modification and the change in the logarithm of the folding rate while there was a much weaker relationship between the variation in the size and shape of the residues. This argues that the transition state involves the sequestering of hydrophobic groups with tight specific packing occurs on the downhill side of the barrier (*50*). Amide H/D isotope effect studies also indicate significant H-bonded structure in the transition state (*51*). Taken together these studies suggest that the TSE of NTL9 contains a fairly well structured backbone stabilized by a more loosely packed hydrophobic core. In this model most of the charged residues are expected to be fluxional, solvated, and not involved in specific native interactions. In other words, in terms of solvation and conformational flexibility, the charged residues are in a more DSE like environment rather than a native like environment. There is at least one fairly specific interaction involving charged residues which appears to be formed in the TSE. Mutational analysis of a set of Lys to Met and acidic to amide mutants showed that K12 and D8 are involved in non-native interactions in the DSE and indicated that these interactions are still present, although slightly weaker, in the TSE (*31, 32*). The presence of non-native interactions in the TSE or DSE can not be inferred directly from ionic strength dependent studies, however this does not affect the REFER analysis. It can, however, affects the interpretation of the origins of any REFER. The persistence of this interaction contributes to the "DSE-like" character of the

salt sensitive interactions in the transition state.

A key question of interest is whether or not the salt dependence of the stability and activation free energy of NTL9 is due to screening of electrostatic interactions. It is clearly not easy to distinguish between a dependence on I vs. a dependence of $\sqrt{I}$ or $Ae^{-\sqrt{I}}$ but never the less we believe that part of the salt effects arise from screening especially at low salt. First the control experiments are consistent with a dependence on total ionic strength and not the composition of the ion. This argues against specific ion binding as the origin of the observed effects. Secondly, the primary ions used, Na^+, K^+, Cl^- and Br^- all fall in the middle of the Hofmeister series and are relatively nonchaotropic and nonkosmotropic. This Hofmeister effects are expected to be modest at least for the lowest salt concentrations. They are expected to play a more important role as the salt concentration increases. On the other hand pH dependent studies yield an α_{pH} value of 0.33 which is clearly smaller than α_I. pH dependent studies report on the development of electrostatic interactions and do not suffer from the same problems in interpretation that ionic strength dependent studies do. The larger value of α_I related to α_{pH} could arise from a contribution of non-electrostatic interactions to the REFER, i.e. Hofmeister effects. These would be expected to increase α_I since, as described above, hydrophobic interactions play a key role in the folding of NTL9. Alternatively α_I could be dominated by electrostatic screening but could still be different from α_{pH} because some interactions in the TSE or native state may not be screened by salt. Along these lines it is interesting to note that the one specific pairwise charge-charge interaction in the native state, the salt bridge between the N-terminus and D23, has been shown to be insensitive to screening by salt (*43*) Clearly the common practice of assuming that ionic strength effects are due only to the screening of electrostatic interactions and that salt screens all electrostatic interactions equally is not correct..

The analysis of the cross-interaction parameters is consistent with a salt induced Hammond effect. Finally we note that the study described here illustrates the power of the REFER approach even when it is impossible to a prior determine if $\ln k_f$ or lnK follows Debye-Hükel like relationship. In addition, the REFER analysis allows potential transition state movement to be probed and the variation of ionic strength provides a convenient method for measuring self and cross interaction parameters.

Table 4-1. Summary of the kinetic and thermodynamic data. Studies were conducted in 20mM NaAc, pH 5.5, 25°C.

Added Salt[1]	ΔG^{0} [2] (kcal/mol)	m_{eq} (kcal/mol M^{-1})	k_f (s^{-1})	m_f (kcal/mol M^{-1})	β_T
No salt	4.16±0.040	0.65±0.010	820±30.0 760±20.0 750±30.0	0.44±0.010 0.45±0.010 0.41±0.010	0.67±0.010 0.69±0.010 0.63±0.010
0.025M NaCl	4.21±0.120	0.65±0.020	820±40.0	0.44±0.010	0.67±0.010
0.05M NaCl	4.15±0.070	0.63±0.010	910±20.0 860±30.0	0.45±0.010 0.44±0.010	0.71±0.010 0.69±0.010
0.1M NaCl	4.34±0.080	0.66±0.010	950±30.0 990±40.0	0.46±0.010 0.47±0.010	0.69±0.010 0.71±0.010
0.15M NaCl	4.20±0.070	0.63±0.010	840±30.0	0.42±0.020	0.66±0.020
0.25M NaCl	4.62±0.040	0.67±0.010	900±30.0	0.40±0.010	0.59±0.010
0.5M NaCl	4.70±0.040	0.66±0.010	1190±70.0	0.43±0.010	0.65±0.010
0.75M NaCl	5.04±0.090	0.68±0.010	1450±140.0	0.42±0.010	0.61±0.010
1M NaCl	5.05±0.050	0.66±0.010	1880±170.0	0.43±0.010	0.65±0.010
1.5M NaCl	5.53±0.060	0.66±0.010	2060±110.0 2580±120.0	0.40±0.010 0.42±0.010	0.60±0.010 0.63±0.010
0.1M NaAc	4.50±0.320	0.66±0.060	1060±30.0	0.46±0.010	0.69±0.040
0.1M NaF	4.47±0.200	0.67±0.030	1040±30.0	0.46±0.010	0.68±0.010
0.1M KCl	4.41±0.200	0.68±0.030	1030±30.0	0.47±0.010	0.69±0.010
0.75M KCl	4.99±0.050	0.72±0.080	1370±140.0	0.48±0.020	0.66±0.040
0.033M Na_2SO_4	4.45±0.120	0.65±0.030	860±40.0	0.46±0.010	0.70±0.020
0.25M Na_2SO_4	4.75±0.04	0.71±0.060	1240±90.0	0.43±0.010	0.60±0.040

(1) All samples contained 20mM NaAc in addition to the added salt. Thus the ionic strength is 20mM larger than the values listed. The numbers after the ± symbol represent the standard error to the fit.

(2) Determined from equilibrium unfolding as described in the Materials and Methods.

