Stony Brook University

OFFICIAL COPY

The official electronic file of this thesis or dissertation is maintained by the University Libraries on behalf of The Graduate School at Stony Brook University.

© All Rights Reserved by Author.

Protein Folding and Stability

Distinguishing Folded from Unfolded State Effects

A Dissertation Presented

by

Shifeng Xiao

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University

May 2012

Stony Brook University

The Graduate School

Shifeng Xiao

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation.

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

Erwin London, Ph.D., Chairperson
Professor, Department of Chemistry and Department of Biochemistry and Cell Biology, Stony Brook University

Peter J. Tonge, Ph.D., Third Member
Professor, Department of Chemistry, Stony Brook University

Markus A. Seeliger, Ph.D., Outside Member
Assistant Professor, Department of Pharmacological Sciences, Stony Brook University

This dissertation is accepted by the Graduate School

Charles Taber
Interim Dean of the Graduate School

Abstract of the Dissertation

Protein Folding and Stability

Distinguishing Folded from Unfolded State Effects

by

Shifeng Xiao

Doctor of Philosophy

in

Chemistry

Stony Brook University

2012

The villin headpiece subdomain (HP36) is one of the smallest naturally occurring protein domains that folds cooperatively in the absence of disulfide bonds or ligand binding. It has a simple topology consisting of three α-helices that form a hydrophobic core. Kinetic studies have shown that the subdomain folds on the microsecond time scale, making it one of the fastest folding proteins. Its simple topology, small size, and rapid folding have made it a very popular model for computational, theoretical, and experimental studies.

HP36 has a well packed hydrophobic core comprised in part of an unusual set of three closely packed phenylalanine residues F47, F51, F58. Aromatic aromatic interactions have been conjectured to play a critical role in specifying the subdomain fold and have been proposed to play a general role in stabilizing small proteins. The modest stability of HP36 has hindered studies of core packing since multiple mutations can lead to constructs which fail to fold and even single mutants can result in poorly folded variants. Using a hyperstable mutant of HP36 as

the new background, generated by targeting surface residues, I show that aromatic aromatic interactions are not required for specifying the subdomain fold, although they have effects on the stability.

Proline-aromatic interactions involving P62 and W64 have been proposed to play a critical role in specifying the subdomain fold by acting as gatekeeper residues, i.e. as residues absolutely essential for specifying the fold. Mutation studies based on the same new background reveal that proline-aromatic interactions are not required for specifying the subdomain fold. These studies argue against the concept of specific gatekeeper residues. The implications for protein folding are discussed.

To probe unfolded state electrostatic interactions, the pH-dependent stability of wildtype HP36 and three mutants, K48M, K65M and K70M, all of which significantly increase the stability of the protein were examined. The increased stability of the K48M mutant is due to the removal of favorable electrostatic interactions in the unfolded state, the increased stability of the K65M mutant is due to the reduction of the desolvation penalty at the mutation site upon folding, while the increased stability of the K70M mutant is due to the introduction of a new hydrophobic interaction between the methionine and the hydrophobic core in the native state. The unfolded state electrostatic interactions were confirmed by double mutant thermodynamic cycle analysis and by using a method to estimate residue-specific unfolded state pKa values. The results demonstrate that electrostatic as well as hydrophobic interactions play an important role in the unfolded state, and illustrate an approach for distinguishing native state effects from unfolded state effects. This work also has interesting implications for studies which attempt to stabilize proteins by targeting surface electrostatics since it shows that the mechanism of stabilization may be much more complicated than generally anticipated.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AUC	Analytical unltracentrifugation
CD	Circular dichroism
C_M	Midpoint concentration of denaturant induced unfolding transition
DM	The hyperstable double mutant of HP36, N68AK70M
DSE	Denatured state ensemble
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
F	The folded state
HPLC	High performance liquid chromatography
HP21	The first 21-residue peptide fragment from HP36
HP35	The C-terminal subdomain of villin headpiece
HP36	The protein comprising HP35 with an additional N-terminal methionine
HP67	Villin headpiece without the first 9 residues
HP76	Villin headpiece
H/D	Hydrogen to Deuterium
I	The intermediate state
ICE	Integrated continuum electrostatics
LB	Luria-Bertani
M	Molarity
MALDI	Matrix assisted laser desorption and ionization
mM	Millimolarity
MD	Molecular dynamics

MT	Mutant
N	The native state
NIH	National institute of health
NTL9	The N-terminal domain of protein L9
NMR	Nuclear magnetic resonance
PB	Poisson Boltzmann
PDB	Protein data bank
ppm	Parts per million
R	Universal gas constant
ref	Reference
T	Temperature
TFA	Trifluoroacetic acid
T_m	Midpoint of thermal unfolding transition
TOCSY	Total correlation spectroscopy
TS	The transition state
T-jump	Temperature jump
U	The unfolded state
UV	Ultraviolet
WT	Wildtype
ΔC_p^o	Heat capacity change upon unfolding
ΔG^o_{N-D}	Standard Gibbs free energy of unfolding
ΔH^o	Standard enthalpy of unfolding
ΔS^o	Standard entropy of unfolding

[den]	Concentration of denaturant
[θ]	Mean residue ellipticity
3D	Three dimensional

Acknowledgments

I would like to thank my advisor Prof. Daniel Raleigh for guiding and supporting my research works. He can always propose great ideas to perform experiments, to solve experimental problems, to analyze data, to interpret experimental phenomena, etc. I am also grateful for his encouragement and patient revisions in my manuscript writings and presentation slides. He is really an excellent advisor in training doctoral students.

I thank Prof. Erwin London, Prof. Peter Tonge, and Prof. Markus Seeliger for serving as my dissertation committee members. I thank them for their valuable suggestions on my research projects and research proposal writing as well as their fast response to my meeting scheduling and reference letter request.

I appreciate the friendship with the past and present members of the Raleigh group: Dr. Yuan Bi, Dr. Benben Song, Dr. Ruchi Gupta, Dr. Bing Shan, Dr. Humeyra Taskent, Dr. Fanling Meng, Wenli Meng, Peter Marek, Vadim Patsalo, Ping Cao, Ivan Peran, Bowu Luan, Hui Wang, Ling-Hsien Tu, Matthew Watson, Harris Norr, Rehana Akter. I specially want to thank Dr. Yuan Bi, Dr. Bing Shan and Vadim Patsalo for several collaborative projects.

Last but not the least I would like to thank my family for their encouragement, support and understanding. Especially, I thank my wife Jing for her love and being always on my side. My life has become much more colorful since I got married to her.

Publications resulting from work described in this thesis

1. **Xiao, S.**, Bi, Y., Shan, B., and Raleigh, D. P. (2009) Analysis of core packing in a cooperatively folded miniature protein: the ultrafast folding villin headpiece helical subdomain. *Biochemistry* 48, 4607-4616.
2. **Xiao, S.** and Raleigh, D. P. (2010) A critical assessment of putative gatekeeper interactions in the villin headpiece helical subdomain. *J. Mol. Biol.* 401, 274-285.
3. **Xiao, S.**, Patsalo, V., Shan, B., Bi, Y., Green, D., and Raleigh, D. P. Rational modification of protein stability by targeting surface sites: simple approaches, complicated results. (*to be submitted*)

Additional Publications

1. Choi, U. B., **Xiao, S.**, Wollmuth, L. P., and Bowen, M. E. (2011) Effect of Src phosphorylation on the disordered C-terminal domain of the NMDA receptor subunit GluN2B. *J. Biol. Chem.* 286, 29904-29912.
2. Davis, C. M., **Xiao, S.**, Raleigh, D. P., and Dyer, B. R. Heterogeneity in the ultrafast folding of a 10-residue hairpin. (*to be submitted*)

1. Introduction

1.1 Protein folding and stability

1.1.1 The importance of protein folding

Proteins are the most important of all the molecules found in living organisms. They are used to support the skeleton, control senses, move muscles, digest food, defend against infections and process emotions. They are the most abundant component within a cell (more than half the dry weight of a cell is made up of proteins) and they have a range of indispensable roles. Protein folding is the physical process by which a polypeptide folds into its characteristic and functional three-dimensional structure from a random coil. Each protein exists as an unfolded polypeptide or random coil when translated from a sequence of mRNA to a linear chain of amino acids. Amino acids interact with each other to produce a well-defined three-dimensional structure, the folded protein, known as the native state. The resulting three-dimensional structure is determined by the amino acid sequence (*1*). The correct three-dimensional structure is essential to function. Failure to fold into native structure produces inactive proteins which can be toxic. Several neurodegenerative and other diseases such as Alzheimer's disease, Parkinson's disease, Huntington's disease, type-II diabetes, Creutzfeldt-Jakob disease and others are believed to result from the accumulation of amyloid fibrils formed by misfolded proteins (*2*). Understanding of protein folding pathways will help to explain how these diseases arise and develop cures for them. Proteins also have applications in biomaterials and biotechnology and an improved understanding of their folding and stability will aid these effects.