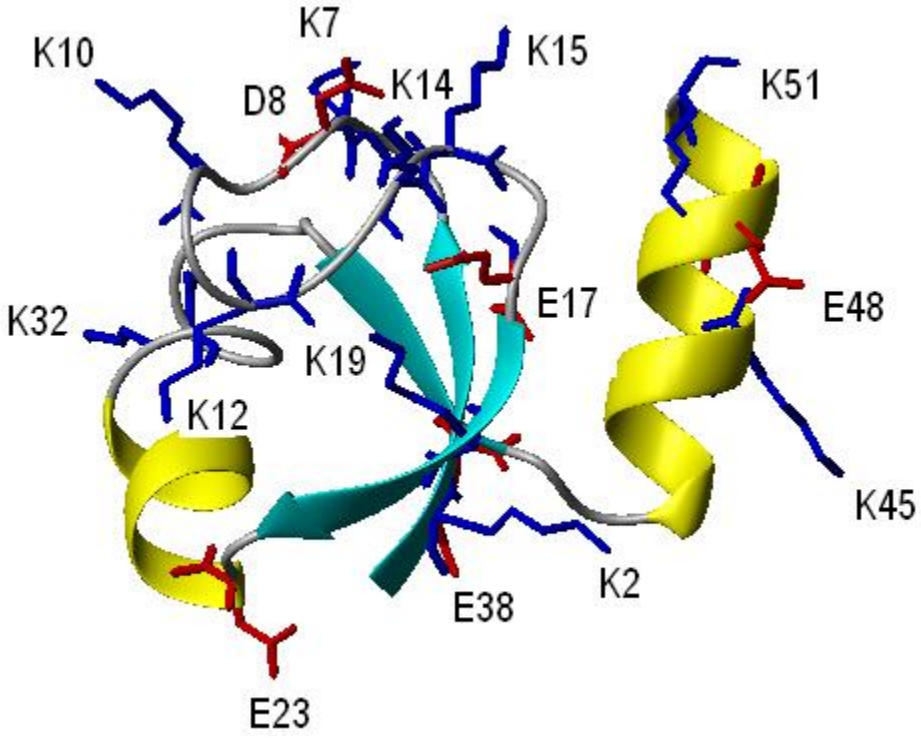


Figure 4-1. A ribbon diagram of NTL9. The acidic and basic residues are shown in stick format. The diagram was constructed using the pdb file 2HBB and the program MolMol. The last few residues of NTL9 are disordered in the X-ray structure. Thus the position of Glu54 and the C-terminus are not shown.

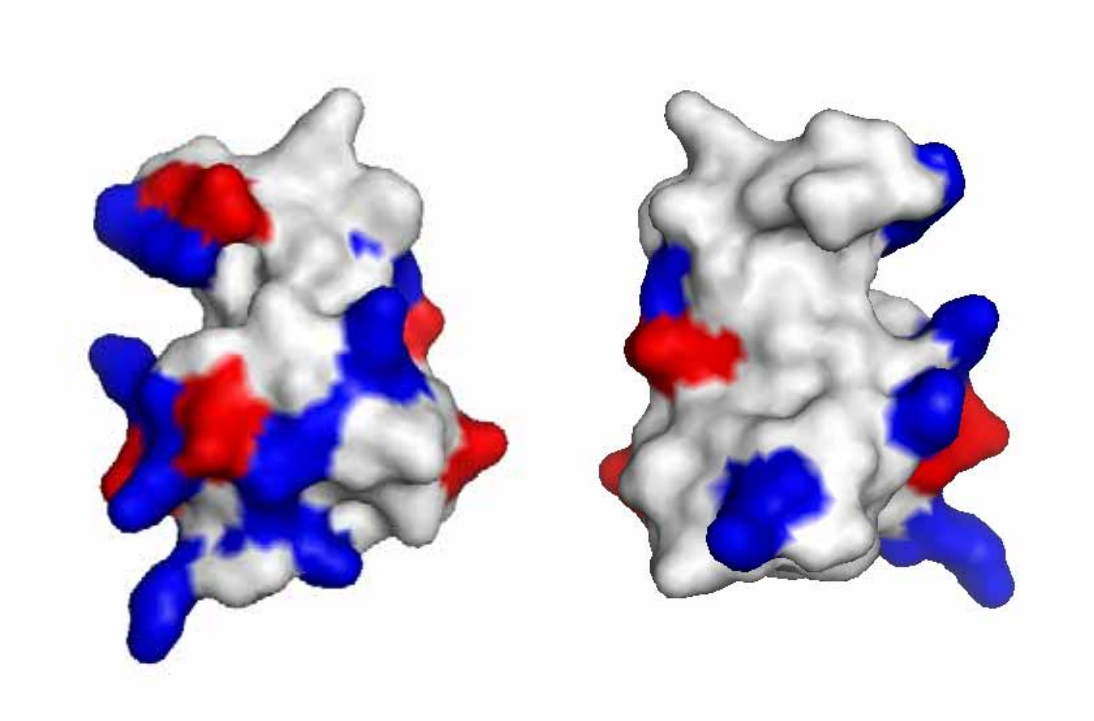


Figure 4-2. Color coded surface representation of NTL9 showing the distribution of positively charged (blue) and negatively charged (red) residues. Panels A and B differ by 180^{o} rotation around the z-axis.

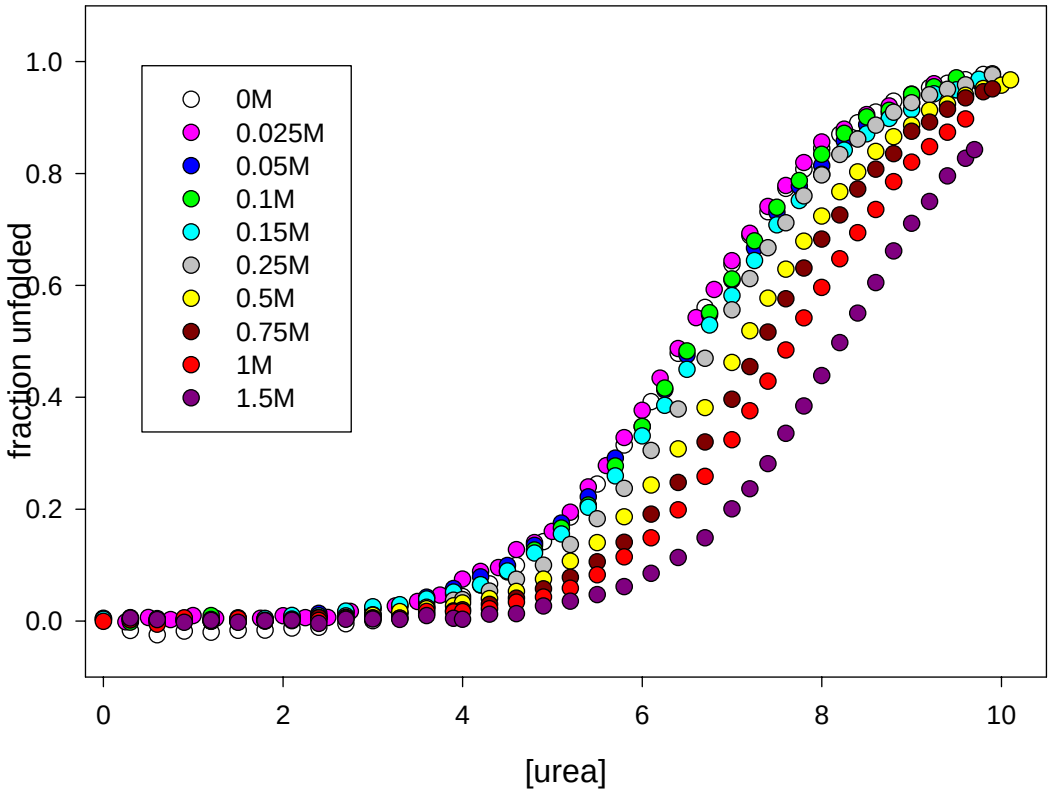


Figure 4-3. Plots of the fraction of unfolded vs. [urea] at various concentrations of NaCl. The ionic strength increases from left to right. Measurements were made in 20mM sodium acetate, pH 5.5, 25^{o}C. The concentration of added NaCl were 0M, 0.025M, 0.05M, 0.1M, 0.15M, 0.25M, 0.5M, 0.75M, 1M, 1.5M from left to right.