1.1.2 Mechanism and pathway of protein folding

A 100-residue protein would have 2^{100} conformations if there are two configurations for

each amino acid in the absence of excluded volume effects. If only 10^{-11} s were required to convert from one conformation to the other, a random search of all conformations would require 10^{20} s or 10^{13} years. Many *in vitro* studies showed that the folding process typically occurs on a milliseconds-to-seconds time scale (*3-5*). So there must be a way for a polypeptide to reach the native state in a time many orders of magnitude shorter than that required for a random search. This led Levinthal to suggest "pathways" for protein folding (*6*).

Protein folding processes are often classified as either a two-state (*7*) or a multi-state process (*8, 9*). The simplest case is two-state folding which does not involve any intermediates.

$$U \rightleftharpoons F$$

where U is the denatured state and F is the native state. Concepts of two-state folding were developed by Brandts who characterized the thermodynamics of the reversible thermal unfolding of chymotrypsinogen (*10*) and Tanford who characterized the guanidinium chloride induced unfolding reactions of several small proteins (*11, 12*). Some proteins may form stable intermediate states which can be on-pathway or off-pathway.

$$U \rightleftharpoons I_1 \rightleftharpoons I_2 \rightleftharpoons \cdots \rightleftharpoons F$$

where I means an intermediate state. Characterization of partly folded intermediates is a key approach to understanding the protein folding pathways. In 1971 two laboratories first reported that fast reaction kinetics revealed folding intermediates (*13, 14*). Intermediates can be studied by characterizing transient intermediates during refolding experiments or by studying equilibrium species populated by the use of non-native conditions or mutating residues (*15-18*). Among the traditional experimental methods, hydrogen exchange is a unique tool that allows the detection and characterization of intermediates. The parts of the structure that form rapidly, will be protected quickly, and thus not exchanged, whereas areas that fold late in the pathway will be

exposed to the exchange for longer periods of time. Recently, Vendruscolo (*19*) and Dokholyan (*20*) developed new computational methods to incorporate hydrogen exchange factors from NMR experiments as constraints into molecular dynamics (MD) simulations to characterize the conformations of intermediate states. For example, Vendruscolo and co-workers defined the structure of the exchange-competent intermediate state of Im7 by using equilibrium hydrogen-exchange protection factors and then used this ensemble to predict Φ-values and chemical shifts. These two species were predicted with high accuracy from the determined exchange-competent intermediate (*19*).

There are several models used to rationalize folding pathways. The diffusion/collision models (*21, 22*) and framework models (*23, 24*) state that partially stable helices form early through fluctuations, reducing the conformational search, so they can assemble into tertiary structures. In this view, secondary structure fluctuations precede collapse and assembly, and assume local interactions are the main determinants of protein structure. An alternative view, the so-called hydrophobic collapse model (*25*), proposes that collapse drives secondary structure formation, rather than the reverse. The nonlocal interactions that drive collapse processes in heteropolymers can give rise to protein structure, stability, and folding kinetics. Later on Fersht developed the nucleation-condensation mechanism which unites features of both the framework and collapse mechanisms (*26, 27*). In this scenario, a cluster of key residues need to form their native contacts in order for the folding reaction to proceed fast into the native state. However, critical reviews show that residues in the nucleus can be mutated and the proteins still fold (*28*). The so-called “new view” of protein folding kinetics which replaces the concept of folding pathways with energy landscape has emerged (*29-31*). In the so-called funnel model, each individual protein molecule from the top of the funnel, the unfolded state, may follow its own

trajectory, but they all may eventually reach the same point at the bottom of the funnel, the native state (Figure 1-1). Several possible trajectories shown in Figure 1-1 are: fast folding from unfolded state to the native state (yellow), slow folding that crosses the high energy barrier (green), and slow folding that returns to a less folded state and then follows the valley for fast folding (red). This view readily explains both the process of reaching a global minimum in free energy and doing so quickly by multiple folding routes on funnel-like energy landscapes.

1.1.3 Thermodynamics of protein folding

The stability of a protein is defined by the Gibbs free energy difference between the native state (N) and the unfolded state (D) (Figure 1-2):

$$\Delta G^{o}_{N-D} = G^{o}_{D} - G^{o}_{N} \tag{1.1}$$

The free energy can also be defined as:

$$\Delta G^{o}_{N-D} = \Delta H^{o}_{N-D} - T\Delta S^{o}_{N-D} \tag{1.2}$$

where ΔH^{o}_{N-D} is the enthalpy difference and ΔS^{o}_{N-D} is the entropy difference.

The enthalpy of a system is the integrated sum of the heat energy required to reach that state from 0 K:

$$H = \int_{0}^{T} Cp.dT + H_0 \tag{1.3}$$

where H_0 is the ground state energy (at 0 K) and C_p is the heat capacity constant.

Integration of equation 1.3 leads to:

$$\Delta H^{o}_{N-D} = \Delta C^{o}_{P}(T - T_o) + \Delta H^{o}_{N-D}(T_o) \tag{1.4}$$

where $\Delta H^{o}_{N-D}(T_o)$ is the enthalpy difference at a reference temperature T_o and ΔC_p is the heat capacity difference between the native state and the unfolded state which is related to the change

of solvent accessible surface area upon unfolding (*32*).

The change in entropy is given by:

$$\Delta S^o = \int_o^T \frac{C_p^o}{T}.dT \tag{1.5}$$

Integration of equation 1.5 will lead to:

$$\Delta S_{N-D}^o = \Delta C_P^o \ln\left(\frac{T}{T_o}\right) + \Delta S_{N-D}^o(T_o) \tag{1.6}$$

where $\Delta S_{N-D}^o(T_o)$ is the entropy difference at a reference temperature T_o.

Substituting equation 1.4 and 1.6 into equation 1.2 gives the Gibbs-Helmholtz equation:

$$\Delta G_{N-D}^o(T) = \Delta H^o(T_o) - T\Delta S^o(T_o) + \Delta C_P^o\left[T - T_o - T\ln\left(\frac{T}{T_o}\right)\right] \tag{1.7}$$

According to equation 1.7, the free energy of protein unfolding is temperature dependent. We can define the T_o as the mid-point of the thermal unfolding process, T_m, at which ΔG_{N-D}^o=0 and $\Delta S^o(T_m) = \Delta H^o(T_m)/T_m$. So we can rewrite equation 1.7 as:

$$\Delta G_{N-D}^o(T) = \Delta H^o(T_m)\left(1 - \frac{T}{T_m}\right) + \Delta C_P^o\left[T - T_m - T\ln\left(\frac{T}{T_m}\right)\right] \tag{1.8}$$

Thermal unfolding curves can be obtained by recording the signal as a function of temperature monitored by CD, fluorescence, NMR, etc. In a two-state folding process, ΔG_{N-D}^o can be also written as:

$$\Delta G_{N-D}^o = -RT\ln K_{eq} = -RT\ln\left(\frac{P_D}{1-P_D}\right) = -RT\ln\left(\frac{y - y_N}{y_D - y}\right) \tag{1.9}$$

where K_{eq} is the equilibrium constant, P_D, the fraction of unfolded state, y, the observed signal, y_N, the native state signal, y_D, the unfolded state signal.

Typically, the thermal unfolding curve can be fit to equation 1.10 using non-linear least squares to get T_m, $\Delta H^o(T_m)$ and ΔC_p.

$$y(T) = \frac{a_n + b_n T + (a_d + b_d T) e^{-\Delta G^\circ_{N-D}(T)/RT}}{1 + e^{-\Delta G^\circ_{N-D}(T)/RT}} \quad (1.10)$$

where a_n and b_n represent the intercept and slope of the pre-transition, a_d and b_d represent those of the post-transition. $\Delta G^o_{N-D}(T)$ can be given by equation 1.8.

Protein unfolding can also be induced by chemical denaturants such as urea or guanidine hydrochloride. The unfolding free energy at a certain concentration of denaturant is linearly related to the denaturant concentration (*33*):

$$\Delta G^o_{N-D}([den]) = \Delta G^o_{N-D}(H_2O) - m \times [den] \quad (1.11)$$

where $\Delta G^o_{N-D}(H_2O)$ is the protein stability in water and *m* is the slope which represents the change of buried surface area upon unfolding (*32*). The chemically induced denaturation curve can be fit to equation 1.12 to give $\Delta G^o_{N-D}(H_2O)$ and the m value.

$$y([den]) = \frac{a_n + b_n[den] + (a_d + b_d[den]) e^{-\Delta G^\circ_{N-D}([den])/RT}}{1 + e^{-\Delta G^\circ_{N-D}([den])/RT}} \quad (1.12)$$

where a_n, b_n, a_d, b_d are defined in equation 1.10 and $\Delta G^o_{N-D}([den])$ can be given by equation 1.11.

1.1.4 Kinetics of protein folding

Understanding the protein folding kinetics can provide insights into folding mechanisms. Up until the last 10 to 15 years, folding kinetics were studied almost exclusively using stopped-flow techniques. In this experiment, protein folding is initiated by rapidly mixing a chemically denatured protein solution with a buffer to dilute the denaturant and restore

native-like conditions. The kinetics of folding are then monitored with one of the several optical spectroscopic methods. Similarly, unfolding can be initiated by mixing a native protein solution with concentrated denaturant. Unfortunately, the dead time of common stopped-flow devices is usually around 1 millisecond. Some studies on synthetic polymers showed that elementary processes such as α-helix formation are too fast to be observed in stopped-flow experiments (*34*). More recently, some fast kinetic methods have been developed which can be roughly classified into three categories: photochemical triggering, temperature or pressure jump, and ultrarapid mixing methods.

The first fast-folding study employed a photochemical trigger—the photodissociation of carbon monoxide from denatured cytochrome *c* (*35*). The experiment has unlimited time resolution because photodissociation occurs in less than 1 picosecond, faster than protein conformational change. Another kind of photochemical trigger uses a photolabile disulfide which does not require a metal site (*36, 37*).

Temperature jumps are a much more generally applicable method because they can be used to perturb the folding equilibrium for any process that produces a significant enthalpy change. This was first done by heating a solution of a dye with a mode-locked picosecond laser (*38*). The time resolution was ~70 ps. It is now done by direct heating of water by direct excitation of water modes using a laser pulse. The Eigen T-jump method of resistive heating with an electrical discharge has also been employed to study folding kinetics with time resolution ~10 μs (*39, 40*). Another perturbation method, pressure jump, has been used to study folding kinetics (*41*). In this experiment, a stack of piezo-electric crystals was used to change the pressure 100-200 bar in 50-100 μs, rapidly (un)folding the protein.

There has been considerable effort to improve the time resolution in mixing experiments by

continuous flow methods based on turbulent mixing (*42-44*) and hydrodynamic focusing (*45*, *46*). The dead time of these instruments is 50-200 μs.

Protein folding kinetics can also be studied at equilibrium using dynamic NMR (*47*, *48*). Usually, an NMR line changes its position by Δν upon transition from the folded state to the unfolded state because of a change of the chemical environment of the nuclei. With high-field NMR spectrometers, rate constants of exchange between F and U in the submillisecond time scale may be estimated using line broadening measurements of suitable ^{1}H lines with changes of chemical shifts upon unfolding by a few ppm. NMR relaxation dispersion experiments have also been applied and are more general (*49*).

1.2 Major forces in protein folding and stability

1.2.1 Hydrogen bonding

A hydrogen bond contains both positive (H-donor) and negative (H-acceptor) partial charges. It represents a combination of covalent and electrostatic interactions, but the main part is the electrostatic attraction between hydrogen donor and acceptor. Hydrogen bonding plays an important role in determining protein structure. In the secondary structure of proteins, hydrogen bonds form between the backbone oxygens and amide hydrogens. An α-helix is formed when the spacing of the amino acid residues participating in a hydrogen bond occurs between positions *i* and *i* + 4 (*50*). A β-sheet is formed when two strands are joined by hydrogen bonds involving alternating residues on each participating strand. Hydrogen bonds also play a part in forming the tertiary structure of protein through interactions of residue sidechains. Fersht and coworkers have reported favorable H-bond energies in protein engineering experiments, with values ranging from 0.5 to 1.8 kcal/mol for polar partners and 3.5 to 4.5 kcal/mol when one of the partners bears

a charge (*51*), but other groups have argued that H-bonds are not stabilizing or that the effects are smaller. Part of the disagreement has to do with the reference state and the proper accounting for desolvation effects. An ideal experiment would involve replacing the donor and acceptor by non-polar isosteres. Variable views on the importance of hydrogen bonds can be found in articles by Pace (*52*), Dill (*53*), Honig (*54*) and others.

1.2.2 Hydrophobic interaction

The hydrophobic interaction is an entropic effect originating from the disruption of highly dynamic hydrogen bonds between molecules of liquid water by the nonpolar solute. Hydrophobicity is usually expressed as the Gibbs free energy of transfer from water into the reference state (*55*). The transfer of the sidechains of hydrophobic amino acid residues, for example, leucine, isoleucine, and valine, from cyclohexane into water is energetically costly. Thus, the burial of hydrophobic sidechains in the folding process is energetically favorable.

The Gibbs free energy of transfer of a non-polar molecule from a reference state, such as cyclohexane, into water is composed of enthalpic and entropic terms:

$$\Delta G^{o}_{\text{hyd}} = \Delta H^{o}_{\text{hyd}} - \text{T}\Delta S^{o}_{\text{hyd}} \qquad (1.13)$$

At room temperature, $\Delta H^{o}_{\text{hyd}}$ for the transfer is small and $\Delta G^{o}_{\text{hyd}}$ is dominated by the entropy term (*56*). This is mainly because the formation of ordered water cage around non-polar residues is an entropically costly process.

Different chemical groups make different contributions to the Gibbs free energy. For example, for non-cyclic structures, the Gibbs free energy of transfer from vapor into water for the groups $–CH_3$, $–CH_2–$, –CH<, >C<, >C=O, $–NH_2$, –OH, and –NH– are 0.89, 0.17, –0.39, –1.16, –5.51, –5.63, –6.10, and –6.13 kcal mol^{-1} respectively (*55*).

1.2.3 Electrostatic interactions

Electrostatic interactions are the strongest forces among the non-covalent interactions. In folded proteins, pairs of neighboring, oppositely-charged residues often interact to form salt bridges. Although they are sometimes not critical for protein folding, they can contribute to protein stability (*57-61*). Whether or not they make a net favorable contribution is still debated. Kumar and Nussinov examined 222 non-equivalent salt bridges derived from 36 non-homologous high-resolution monomeric protein crystal structures (*62*). They concluded that most of the salt bridges are stabilizing towards proteins, regardless of whether they are buried or exposed, isolated or networked, hydrogen bonded or non-hydrogen bonded. For most salt bridges, the desolvation cost for bridge formation is paid off primarily by the electrostatic interaction of the salt-bridging residues with each other, and the interaction of the salt bridge with the rest of the protein is of secondary importance. Other people think differently. Hendsch and Tidor calculated the electrostatic contribution to the free energy of folding for 21 salt bridges in 9 proteins and they showed that the majority were electrostatically destabilizing because the unfavorable desolvation contribution was not fully compensated by favorable interactions within the salt bridge (*63*). The geometrical positioning of the interacting side-chains in the salt bridges with respect to each other is a very important determinant of salt bridge strength.

1.3 The denatured state of proteins

1.3.1 The significance of the denatured state

The denatured state of proteins is the reference state for thermodynamic studies and the starting state of protein folding (Figure 1-2). The folded state can be characterized by NMR and X-ray crystallography, and the late intermediates and transition states have been analyzed by

protein engineering and quench-flow NMR techniques (*64*). However, people know relatively little about the unfolded states, especially those formed under native-like conditions (*65*). Unveiling the structural and dynamic properties of denatured proteins is crucial for understanding the protein folding and misfolding problems. For example, the computational determination of thermodynamic stability requires an accurate approximation of the denatured state. NMR hydrogen exchange experiments also rely upon models of denatured state to produce the intrinsic hydrogen exchange rate (*66*). Residual structure in the denatured state is thought to reduce the initial conformational search and thus accelerate protein folding. Protein stability can be increased by destabilizing the denatured state, so the denatured state can be used as a target in protein engineering (*67*). A number of human diseases such as Alzheimer's disease, Parkinson's disease, Huntington's disease, type-II diabetes, Creutzfeldt-Jakob disease, etc are believed from protein aggregation and amyloid fibrils formed from unfolded states or partially unfolded states (*2*). Mutational studies indicate that changes in the denatured state structures may result in modulation of the kinetics of amyloid formation (*68*).

1.3.2 The "structure" of denatured state

It was initially believed that the denatured state of proteins is a featureless random coil. According to Flory's random coil theory (*69*), the size of a random coil polymer, its radius of gyration R_g, depends on the length of polymer chain, N.

$$R_g = R_0 N^{\nu} \tag{1.14}$$

where R_0 is the scaling constant and ν is the power law scaling exponent. Flory predicted the exponent to be 0.6. Tanford *et al* first confirmed this random coil scaling behavior for 12 proteins denatured by 6M guanidine HCl (*70*). They obtained the scaling exponent 0.67 ± 0.09. Recently,

Kohn *et al* showed the scaling behavior for 17 denatured proteins using small angle X-ray scattering and they found the exponent to be 0.598 ± 0.029 (*71*). In addition, a computational study confirmed this behavior by generating an ensemble of protein conformations whereby only steric interactions between amino acids were considered for four proteins. A scaling exponent of 0.58 ± 0.02 was obtained (*72*). This work led to the concept of unfolded states as random coil.

However, proteins often exhibit native conformational bias and retain a certain amount of residual structures even under denaturing conditions. More and more experimental evidence supports residual native-like structural elements in the denatured state (*73-86*). For example, Bu *et al* showed residual helical structure and tertiary-like interactions even in the absence of disulfide bonds and under highly denaturing conditions by applying quasi-elastic neutron scattering on α-lactalbumin (*74*). Through a series of 3D NMR experiments, Meng *et al* determined that there is significant native-like structure in HP21 which is a good model for structure in the unfolded state of HP36 in the absence of denaturant (*80*). A large number of experiments suggested structure in the denatured state of Staphylococcal nuclease (*83-86*). Several computational studies have also reported some specific interactions in conformational biasing toward the native state in the denatured state (*87-89*). For example, using a simple force field with only steric and hydrogen bond interactions, Pappu *et al* showed that protein denatured states have a strong preference for the native structure (*87*).