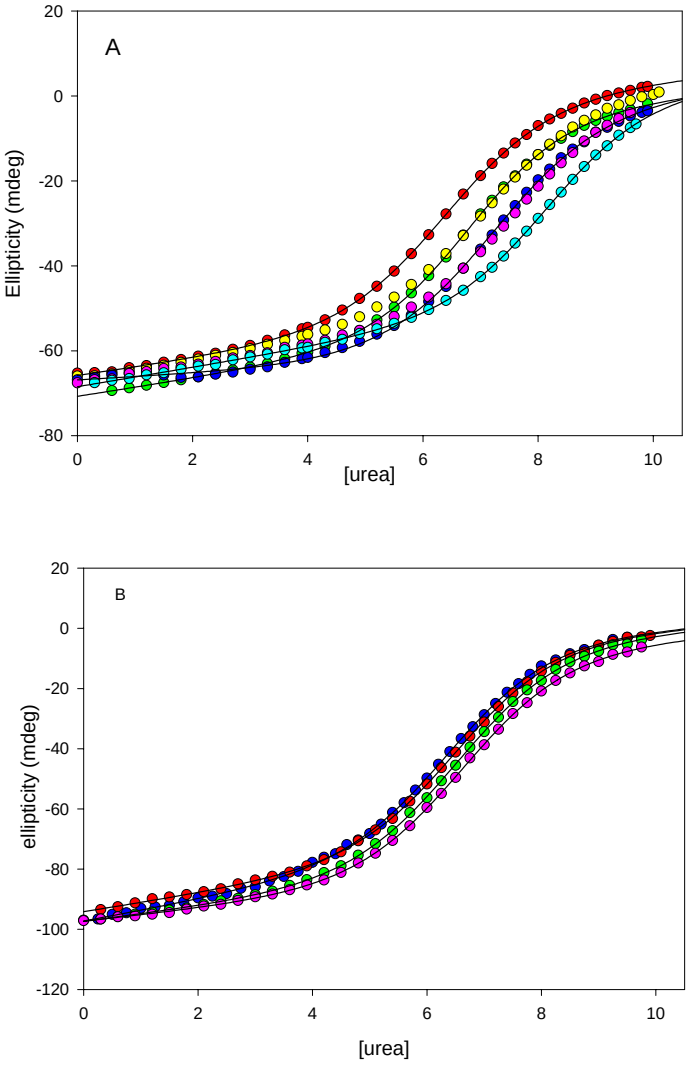


Figure 4-4. Plots of the measured elipticity vs. [urea] showing the fits. (A) from left to right 0M, 0.25M, 0.5M, 0.75M, 1M and 1.5M added NaCl; (B) from left to right 0.025M, 0.05M, 0.1M and 0.15M added NaCl. All experiments were conducted at pH 5.5, 25^{o}C in 20mM NaAc. Data was analyzed as described in the Materials and Methods. Direct curve fitting was used to estimate ΔG^{0} for data at and below 1M salt.

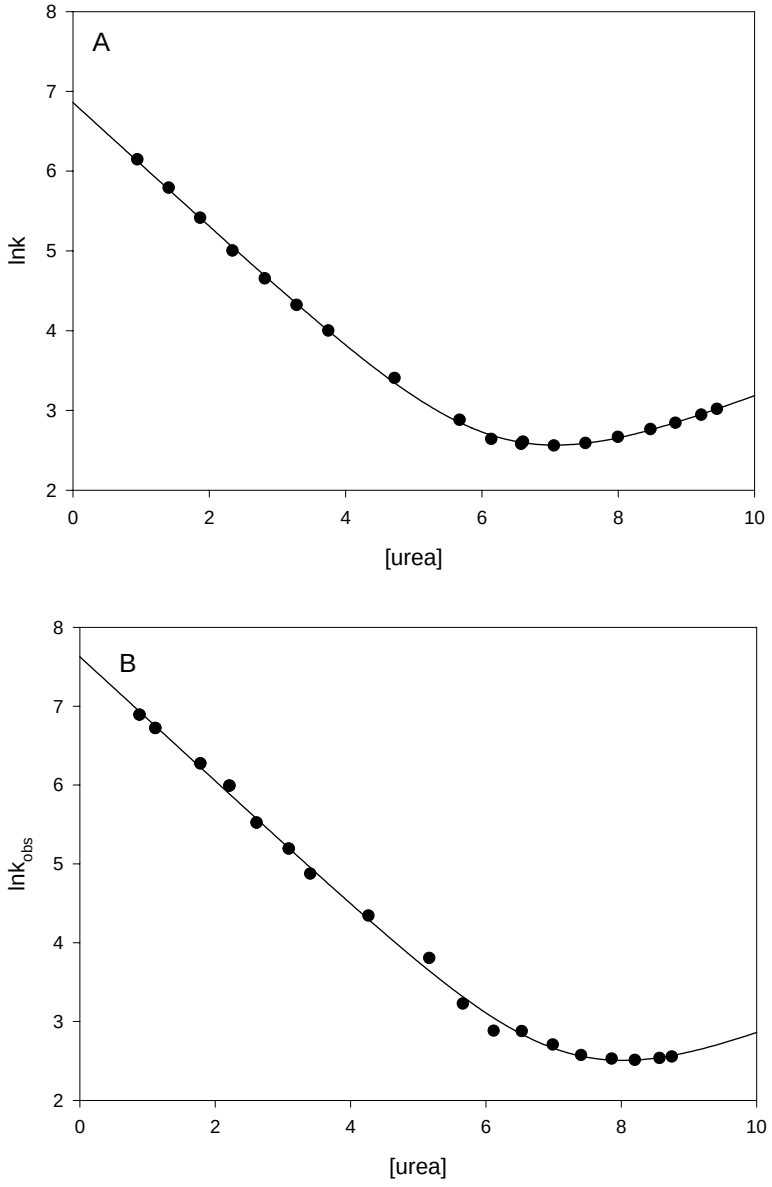


Figure 4-5. Plots of lnk_{obs} vs. [urea], "chevron plots", for NTL9. (A) 100mM NaCl; (B) 1M NaCl. All experiments were conducted at pH 5.5, 25°C in 20mM sodium acetate. Thus the ionic strength for (A) was 0.12M and 1.02M for (B).

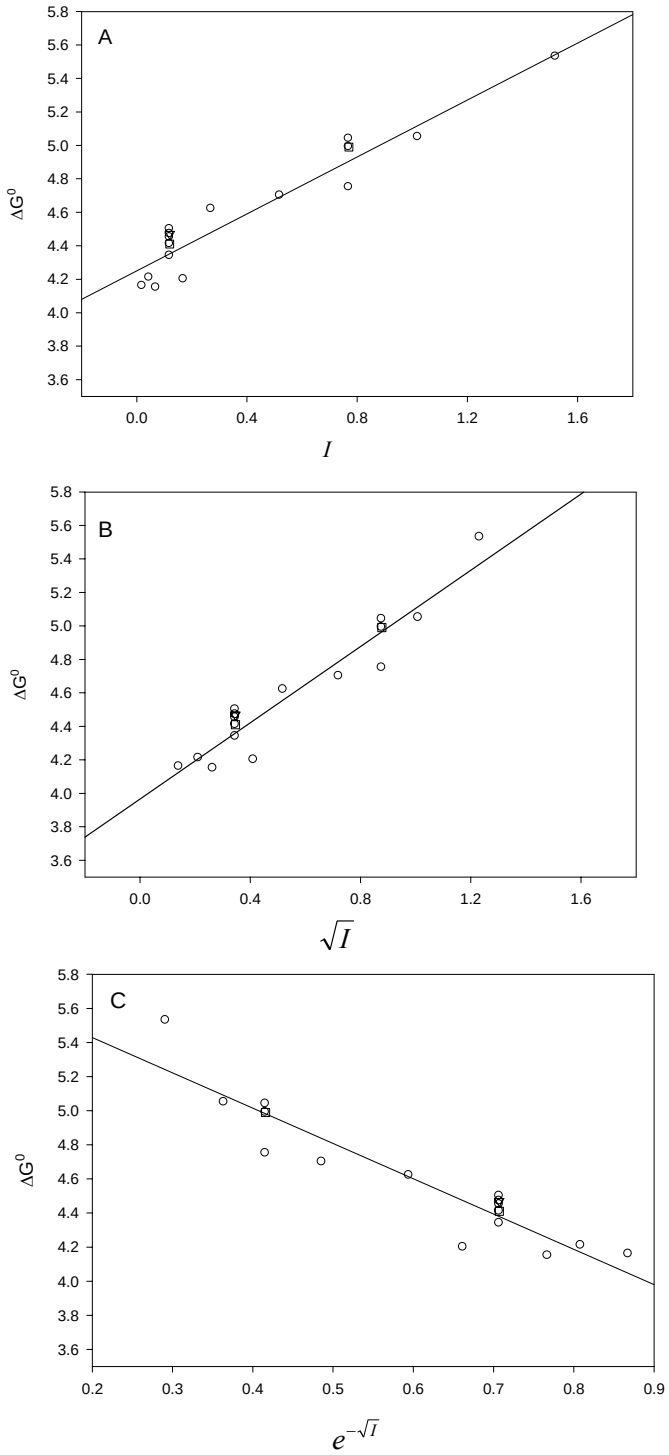


Figure 4-6. Plots of ΔG^0 vs. (A) I, (B) $\sqrt{I}$, (C) $Ae^{-\sqrt{I}}$. Data were collected at pH 5.5 in 20mM sodium acetate, 25°C. NaCl dependent data is indicated as (○), KCl dependent data is indicated as (□) and NaF dependent data as (▽).

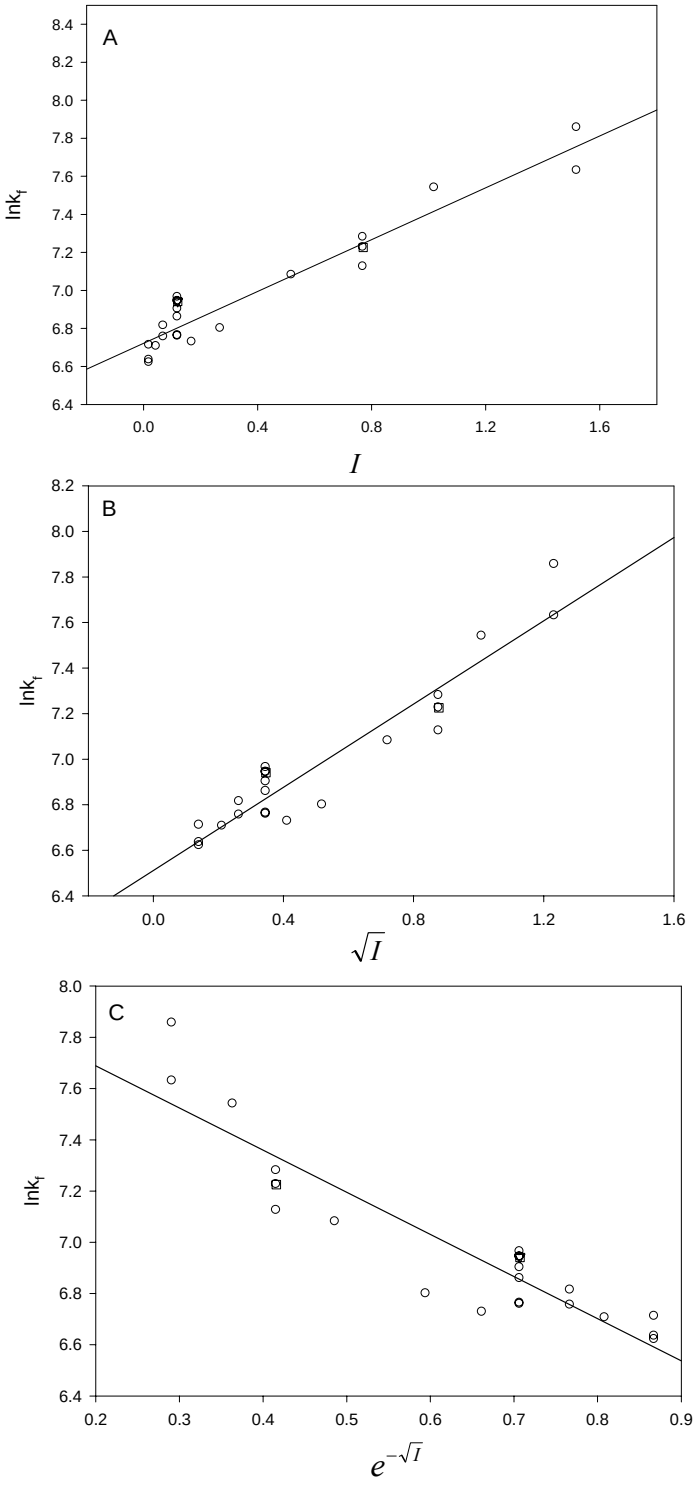


Figure 4-7. Plots of $\ln k_f$ vs. (A) I, (B) $\sqrt{I}$, (C) $Ae^{-\sqrt{I}}$. Data were collected at pH 5.5 in 20mM sodium acetate, 25°C. NaCl dependent data is indicated as (○), KCl dependent data is indicated as (□) and NaF dependent data as (▽).