There is no single denatured state. In strongly denaturing conditions, the most populated denatured species are highly unfolded. Under native conditions, the most populated unfolded species are more compact (*90, 91*). The compact denatured states have some structure that is sequence dependent and native-like. They have disorder in both the sidechains and the backbone (*92*). They have radii slightly greater than the native state (*93-95*). They also form local contacts,

particularly helices and turns (*80, 95, 96*). Furthermore, multiple or diffuse hydrophobic clusters are often formed, but no well-defined hydrophobic core (*97-99*). Hydrophobic clustering involves a much larger solvent exposed surface area than the native hydrophobic core.

It is often assumed that electrostatic interactions play a minor role in the denatured state because the charge residues are thought to be well solvated in the denatured state. However, several recent studies have shown that both favorable and unfavorable electrostatic interactions can be formed in the denatured state and they can have a large effect on the energetics of denatured state (*60, 100-105*).

1.3.3 pH-dependent stability to probe denatured state electrostatics

To probe electrostatic interactions in the denatured state, pH-dependent stability analysis is a powerful approach which can be analyzed by using the Wyman-Tanford linkage relationship (*12*):

$$\frac{\partial \Delta G^o}{\partial pH} = 2.303RT\Delta Q \tag{1.15}$$

where ΔQ is the difference between protons bound in the native state and the denatured state. ΔG^o is the free energy of unfolding. The net charge Q can be obtained from the degree of protonation, $\theta_{i,A}$, of the individual titratable residue by Henderson-Hasselbalch equation:

$$\theta_{i,A} = \frac{10^{(pK_i^A - pH)}}{1+10^{(pK_i^A - pH)}} \tag{1.16}$$

where pK_i^A is the pKa of residue *i* in A state (either native or denatured state). It should be noted that the linkage relationship (equation 1.15) is always valid but equation 1.16 is not valid for strong interactions between titratable sites (*106*). Assuming the titration behavior can be described as the sum of individually titrating sites, the pH-dependent stability can be obtained by

integrating equation 1.15 and applying equation 1.16.

$$\Delta\Delta G°(pH - pH6) = RT\sum_{i=1}^{j}\ln\left[\frac{(1+10^{(pH-pK_{i,N})})(1+10^{(6-pK_{i,D})})}{(1+10^{(6-pK_{i,N})})(1+10^{(pH-pK_{i,D})})}\right] \quad (1.17)$$

where i is the residue, j is the number of acidic residues, $pK_{i,N}$ is the native state pKa value of residue i, $pK_{i,D}$ is the denatured state pKa value of residue i, and pH6 is chosen as a reference pH.

In proteins, the pKa values of the ionizable groups may be substantially raised or lowered by environment effects. The three most important effects are (a) dehydration, (b) charge-charge interaction, and (c) charge-dipole interaction (Figure 1-3). (a) It is energetically unfavorable to transfer a charged group from water to the interior of a protein where the dielectric constant ($\epsilon_{protein}$) is lower. The neutral form of the ionizable group is favored, and thus the pKa values of acidic residues will be increased and those of basic residues will be lowered when the groups are buried in a folded protein. (b) The pKa values of acidic residues will be raised by a negatively charged environment and decreased by a positively charged environment. (c) The effect of hydrogen bonding on the pKa values depends on the interactions with protonated or deprotonated form. The pKa will be increased when the interactions are more favorable with the protonated state and decreased when the interaction with deprotonated state are more favorable.

pKa values for the native state of protein can be measured by following the NMR chemical shifts of the acidic residues as a function of pH, while those for the denatured state cannot be experimentally measured in stable proteins. To the best of my knowledge, the drkN SH3 domain and the V3AI4A mutant of NTL9 are the only examples of cooperatively folding globular proteins for which denatured state pKa's have been directly determined under native conditions (*107, 108*). Most studies make use of random coil pKa's or model compound pKa's (*109*) or pKa's from strongly destabilizing mutants (*108*) for the denatured state pKa's to calculate the

$\Delta\Delta G^{o}$. Comparing the pH-dependent stability curve calculated with equation 1.17 to the experimental curve will indicate whether the model can account for the denatured state. A discrepancy between the two proves that there must be interactions in the denatured state that perturb the denatured state pKa values from the model compound. Thus, denatured state electrostatic interactions can be probed and models tested. As illustrated in Figure 1-4, if there is electrostatic interaction in the denatured state, the calculated pH-dependent stability curve deviates from the experiments. If there is no electrostatic interaction in the denatured state, the calculated pH-dependent stability curve agrees well with the experimental points. It should be noted that agreement between the calculated and experimental curves does not prove that the chosen model denatured state pKa's provide an accurate description of the titration behavior of the denatured state because the favorable and unfavorable electrostatic interactions may cancel out each other's effects (*110*).

1.4 The villin headpiece domain and its subdomain

Villin is an F-actin binding protein which plays an important role in the formation of microvilli of the gut and kidney (*111*). Villin headpiece (HP76) is located at the extreme C-terminus of villin which contains 76 amino acid residues and retains full F-actin-binding activity in isolation (*112*). HP67, generated by removing the first 9 residues in the N-terminal headpiece, is the smallest biologically active unit of villin (*113*). Its structure contains two subdomains: an N-terminal subdomain (P10-H41) and a C-terminal subdomain (L42-F76). The numbering system corresponds to the full-length protein. The N-terminal subdomain consists of loops, turns, and a short stretch of α-helix. It cannot fold in isolation but its folded form is essential for the specific F-actin binding activity. The C-terminal subdomain (HP35) consists of

three α-helices and can fold by itself (*113*). The sequence alignment of HP35 and its homology sequences is shown in Figure 1-5.

The helical C-terminal subdomain of the villin headpiece, denoted as HP36, comprises HP35 with an additional N-terminal methionine from the expression construct. Met is designated as residue 41. HP36 is one of the smallest naturally occurring protein domains that folds cooperatively in the absence of disulfide bonds or ligand binding (*114*). The subdomain has a simple topology consisting of three α-helices that form a hydrophobic core (Figure 1-6). Kinetic studies have shown that the subdomain folds on the microsecond time scale, making it one of the fastest folding proteins (*115, 116*). Its small size, simple topology and fast folding have made it a very popular model for computational, theoretical, and experimental studies (*117-130*).

1.5 The aims of this thesis

This thesis describes the folding, stability, and dynamics of the villin headpiece subdomain (HP36). The role of aromatic aromatic interactions in the protein fold and stability was examined by characterizing a complete set of single, double, and triple Phe to Leu mutants based on a hyperstable variant as the new wildtype background. The role of the putative gatekeeper interactions between Pro62 and Trp64 was studied by characterizing a set of mutants involving the two residues on the same new background. To find out the mechanisms of stability increase for three stabilized mutants, folded state and unfolded state electrostatic interactions were investigated by protein mutagenesis, pH-dependent stability measurements, molecular dynamic simulations, double mutant cycle analysis, and estimation of residue-specific unfolded state pKa values.

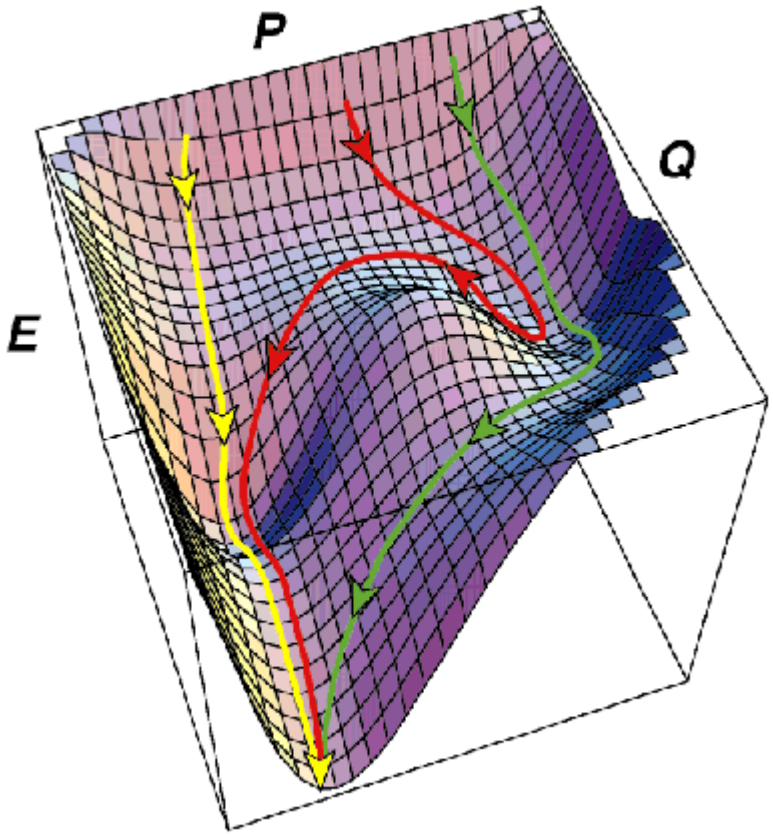


Figure 1-1. A three dimensional energy surface for protein folding. E refers to the energy of the system, P is a measure of available conformational state, and Q is the fraction of native contacts. Adapted from ref. (*131*).