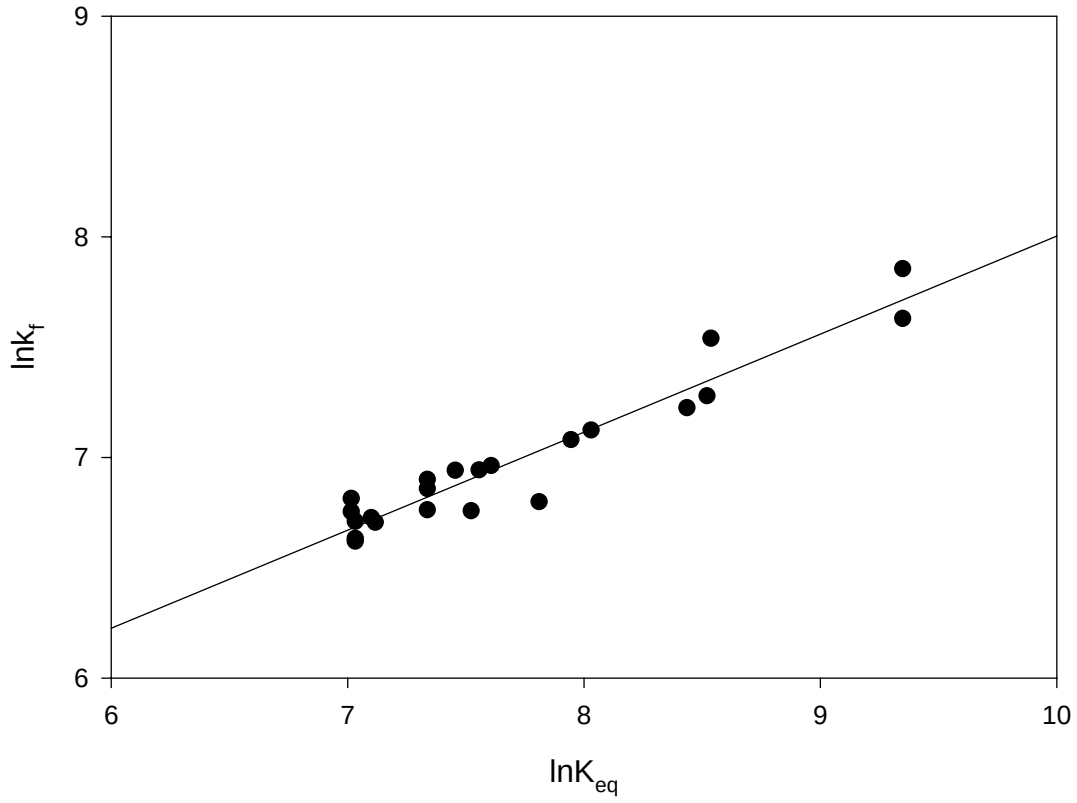


Figure 4-8. A Leffler plot of lnk_f vs. lnK for the ionic strength dependent data. The slope is 0.45 with r^2 =0.918.

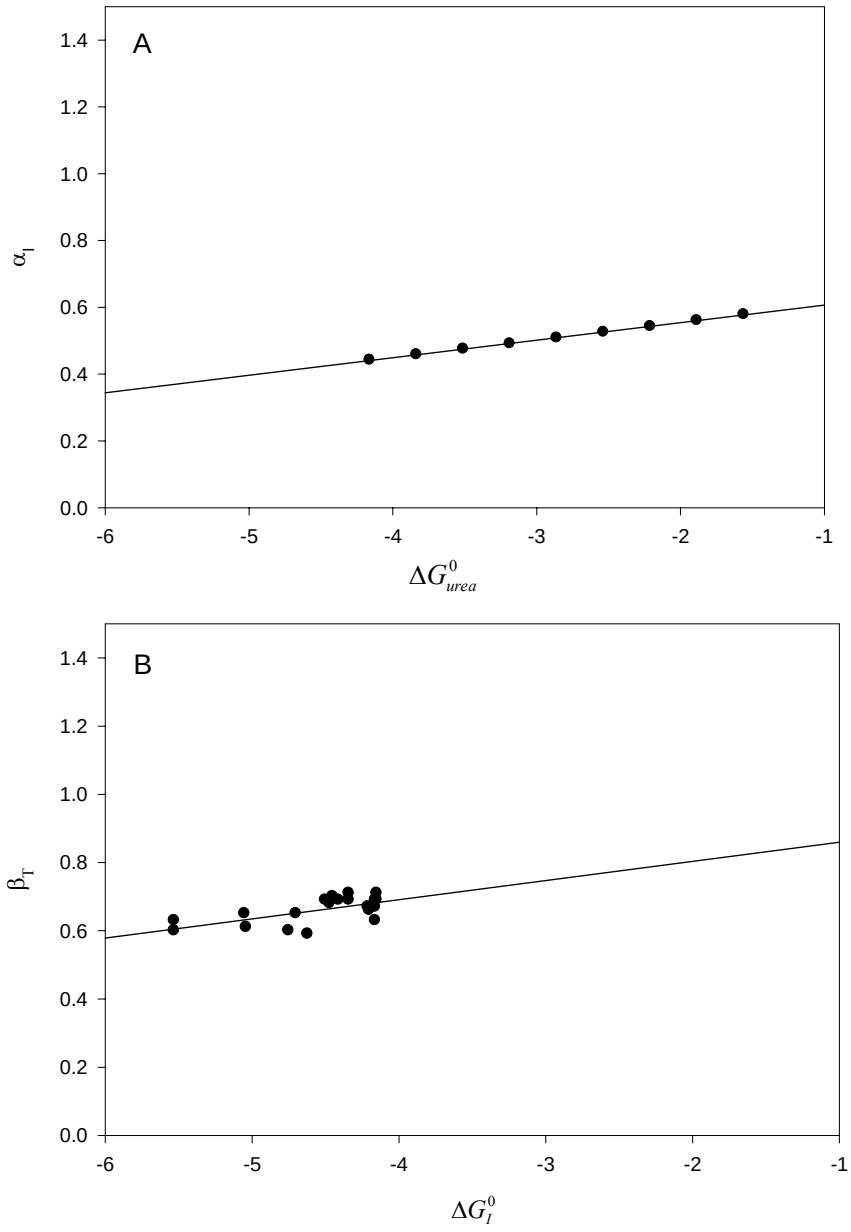


Figure 4-9. Analysis of cross-interaction parameters (A) a plot of α_I vs. ΔG^0_{urea}, slope = 0.052 mol kcal^{-1}, r^2 = 0.9997 (B) a plot of β_T vs. ΔG^0_I, slope = 0.056 mol kcal^{-1}, r^2 =0.404.

4.5 References

1. Matthews, C. R. (1987) Effect of point mutations on the folding of globular proteins, *Methods Enzymol. 154*, 498-511.
2. Goldenberg, D. P., Frieden, R. W., Haack, J. A., and Morrison, T. B. (1989) Mutational analysis of a protein-folding pathway, *Nature 338*, 127-132.
3. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding, *J. Mol. Biol. 224*, 771-782.
4. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding, *J. Mol. Biol. 254*, 260-288.
5. Jackson, S. E. (1998) How do small single-domain proteins fold?, *Fold Des. 3*, R81-91.
6. Sanchez, I. E., and Kiefhaber, T. (2003) Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states, *J. Mol. Biol. 327*, 867-884.
7. Matouschek, A., and Fersht, A. R. (1993) Application of physical organic chemistry to engineered mutants of proteins: Hammond postulate behavior in the transition state of protein folding, *Proc. Natl. Acad. Sci. U.S.A. 90*, 7814-7818.
8. Schatzle, M., and Kiefhaber, T. (2006) Shape of the free energy barriers for protein folding probed by multiple perturbation analysis, *J. Mol. Biol. 357*, 655-664.
9. Sanchez, I. E., and Kiefhaber, T. (2003) Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding, *Biophys. Chem. 100*, 397-407.
10. Tanford, C. (1970) Protein denaturation. Part C. Theoretical models for the mechanism of denaturation, *Advan. Protein Chem. 24*, 1-95.
11. Pohl, F. M. (1976) Temperature-dependence of the kinetics of folding of chymotrypsinogen A, *Febbs. lett. 65*, 293-296.
12. Scalley, M. L., and Baker, D. (1997) Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability, *Proc. Natl. Acad. Sci. U.S.A. 94*, 10636-10640.
13. Kuhlman, B., Luisi, D. L., Evans, P. A., and Raleigh, D. P. (1998) Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the protein L9, *J. Mol. Biol. 284*, 1661-1670.
14. Pappenberger, G., Saudan, C., Becker, M., Merbach, A. E., and Kiefhaber, T. (2000) Denaturant-induced movement of the transition state of protein folding revealed by high-pressure stopped-flow measurements, *Proc. Natl. Acad. Sci. U.S.A. 97*, 17-22.

15. Tan, Y. J., Oliveberg, M., and Fersht, A. R. (1996) Titration properties and thermodynamics of the transition state for folding: comparison of two-state and multi-state folding pathways, *J. Mol. Biol. 264*, 377-389.
16. Oliveberg, M., and Fersht, A. R. (1996) Formation of electrostatic interactions on the protein-folding pathway, *Biochemistry 35*, 2726-2737.
17. Raleigh, D. P., and Plaxco, K. W. (2005) The protein folding transition state: what are Phi-values really telling us?, *Protein pept. lett. 12*, 117-122.
18. Cho, J. H., and Raleigh, D. P. (2006) Denatured state effects and the origin of nonclassical phi values in protein folding, *J. Am. Chem. Soc. 128*, 16492-16493.
19. Luisi, D. L., and Raleigh, D. P. (2000) pH-dependent interactions and the stability and folding kinetics of the N-terminal domain of L9. Electrostatic interactions are only weakly formed in the transition state for folding, *J. Mol. Biol. 299*, 1091-1100.
20. Cavagnero, S., Debe, D. A., Zhou, Z. H., Adams, M. W., and Chan, S. I. (1998) Kinetic role of electrostatic interactions in the unfolding of hyperthermophilic and mesophilic rubredoxins, *Biochemistry 37*, 3369-3376.
21. Hendsch, Z. S., and Tidor, B. (1994) Do salt bridges stabilize proteins? A continuum electrostatic analysis, *Protein Sci. 3*, 211-226.
22. Lopez-Arenas, L., Solis-Mendiola, S., and Hernandez-Arana, A. (1999) Estimating the degree of expansion in the transition state for protein unfolding: analysis of the pH dependence of the rate constant for caricain denaturation, *Biochemistry 38*, 15936-15943.
23. Tissot, A. C., Vuilleumier, S., and Fersht, A. R. (1996) Importance of two buried salt bridges in the stability and folding pathway of barnase, *Biochemistry 35*, 6786-6794.
24. Yang, A. S., and Honig, B. (1993) On the pH dependence of protein stability, *J. Mol. Biol. 231*, 459-474.
25. Spencer, D. S., Xu, K., Logan, T. M., and Zhou, H. X. (2005) Effects of pH, salt, and macromolecular crowding on the stability of FK506-binding protein: an integrated experimental and theoretical study, *J. Mol. Biol. 351*, 219-232.
26. Goto, Y., Calciano, L. J., and Fink, A. L. (1990) Acid-induced folding of proteins, *Proc. Natl. Acad. Sci. U.S.A. 87*, 573-577.
27. Kuhlman, B., Luisi, D. L., Young, P., and Raleigh, D. P. (1999) pKa values and the pH dependent stability of the N-terminal domain of L9 as probes of electrostatic interactions in the denatured state. Differentiation between local and nonlocal interactions, *Biochemistry 38*, 4896-4903.
28. Meeker, A. K., Garcia-Moreno, B., and Shortle, D. (1996) Contributions of the ionizable amino acids to the stability of staphylococcal nuclease, *Biochemistry 35*, 6443-6449.
29. Oliveberg, M., Arcus, V. L., and Fersht, A. R. (1995) pKA values of carboxyl groups in the native and denatured states of barnase: the pKA values of the