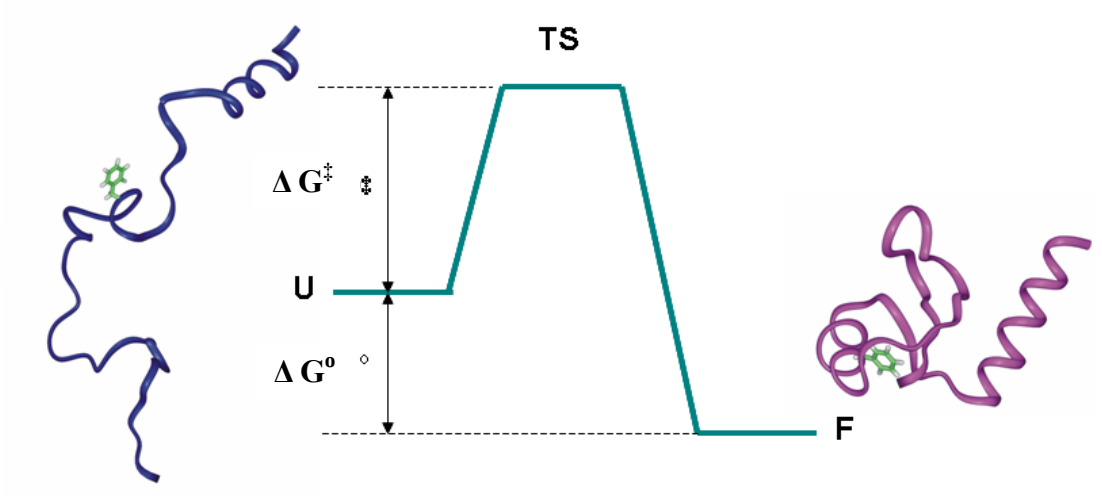


Figure 1-2. A free energy diagram of a two-state folding reaction. U means the unfolded state, TS, the transition state, F, the native state. $\Delta G^{\ddagger}$ and ΔG^{o} represent the activation free energy and the unfolding free energy, respectively.

(a). Dehydrogen (Born Effect):

$-COOH \rightleftharpoons -COO^- + H^+$ ↑ nonpolar environment: pKa ↑

$-NH_3^+ \rightleftharpoons -NH_2 + H^+$ ↑ nonpolar environment: pKa ↓

(b). Charge-charge interactions (Coulombic):

$-COOH \rightleftharpoons -COO^- + H^+$ ↑ positive charge: pKa ↓

$-NH_3^+ \rightleftharpoons -NH_2 + H^+$ ↑ positive charge: pKa ↓

(c). Charge-dipole interactions (Hydrogen bonding):

$-COOH \rightleftharpoons -COO^- + H^+$ ↑ hydrogen bonding to protonated form: pKa ↑

$-NH_3^+ \rightleftharpoons -NH_2 + H^+$ ↑ hydrogen bonding to protonated form: pKa ↑

Figure 1-3. Factors that affect pKa values. (a) pKa change due to Born effect when an ionizable group is buried in the interior of the protein where the dielectric constant is lower than that of water. The lower dielectric constant favors the neutral form of the ionizable group. (b) the pKa values of all the ionizable groups will be lowered by a positively charged environment and increased by a negatively charged environment. (c) the pKa values will be increased when hydrogen bonding is tighter to the protonated form and lowered when it is tighter to the deprotonated form. Adapted from ref. (*132*).

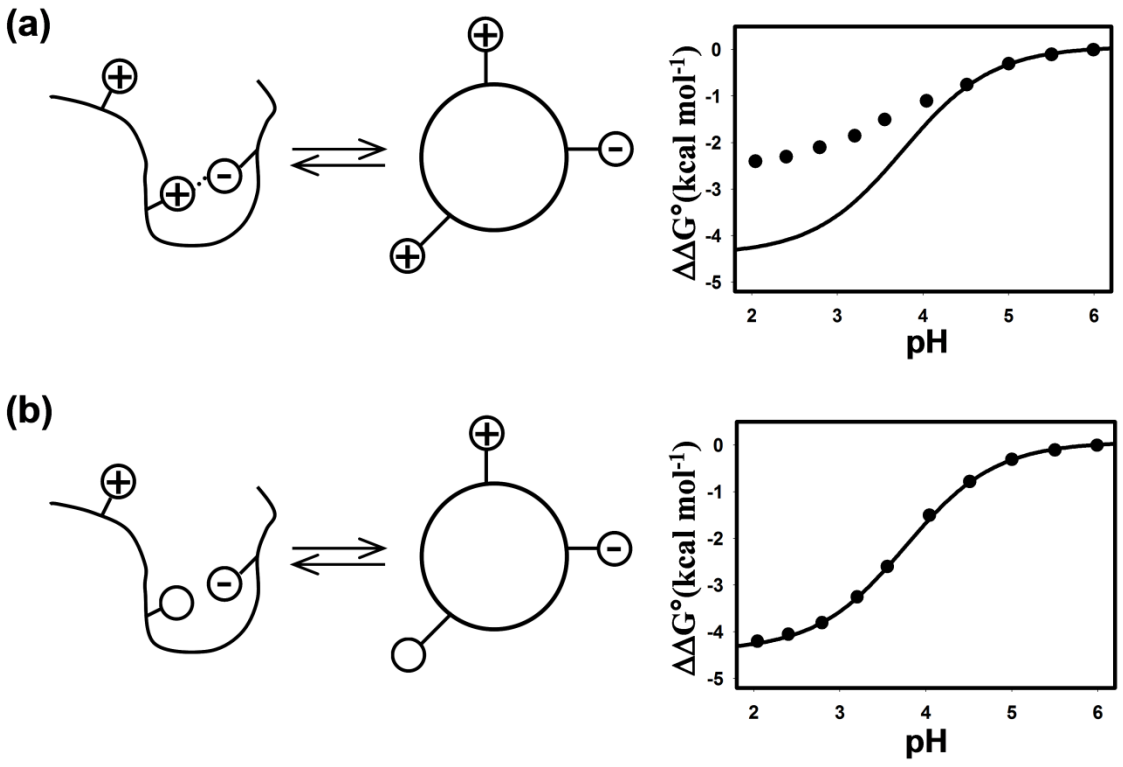


Figure 1-4. A diagram illustrating the use of pH dependent stability measurements to probe denatured state electrostatic interactions. (a) A protein with electrostatic interactions in the denatured state. The calculated stability curve determined using model compound pKa's deviates from the experimental points. (b) A protein with no electrostatic interactions in the denatured state. The calculated stability curve agrees well with the experimental points. Adapted from ref. (*110*).

	42	52	62	72
Chicken villin	LSDEDFKAVF	GMTRSAFANL	PLWKQQNLKK	EKGLF
Mouse villin	LSTEDFTRAL	GMTPAAFSAL	PRWKQQNIKK	EKGLF
Human villin	LSIEDFTQAF	GMTPAAFSAL	PRWKQQNLKK	EKGLF
Bovin villin	LSIEDFTRAL	GMTPSAFWAL	PRWKQQNLKK	EKGLF
X. laevis villin	LSDAEFAAIL	GMPKSQFSQL	PKWKQQNLKK	EKGLF
D. discoideum villin	LSDEEFLSTF	KMTKEIFQKT	PAWKTKQLRV	DNGLF
A. thaliana villin1	LTEKEFEERF	GMAKSEFYAL	PKWKQNKLKI	SLHLF
A. thaliana villin2	LSEEEFQSVF	GIEKEAFNNL	PRWKQDLLKK	KFDLF
A. thaliana villin3	LSEVEFKTVF	GMEKESFYKL	PGWKQDLLKK	KFNLF
A. thaliana villin4	LSSEEFKEKF	GMTKEAFYKL	PKWKQNKFKM	AVQLF
Human advillin	LSEQDFVSVF	GITRGQFAAL	PGWKQLQMKK	EKGLF
Bovin advillin	LSEKDFVSVF	GITRGQFAAL	PGWKQLQMKK	EKGLF
Mouse advillin	LSEQDFVSVF	GITRGQFTAL	PGWKRLQLKR	ERGLF
Chicken supervillin	LSDEDFEVAL	EMTREEYNAL	PSWKQVNLKK	AKGLF
Mouse supervillin	LTDEDFEFAL	DMSRDEFNAL	PTWKQVNLKK	SKGLF
Human supervillin	LTDEDFEFAL	DMTRDEYNAL	PAWKQVNLKK	AKGLF
Human dematin	LSAEDFSRVF	AMSPEEFGKL	ALWKRNELKK	KASLF
Mouse dematin	LSAEDFSRVF	AMSPEEFGKL	ALWKRNELKK	KASLF
Human actin-binding LIM1	LAPEVFREIF	GMSIQEFDRL	PLWRRNDMKK	KAKLF
Human actin-binding LIM2	LSPEEFQEVF	GMSIEEFDRL	ALWKRNDLKK	KALLF
Human actin-binding LIM3	LSQEEFYQVF	GMTISEFDRL	ALWKRNELKK	QARLF
Mouse actin-binding LIM1	LAPEVFWEIF	GMSIQEFDKL	PLWRRNDMKK	KAKLF
Mouse actin-binding LIM2	LSPEEFQEVF	GMSIEEFDRL	ALWKRNDLKK	KALLF
Mouse actin-binding LIM3	LSQEEFYQVF	GMTISEFERL	ALWKRNELKK	QARLF
Rat actin-binding LIM2	LSPEEFQEVF	GMSIEEFDRL	ALWKRNDLKK	KALLF
D. discoideum villidin	LNDEDFEKVF	KMTRTEWLKI	PAWKREGIKK	ELFLF
D. discoideum talin	LSDEEFKAVF	NCERSELAAM	PTWKRNNIKT	KLGLF
D. melanogaster quail	LTHDDFVSVF	NMSFYEFDEL	PKWKKMELKK	QFKLF
A. gambiae quail	LTHDDFVTVF	SMTYHEFEEL	PKWKQVELKK	QKKLF

Figure 1-5. Sequence alignment of chicken villin headpiece subdomain and its homology sequences. The alignment was generated using BLAST (http://blast.ncbi.nlm.nih.gov/). The left side is the protein name and the right side is the corresponding sequence. Note that the numbering system is according to the full length headpiece.