denatured state are on average 0.4 units lower than those of model compounds, *Biochemistry 34*, 9424-9433.
30. Pace, C. N., Alston, R. W., and Shaw, K. L. (2000) Charge-charge interactions influence the denatured state ensemble and contribute to protein stability, *Protein Sci. 9*, 1395-1398.
31. Cho, J. H., Sato, S., and Raleigh, D. P. (2004) Thermodynamics and kinetics of non-native interactions in protein folding: a single point mutant significantly stabilizes the N-terminal domain of L9 by modulating non-native interactions in the denatured state, *J. Mol. Biol. 338*, 827-837.
32. Cho, J. H., and Raleigh, D. P. (2005) Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins, *J. Mol. Biol. 353*, 174-185.
33. de Los Rios, M. A., and Plaxco, K. W. (2005) Apparent Debye-Huckel electrostatic effects in the folding of a simple, single domain protein, *Biochemistry 44*, 1243-1250.
34. Otzen, D. E., and Oliveberg, M. (1999) Salt-induced detour through compact regions of the protein folding landscape, *Proc. Natl. Acad. Sci. U. S. A. 96*, 11746-11751.
35. Chen, Y. R., and Clark, A. C. (2003) Equilibrium and kinetic folding of an alpha-helical Greek key protein domain: caspase recruitment domain (CARD) of RICK, *Biochemistry 42*, 6310-6320.
36. Broering, J. M., and Bommarius, A. S. (2005) Evaluation of Hofmeister effects on the kinetic stability of proteins, *J. Phys. Chem. B. 109*, 20612-20619.
37. Dominy, B. N., Perl, D., Schmid, F. X., and Brooks, C. L., 3rd. (2002) The effects of ionic strength on protein stability: the cold shock protein family, *J. Mol. Biol. 319*, 541-554.
38. Scholtz, J. M. Y., E. J.; Stewart, J. M.; Baldwin, R. L. (1991) A Neutral, Water-Soluble, α-Helical Peptide: The Effect of ionic Strength on the Helix-oil Equilibrium, *J. Am. Chem. Soc. 113*, 5102-5104.
39. Baldwin, R. L. (1996) How Hofmeister ion interactions affect protein stability, *Biophys. J. 71*, 2056-2063.
40. Zhou, H. X., and Dong, F. (2003) Electrostatic contributions to the stability of a thermophilic cold shock protein, *Biophys. J. 84*, 2216-2222.
41. Kao, Y. H., Fitch, C. A., Bhattacharya, S., Sarkisian, C. J., Lecomte, J. T., and Garcia-Moreno, E. B. (2000) Salt effects on ionization equilibria of histidines in myoglobin, *Biophys. J. 79*, 1637-1654.
42. Lee, K. K., Fitch, C. A., and Garcia-Moreno, E. B. (2002) Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein, *Protein Sci. 11*, 1004-1016.
43. Luisi, D. L., Snow, C. D., Lin, J. J., Hendsch, Z. S., Tidor, B., and Raleigh, D. P. (2003) Surface salt bridges, double-mutant cycles, and protein stability: an

experimental and computational analysis of the interaction of the Asp 23 side chain with the N-terminus of the N-terminal domain of the ribosomal protein l9, *Biochemistry 42*, 7050-7060.

44. Kuhlman, B., and Raleigh, D. P. (1998) Global analysis of the thermal and chemical denaturation of the N-terminal domain of the ribosomal protein L9 in H_2O and D_2O. Determination of the thermodynamic parameters, ΔH^0, ΔS^0, and ΔCp^0 and evaluation of solvent isotope effects, *Protein Sci. 7*, 2405-2412.
45. Kuhlman, B., Boice, J. A., Fairman, R., and Raleigh, D. P. (1998) Structure and stability of the N-terminal domain of the ribosomal protein L9: evidence for rapid two-state folding, *Biochemistry 37*, 1025-1032.
46. Pace, C. N. S., B. A. & Thomson, J. A. (1989) Measuring the conformational stability of a protein, *In Protein structure: A Practical Approach (Creighton, T. E., ed), IRL press, Oxford*
47. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding, *Protein Sci. 4*, 2138-2148.
48. Leffler, J. E. (1953) Parameters for the description of transition states, *Science 117*, 340-341.
49. Jencks, W. P. (1985) A primer for the Bema Hapothle. An empirical approach to the characterization of changing transition-state structure, *Chem. Rev. 85*, 511-527.
50. Anil, B., Sato, S., Cho, J. H., and Raleigh, D. P. (2005) Fine structure analysis of a protein folding transition state; distinguishing between hydrophobic stabilization and specific packing, *J. Mol. Biol. 354*, 693-705.
51. Sato, S., and Raleigh, D. P. (2007) Kinetic isotope effects reveal the presence of significant secondary structure in the transition state for the folding of the N-terminal domain of L9, *J. Mol. Biol. 370*, 349-355.

5. Synthesis of ^{13}C, ^{18}O labeled HP36 for Isotope Edited Infrared Experiments

5.1 Introduction

Probing protein folding dynamics at the level of individual residues can provide very useful information to solve the protein folding problem. Recently IR, especially two-dimensional IR has been developed to probe protein structure and dynamics (*1, 2*). The amide I band in FTIR absorbance spectrum is sensitive to the secondary structure and can be used to follow the protein structure change upon perturbation of environment (*3*). However, the peak is broad since the band is the sum of individual C=O stretchs from each residue of the protein and it does not provide any site-specific information. Fortunately, the carbonyl carbon and carbonyl oxygen can be isotopic labeled by ^{13}C and ^{18}O and the $^{13}C=^{18}O$ stretch will absorb at different frequency than the $^{12}C=^{16}O$ stretch. Assuming a harmonic oscillation, the vibrational frequency can be calculated by equation 5-1:

$$\nu = \frac{1}{2}\pi c\sqrt{\frac{k}{\mu}} \qquad 5\text{-}1$$

Where ν is the frequency, c is the velocity of light, k is the force constant and μ is reduced mass which equals $\frac{m_1 m_2}{m_1 + m_2}$. So since the $^{13}C=^{18}O$ stretch has higher reduced mass, the frequency will shift to a lower value.

If only carbonyl carbon is isotopic labeled by ^{13}C, the frequency of $^{13}C=^{16}O$ band will shift about 40cm^{-1} lower than $^{12}C=^{16}O$ band which could still be partly overlapped with it. Another problem with the single ^{13}C labeled method is due to the 1.1% natural abundance of ^{13}C (*4*). Thus in large proteins the isotopic labeled $^{13}C=^{16}O$ band can not be distinguished from natural abundance ^{13}C bands and the interference can also be a complicating factor for midsized and small proteins. Since ^{18}O has only 0.002% natural abundance and ^{13}C, ^{18}O double substitution can shift the C=O stretch frequency to about 75 cm^{-1} lower, the isotopic double labeled method will be very useful to characterize protein dynamics at specific sites.

The villin headpiece subdomain (HP36) was chosen to incorporate a single $^{13}C=^{18}O$ label into the peptide backbone because it is a small helical protein which undergoes two state folding (*5*). In HP36, the $^{12}C=^{16}O$ stretch absorbs at 1645cm^{-1} while $^{13}C=^{18}O$ stretch shows absorption at 1572cm^{-1} which is well resolved. Thus by monitoring the $^{13}C=^{18}O$ band we can get the dynamic and structural information of the isotopic labeled specific site.

5.2 Results

Three residues were chosen to be isotopic labeled in different secondary fragments of HP36 to compare the dynamics from different location. They are Ala 49, Ala 57 and Leu-69, which are located in three different helices. Figure 5-1 shows the structure of HP36 with the sites of the residues we double labeled. Ala 49 is located on the protein surface while Ala 57 is partially buried and Leu 69 is buried in the hydrophobic core. Table 5-1 gives the solvent accessible surface area of the carbonyl carbon, carbonyl oxygen, backbone and sidechain of these residues.