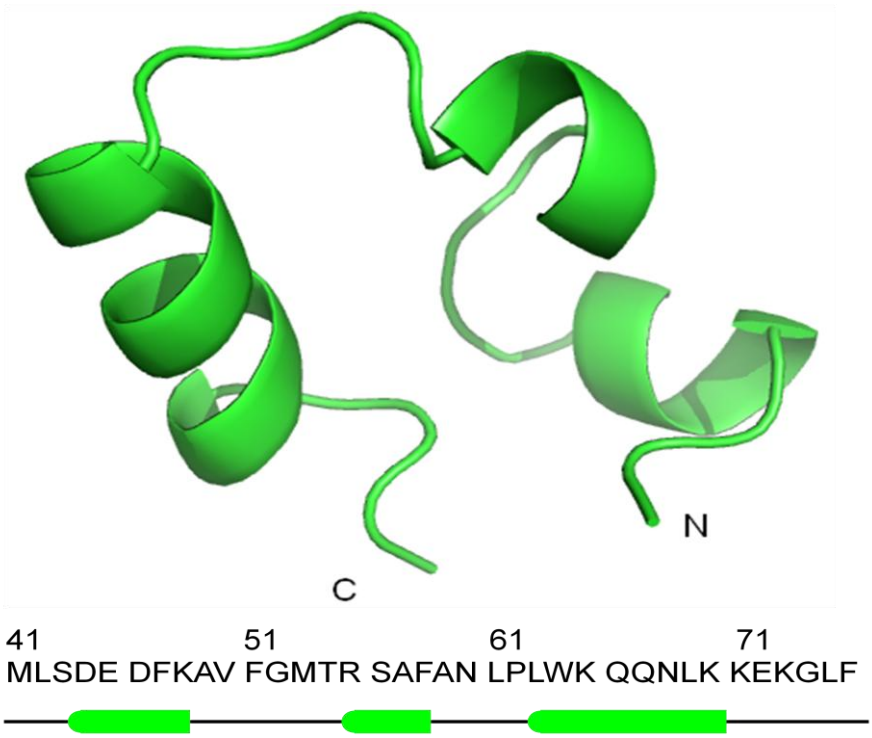


Figure 1-6. Ribbon diagram and primary sequence of villin headpiece subdomain HP36, based on PDB file 1VII. The N- and C- termini are labeled. The primary sequence is shown together with a diagram of secondary structure. α-helices are represented as cylinders.

2. Analysis of core packing in a cooperatively folded minature protein: The ultrafast folding villin headpiece helical subdomain

Abstract

The helical subdomain of villin headpiece has a well packed hydrophobic core comprised in part of an unusual set of three closely packed phenylalanine residues F47, F51, F58 (denoted using the numbering of the larger headpiece protein). Aromatic aromatic interactions have been conjectured to play a critical role in specifying the subdomain fold and have been proposed to play a general role in stabilizing small proteins. The modest stability of the subdomain has hindered studies of core packing since multiple mutations can lead to constructs which fail to fold and even single mutants can result in poorly folded variants. Using a previously characterized hyperstable mutant of the domain, generated by targeting surface residues, a complete set of single, double and triple core Phe to Leu mutants were characterized. A highly conserved surface Trp which is part of a Trp-Pro interaction was also examined. All mutants are well folded as judged by CD and NMR and all exhibit sigmoidal urea and thermally induced unfolding transitions thus, proving that aromatic aromatic, aromatic proline or aromatic hydrophobic interactions are not required for specifying the subdomain fold. Double mutant cycle analysis demonstrates that F47 and F51 has the largest interaction free energy in the native state. Mutations which lack F58 are the most destabilized although even the triple mutant is folded. Interestingly, mutation of the central Phe, F51, has the smallest effect on stability even though it makes contact with both F47 and F58 and appears to form the strongest pairwise interaction.

Acknowledgements

The data presented in this chapter has been published (Xiao, S., Bi, Y., Shan, B. and Raleigh, D.P. (2009) *Biochemistry* 48, 4607-4616). This chapter contains direct excerpts from the manuscript with a few adjustments. I thank Mr. Wenli Meng for helpful discussions and Mr. Vadim Patsalo for his help with the contact map calculation. I also thank Professor Mcknight for his advice and continued interest in this project.

2.1 Introduction

There is considerable interest in small cooperatively fold protein domains and subdomains since their small size and often very rapid folding make them highly attractive model systems for computational, theoretical and experimental studies of protein folding. In this regard the helical subdomain of the villin headpiece has emerged as perhaps the most popular system for molecular dynamics (MD) simulations of protein folding. The subdomain is one of the smallest naturally occurring protein domains which have been shown to fold cooperatively. The domain folds on the microsecond time scale and is thus one of the fastest folding proteins known. Its rapid folding coupled with its small size, 35 or 36 residues depending on the construct, and simple topology have made it the focus of a very large number of computational and theoretical studies with more than twenty independent research groups, publishing the results of their folding calculations (*5, 113-115, 117-124, 126, 130, 133-159*). There have been considerably fewer published experimental studies on the kinetics of its folding but we and the NIH group have used laser temperature jump methods to show that it folds on the microsecond time scale (*5, 127, 134, 158*).

The villin headpiece helical subdomain consists of residues Leu42 to Phe76 of villin and is denoted HP35 (Figure 2-1). The subdomain sequence is normally numbered according to the numbering used for the intact villin headpiece. Thus the first residue of HP35 is Leu42 and its C-terminus is residue 76 (*133*). Recombinant versions of HP35 retain the N-terminal Met and are thus 36 residues in length. That construct, which is the parent sequence for the present studies, is denoted HP36 and the N-terminal Met is designated Met41. The domain has a simple topology made up of three α–helices (Figure 2-1). HP35 and HP36 have very well packed hydrophobic cores and contain a striking triad of buried phenylalanine residues at positions 47, 51 and 58 which pack against each other in the native state (*114, 160*). As shown in Figure 1-4, F47 is

completely conserved, position 51 is a Phe in 23 out of 29 known sequences and replaced by Leu in the other 6 sequences, and F58 is conserved in 25 out of 29 sequences and a Leu, Tyr or Trp is found in the other 4 sequences. There are no known examples of double mutants which lack aromatic residues at both positions. The protein also contains a conserved Trp at position 64 which has been proposed to play a critical role in dictating the fold via an aromatic proline interaction (*157*). Interactions involving aromatic residues are observed in many, but not all, naturally occurring or designed minature proteins (*161-163*) and the hypothesis has been advanced that they play a particularly critical role in stabilizing and specifying their folds (*157, 161-164*). Aromatic aromatic interactions have also been proposed to play an essential role in the ordered self assembly of small amyloidogenic polypeptides (*165-168*). The sidechains of the three core aromatic residues in HP36 are the three most buried in the hydrophobic core with percent fractional accessibilities of 2, 6 and 6% respectively which is the ratio of side-chain surface area to "random coil" value per residue (*169*). The "random coil" value of a residue X is the average solvent-accessible surface area of X in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations.

Surprisingly, there have been very few mutational studies of the role of the hydrophobic core in folding or stability of the subdomain, presumably in part because the wildtype is not particularly stable in terms of ΔG^o although, as expected it has a high T_m (*5, 147, 157, 160, 170, 171*). The modest ΔG^o of folding for the wildtype has the unfortunate consequence that even moderately destabilizing single or multiple mutants may prevent the subdomain from folding. The failure to fold could arise because the mutations disrupt critical interactions which are absolutely required to specify the fold or they may simply destabilize the native state so much that only a very very small fraction of molecules are folded. In the latter scenario the small

fraction of molecules which are folded do adopt the wildtype structure while in the former the mutants are incapable of adopting the wildtype structure even under stabilizing conditions because critical interactions are missing. The difference in the two scenarios gets to the heart of the difference between the protein folding vs. the protein stability problems (*172*) and is discussed in more detail in chapter 3 of this thesis. In the case of HP35/ HP36 the failure of certain Phe mutants to fold leads to the question of whether or not aromatic aromatic interactions are required for specifying the fold in this unusual small protein domain. In addition, studies of the conserved Trp at position 64 have lead to the proposal that it participates in critical "gatekeeper" interactions which play a key role in specifying the fold (*157, 173*). Those studies, like the previous analysis of the core aromatics, were conducted in the normal wildtype background and lead to significant destabilization.