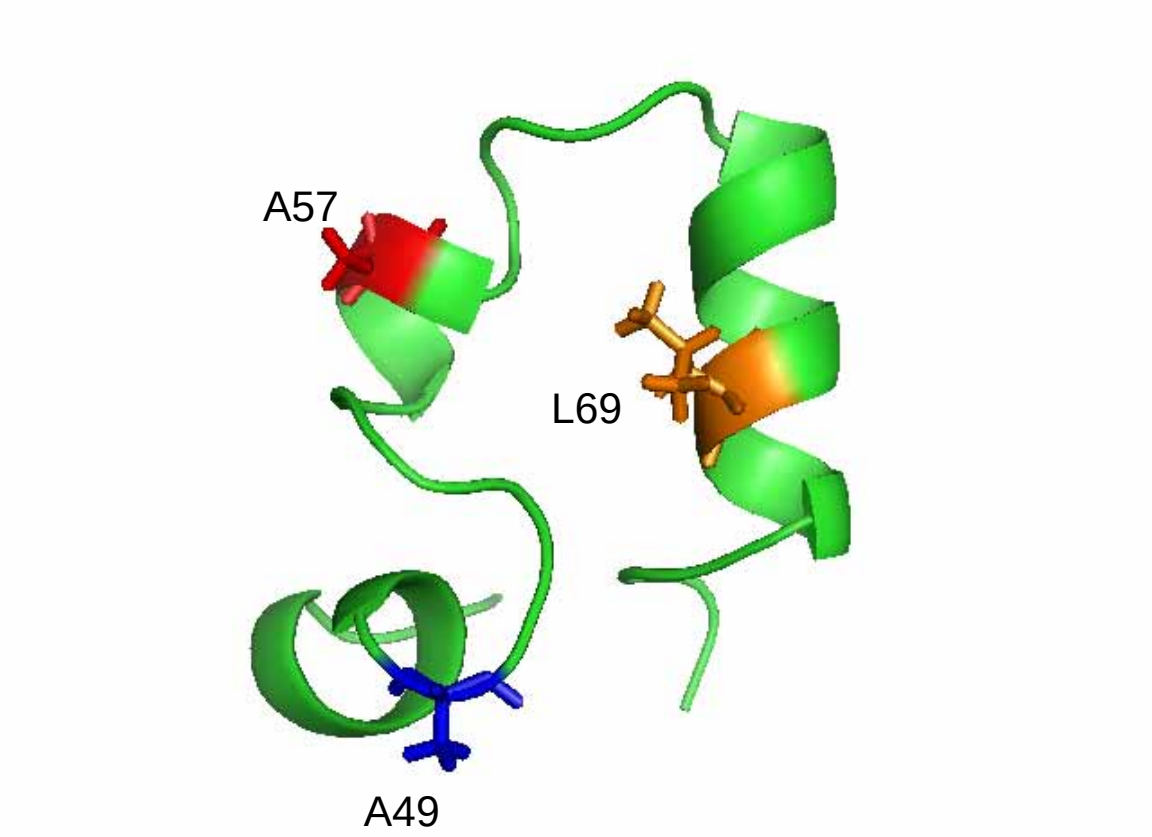


Figure 5-1 Ribbon diagram of villin headpiece HP36. The NMR structure was used to make the figure (PDB code: 1VII). The position of Ala 49, Ala 57 and Leu 69 are shown in ball-and-stick format and colored by blue, red and orange individually. The figure was created by PyMol.

	Carbonyl carbon	Carbonyl oxygen	backbone	sidechain
Ala 49	0.5 Å	19 Å	36 Å	62 Å
Ala 57	0	2 Å	8 Å	34 Å
Leu 69	0	0	0.2 Å	43 Å
Ala-*Ala-Ala	0	24 Å	40 Å	72 Å
Ala-*Leu-Ala	0	24 Å	38 Å	153 Å

Table 5-1 Calculated solvent accessible surface area of carbonyl carbon, carbonyl oxygen, backbone and sidechain of the labeled residues in HP36. The backbone exposure refers to the amide proton and nitrogen, α-carbon, carbonyl carbon and oxygen. For comparison the values for Ala and Leu are given for the central residue in a Ala-X-Ala tripeptide which adopts an extended conformation (φ, ψ=180^{o}).
*A tri-peptide was constructed by PyMol. The surface area of the middle residue was calculated.

We synthesized HP36 ^{13}C, ^{18}O-labeled A49, HP36 ^{13}C, ^{18}O-labeled A57 and HP36 ^{13}C, ^{18}O-labeled L69 individually using a ABI 433 A peptide synthesizer and standard Fmoc chemistry. The first residue in the C-terminus and residues with β-branch sidechains and the following residues need to be double coupled. The sequence of HP36 is MLSDEDFKAVFGMTRSAFANLPLWKQQNLKKEKGLF. So residues F76, P62, L61, T54, M53, V50 and A49 were double coupled.

Peptides were cleaved from the PAL-PEG-PS resin with a 91% trifluoroacetic acid (TFA), 3%anisole, 3%thioanisole, 3%ethanedithiol mixture. Purification was achieved by HPLC using a water/acetonitrile gradient containing 0.1% (v/v) trifloroacetic acid on a semi-preparative C18 column (Vydac) with a flow rate of 5 ml/min. A gradient program of increasing buffer B concentration from 20% to 70% in 50 minutes was used for all three proteins and all of them elute at the same retention time which is 21 minute. The buffer A is 99.9% water with 0.1% TFA; buffer B is 80% acetonitrile and 19.9% water with 0.1% TFA. Peptide identity was confirmed by mass spectrum and all peptides were shown greater than 95% pure in analytical HPLC runs. The molecular weights of HP36 ^{13}C, ^{18}O-labeled A49, HP36 ^{13}C, ^{18}O-labeled A57 and HP36 ^{13}C, ^{18}O-labeled L69 are 4191.9, 4190.3 and 4190.6. The expected molar mass is 4189.9.

5.3 References

1. Huang, R., Kubelka, J., Barber-Armstrong, W., Silva, R. A., Decatur, S. M., and Keiderling, T. A. (2004) Nature of vibrational coupling in helical peptides: an isotopic labeling study, *J. Am. Chem. Soc. 126*, 2346-2354.
2. Starzyk, A., Barber-Armstrong, W., Sridharan, M., and Decatur, S. M. (2005) Spectroscopic evidence for backbone desolvation of helical peptides by 2,2,2-trifluoroethanol: an isotope-edited FTIR study, *Biochemistry 44*, 369-376.

3. Brewer, S. H., Song, B., Raleigh, D. P., and Dyer, R. B. (2007) Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy, *Biochemistry 46*, 3279-3285.
4. Rosman, K. J. R. a. T., P. D. P. . (1998) Isotopic compositions of the elements 1997, *Pure Appl. Chem. 70*, 19.
5. Brewer, S. H., Vu, D. M., Tang, Y., Li, Y., Franzen, S., Raleigh, D. P., and Dyer, R. B. (2005) Effect of modulating unfolded state structure on the folding kinetics of the villin headpiece subdomain, *Proc. Natl. Acad. Sci. U. S. A. 102*, 16662-16667.