The Raleigh group has described a hyperstable double mutant of the domain which was generated by mutating two surface residues whose sidechains are exposed to solvent. The high resolution crystal structure of the double mutant shows that its structure and core packing are identical to wildtype (*147*). I took advantage of this hyperstable variant to conduct an analysis of core aromatic packing in the villin headpiece helical subdomain and to analyze the role of the surface Trp. Many of the mutants characterized here would so destabilize the wildtype protein that it cannot fold. We show that interactions involving aromatic residues are not required to specify the HP36 fold.

2.2 Materials and methods

2.2.1 Protein expression and purification

The plasmid (pET3a-NTL9-FXa-HP36) containing the gene for HP36 was prepared as

described (*147*). The mutant proteins were expressed as fusion proteins with the N-terminal domain of L9 and the fusion protein was cleaved by Factor Xa and purified as described (*147*). The cleavage temperature for DM F47LF58L, DM F51LF58L and DM F47LF51LF58L was 4 ^{o}C while the cleavage temperature for the other 5 mutants was 23 ^{o}C. The cleavage reaction takes 16-20 h. Proteins were purified by HPLC using a two buffer system. Buffer A was H_2O (0.1% TFA) and buffer B was 90% acetonitrile and 10% H_2O (0.1% TFA). All proteins were eluted at around 48% buffer B and more than 95% pure as judged by analytical HPLC. The yield was 12-16 mg/L LB media. The identities of all proteins were confirmed by matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI-TOF). The observed and expected molecular weights were as follows: DM F47L, observed 4114.6, expected 4115.9; DM F58L, observed 4113.6, expected 4115.9; DM F51L, observed 4115.4, expected 4115.9; DM F47LF51L, observed 4084.9, expected 4081.9; DM F47LF58L, observed 4080.8, expected 4081.9; DM F51LF58L, observed 4081.2, expected 4081.9; DM F47LF51lF58L, observed 4043.7, expected 4047.9; DM W64L, observed 4076.0, expected 4076.9.

2.2.2 NMR

All NMR spectra were acquired on a Varian Instruments Inova 600 MHz spectrometer at 25 ^{o}C with approximately 1 mM protein. Samples were prepared in 90% H_2O, 10% D_2O, 10 mM sodium acetate and 150 mM sodium chloride, at pH 5.0 and referenced against sodium 3-(trimethylsilyl) tetradeuteriopropionate.

2.2.3 Circular dichroism

Wavelength scans were recorded from 260 nm to 195 nm as the average of 5 repeats.

Spectra were collected at 25 ^{o}C with samples of 15 μM to 30 μM protein in 10 mM sodium acetate, and 150 mM sodium chloride at pH 5.0 in a 0.1 cm cuvette. Thermal unfolding experiments were performed from 2 ^{o}C to 96 ^{o}C with a 2 ^{o}C interval and the signals were monitored at 222 nm. A 1.0 cm cuvette was used. The buffer and protein concentration was the same as used in wavelength scan experiments. Thermal unfolding curves were fit to equation 1.10 for two-state folding to obtain T_m and $\Delta H^o(T_m)$. The fit includes terms to describe linear pre and post transition baselines. Urea unfolding curves were collected at 25 ^{o}C and 222 nm using an AVIV Instruments model 202SF CD instrument equipped with an automatic titrator. The concentration of urea was increased from 0 to about 10 M in ~0.25 M steps and was measured by the reflective index. The buffer and protein concentration was the same as used in wavelength scan experiments. Urea unfolding curves were fit to the standard linear extrapolation model in which

$$\Delta G^o \text{ (urea)} = \Delta G^o (H_2O) - m \times [\text{urea}] \qquad (2.1)$$

to obtain m-value and ΔG^o in the absence of denaturant. The fit assumes that the pre and post transition baselines are linear functions of [urea].

2.2.4 Analytical ultracentrifugation

Data were collected with a Beckman Optima XL-A analytical ultracentrifuge at 25 ^{o}C and 280 nm using rotor speeds of 38,000 rpm for 24 hours and 48,000 rpm for another 24 hours. Six channel, 12 mm path length, charcoal-filled Epon cells with quartz windows were used. Samples of 20, 40 and 50 μM protein were prepared in 10 mM sodium acetate, and 150 mM sodium chloride at pH 5.0. The partial specific volume, 0.757 mL/g for double Phe-to-Leu mutants and 0.765 mL/g for triple Phe-to-Leu mutant, and solution density, 1.005 g/mL, were calculated from

the SEDNTERP program. Data were analyzed using the HeteroAnalysis program from the Analytical Ultracentrifugation Facility at the University of Connecticut.

2.3 Results

The N68AK70M double mutant of HP36 was used as the pseudo-wildtype and is designated DM HP36 to distinguish it from the true wildtype. Both N68 and K70 are surface residues and exposed to solvent thus they may be replaced without altering core packing, a feature which has been confirmed by comparison of the high resolution crystal structures of DM HP36 and wildtype HP35 (PDB: 1YRF) (*147*). A set of seven Phe-to-Leu mutants was prepared comprised of the complete set of single mutants, all three double mutants and the triple mutant. The sole Trp was also mutated to Leu. Aromatic to Leu rather than aromatic to Ala mutations were made so that aromatic residues could be removed without deleting significant sidechain volume and thus potentially leading to significant repacking effects. A Phe to Leu substitution does result in a decrease in sidechain volume of approximately 35 $Å^3$. However it is the smallest change associated with mutation to a hydrophobic residue and thus the most conservative (*174*). All of the mutants expressed well and were soluble. CD spectroscopy indicates that all are folded and the far UV CD spectra of all of the mutants are very similar to that recorded for DM HP36 and HP36 (Figure 2-2). DM HP36 contains a set of ring current shifted methyl resonances that are very diagnostic of its fold. Thus ^{1}H NMR spectra provide a good probe of the integrity of the tertiary structure. However, the three buried Phe are responsible for the ring current shifts thus Phe to Leu mutants may exhibit only a subset of the ring current shifted peaks yet still be folded. As expected no ring current shifted methyls were observed in the ^{1}H NMR spectrum of the DM F47LF51LF58L triple mutant however strongly upfield shifted methyl peaks were observed for

each single Phe-to-Leu mutant and for the DM F47LF51L, DM F47LF58L and DM F51LF58L double mutants (Figure 2-3). HP36 contains one Trp residue at position 64. The indole ring is partially exposed to solvent but one face is packed against the body of the protein and the indole N-H resonance appears at the characteristic chemical shift of 10.5 ppm in the folded state. All of the Phe mutants display a single peak for this proton located between 10.4 and 10.5 ppm providing further evidence that the proteins are well folded. In addition, the peaks in the aromatic-amide region of the spectra are well dispersed as expected for a folded protein (Figure 2-3). The ^{1}H NMR spectrum of the W64L mutant displays the characteristic ring current shifted resonances observed in the wildtype indicating that core packing is not disrupted. Previous work reported the characterization of a subset of Phe-to-Leu mutants in a M53L background and showed that the single mutants were monomeric in solution (*160*). Some potential association was detected for the F47LF58L and F51LF58L double mutants in this system using gel filtration, however the M53L background is even less stable than wildtype and thus it was not clear if the hydrodynamic behavior of those mutants represented a rapid equilibrium between the folded and unfolded states or weak self association in their respective native state. Consequently we conducted analytical ultracentrifugation (AUC) studies of each double mutant and the DM F47LF51LF58L triple mutant. Equilibrium AUC gives a molecular weight which is independent on the shape of the protein. Apparent molecular weights were determined from a single species fit of data collection from a 50 μM sample which is the highest concentration among the three samples (Figure 2-4). The apparent molecular masses determined from a single species fit are summarized in Table 2-1. The apparent molecular weights deduced from a single species fit to the AUC data are thus 8 to 20% higher than the true monomeric weights, indicating only weak self associate at worst. The fact that gel filtration data collected for F47LF58L and F51LF58L in

the destabilized M53L background gave apparent molecular weights of ~ 7,500 reflects the fact that the double Phe-to-Leu mutants likely lead to a significant populated of unfolded protein in this system (*160*).

Having established that all of the mutants are folded and well behaved in solution we turned to stability measurements to characterize the effects of the aromatic-to-Leu substitutions. All proteins displayed sigmoidal thermally induced unfolding transitions which can be readily fit to equation 1.12 for two-state unfolding with linear pre- and post- transitions. There are several practical issues which are worth noting. First the DM HP36 pseudo-wildtype is so stable that it is not fully unfolded even at temperatures approaching 100 ^{o}C. A variant with a lower thermal stability could be used as the background for making core mutations, however that leads to the problem that the most destabilizing mutants are not well folded, thus use of DM HP36 is a necessary compromise. Nevertheless the T_m of DM HP36 has been previously estimated from the derivative of the thermal unfolding curve to be 90.6 ^{o}C. A second issue is that small proteins normally have broad unfolding curves because of the small ΔH^o of unfolding. This is the case for HP36. Furthermore mutations which lead to core packing that is less optimal than wildtype should have even smaller ΔH^o's of unfolding. Thus some of the mutant unfolding curves are relatively broad, but again all can be reliably fit to yield T_m and $\Delta H^o(T_m)$. The thermal unfolding curves are shown in Figure 2-5 as plots of CD signal vs. Temperature. We display the data in this way instead of as plots of fraction folded because using fraction folded can disguise problems with baselines, particularly for small proteins. Showing the actual fits gives a more objective view of the data. A more subtle, but more fundamental point is that the analysis is based on a global parameter, mean residue ellipticity, which reports on the overall properties of the molecule. Non-coincidental unfolding curves have been observed for some helical proteins in NMR

monitored unfolding studies even when sigmoidal melting curves are observed using global probes although the basis for the observed effects is extremely controversial (*175, 176*). Unfortunately the unfolding transitions of DM HP36 or its mutants can not be followed throughout the entire unfolding curve by NMR due to a combination of high T_m's, fast exchange between folded and unfolded forms at higher temperature and issues of spectral resolution. Thus we use CD monitored unfolding as a probe of the global integrity of the fold and use the measured T_m's as a semi-quantitative probe of the effects of the mutations on stability.

Each single Phe-to-Leu mutation leads to a less thermal stable subdomain as does the Trp-to-Leu mutation. The F51L mutation has the smallest effect reducing the apparent T_m by approximately 15 ^oC while the F47L or F58L mutants clearly have larger effects reducing the T_m by 21 ^oC and 30 ^oC respectively. The rank order observed here agrees with that observed by McKnight and colleagues in the M53L background (*160*). It is interesting to note that substitution of F58 has a larger effect than substitution of F51 even though F51 is sandwiched by the two other Phe rings. In fact the F51L mutation has the smallest effect on T_m of any of the Phe mutants. The effect of the Trp to Leu mutant on T_m is similar to the effect of F51L confirming that W64 makes stabilizing interactions, as noted previously (*157*), even though it is a surface residue. Precise thermodynamics parameters could not be obtained for the W64 substitution in prior investigations because the mutation leads to incompletely defined native baselines in the background used. In contrast the DM HP36 background leads to well defined baselines for the Trp mutant. The same trends detected in the single mutants are observed in the double mutants. Namely any double mutant which includes the F58L substitution is very destabilizing and the apparent T_m's are reduced by more than 40 ^oC while the T_m of the F47LF51L mutant, i.e. the double mutant which retains Phe58 has a T_m 20 ^oC or more higher than the other double mutants.

The thermodynamics of the unfolding of both of the most destabilized Phe double mutants could not be characterized in earlier studies because of the lower stability of the pseudo-wildtype background used. The data presented here shows unambiguously that the deleted phenylalanines are not required for specifying the fold. Interestingly the triple Phe-to-Leu mutant, DM F47LF51LF58L, is folded and is no less thermal stable than the F47LF58L or F51LF58L double mutants. In fact the T_m of the triple mutant is actually slightly higher than the T_m of F51LF58L double mutant and is comparable to that of the F47LF58L double mutant. This may reflect that the more flexible Leu sidechains can slightly repack in the core of the triple mutant relative to double mutants which contain one Phe. The triple Phe-to-Leu mutant has not been previously characterized and the results are very interesting since they directly demonstrate that the Phe residues are not required for folding nor are they required for the generation of a cooperative (sigmoidal) thermal unfolding transition. Thus while packing among the three phenylalanine is stabilizing, aromatic aromatic interactions are not required for HP36 to fold and Trp-Pro interactions are also not required.

The fits to the thermal unfolding curves also yield estimates of the enthalpy of unfolding at T_m, $\Delta H^o(T_m)$. Since the different proteins have different T_m's, $\Delta H^o(T_m)$ values can not be meaningfully compared to each other. However ΔH^o can be estimated at a common temperature provided ΔCp^o is known. pH dependent studies of wildtype HP36 provide an estimated ΔCp^o of 0.38 kcal mol^{-1} deg^{-1} while analysis of the ΔH^o of surface mutants gives an estimated ΔCp^o of 0.23 kcal mol^{-1} deg^{-1} (*147*). Clearly there is considerable uncertainty in the estimation of ΔCp^o. Furthermore there is no reason that the value of ΔCp^o for the Phe to Leu mutants need be exactly the same. For example, the Phe to Leu mutants might alter residual structure in the unfolded state which could lead to change in ΔCp^o even in the absence of native state effects (*5*). Table 2-2

summarizes the apparent ΔH^o values at 50 ^oC, a temperature chosen because it involves a shorter extrapolation than 25 ^oC or 37 ^oC and hence reduces any potential systematic error associated with uncertainty in ΔCp^o. For the reasons listed above we are hesitant to over interpret the ΔH^o values, however we note that replacement of Phe58 has the largest effect among the single mutants and the double mutants which include F58L also show a larger effect than the double mutant which does not. Again the importance of F58 is highlighted. The conclusion that the F58L mutant has the largest effect on ΔH^o is independent of the choice of ΔCp^o over the range of values considered. The Trp64-to-Leu mutation has the smallest effect upon ΔH^o which is not surprising given that it lies on the surface of the protein.

We also conducted urea induced unfolding studies of all of the mutants (Figure 2-6). Some of the issues that were relevant for the thermal unfolding come to play here as well. First the DM HP36 protein is too stable to give a complete urea unfolding curve, but stability can be estimated using the midpoint, C_M, determined from the derivative of the curve and the wildtype m-value. In addition, the stability of this protein has been examined previously by amide H/D exchange (*147*). All of the mutants give complete unfolding curves however the pretransition baselines are not as optimally defined for the two most destabilizing double mutants or the triple mutant although clear sigmoidal curves are measured. We note this is part of the motivation for using DM HP36 as the background. A less stable "wildtype" would lead to very poorly defined transitions for the destabilizing mutants. This has plagued all prior attempts to analyze the role of the aromatic residues and prevented a complete description of the role of the aromatic triad. The parameters deduced from the urea unfolding, C_M values, apparent ΔG^o values in the absence of urea and m-values, are listed in Table 2-2 which also summarizes the thermal unfolding data. Of the single Phe mutants F47L or F51L have much smaller effects than does F58L. The effect of

the W64L mutation is also much smaller than the F58L substitution. This is entirely consistent with the thermal unfolding data. Fits to the urea unfolding curves also provide m-values which are formerly equal to $\frac{\partial \Delta G^{o}}{\partial\ [urea]}$. M-values are generally believed to be related to the change in solvent accessible surface area between the folded and unfolded forms and changes in m-values have been proposed to reflect, at least in some cases, changes in the unfolded state ensemble (*32, 91, 177*). The m-value for wildtype HP36 has been determined to be 0.52 kcal mol^{-1} M^{-1} under these conditions but the m-value for the hyperstable DM HP36 could not be determined because the protein is too stable. The m-values of the triple mutant and F47LF58L, F51LF58L double mutants are similar but are all smaller than the m-value for the F47LF51L double mutant which retains F58. The m-value of DM W64L is similar to wildtype. M-values are less reliably extracted from fitting broad unfolding curves and thus we avoid attempting a molecular level interpretation of the changes. In addition, recent theoretical work suggests that urea induced unfolding may be more complex than often assumed (*178*).

Analysis of the urea induced unfolding of the double mutants yields conclusions which are similar to those obtained from the thermal unfolding, namely double mutants which remove F58 are drastically destabilized relative to the double mutant which retains F58. In addition the triple Phe to Leu mutant is no less stable than either of the double mutants which lack F58. In fact it is slightly more stable. Mutations can also affect residual interactions in the denatured state ensemble and contribute to changes in stability via denatured state effects. This may be relevant for HP36 since studies with peptide fragments provide evidence for hydrophobic clusters in the unfolded state ensemble which involves some of the Phe residues and which can be partially disrupted by mutation to Leu (*155, 156*). How might such effects impact the interpretation of the experimentally determined changes in T_m and ΔG^{o}? Clearly the conclusion that interactions

involving the aromatic residues are not required for folding is not altered. Transiently populated hydrophobic clustering in the unfolded state is expected to be relatively weak energetically, but it should stabilize the unfolded state ensemble. Thus disrupting such interactions in the absence of any native state effects would stabilize the protein because the free energy of the unfolded state would be increased. The fact that all of the mutations studied here significantly destabilize the protein argues that native state effects dominate. In addition it is worth noting that the same rank order of the effect of single Phe mutations is observed in the DM background, in the single Phe mutant backgrounds, and in the M53L background analyzed previously, strongly suggests that native state effects are dominate and the rank ordering of stability is robust.

Analysis of $\Delta\Delta G^o$ values for the single and double mutants or examination of ΔT_m values provides a measure of the effect of mutation(s) upon stability, but does not provide direct evidence for energetically significant interaction between the aromatic residues. Double mutant cycle analysis was developed to provide a method for measuring the interaction free energy, ΔG_{int}, between two residues by eliminating the effects of mutations upon other interactions (*179*).

$$\Delta G_{int}(A,B)_F - \Delta G_{int}(A,B)_U = \Delta G^o_{WT} - \Delta G^o_A - \Delta G^o_B + \Delta G^o_{AB} \qquad (2.2)$$

where $\Delta G_{int}(A,B)_F$ is the folded state interaction free energy between residue A and B, $\Delta G_{int}(A,B)_U$ is the unfolded state interaction free energy between those two, ΔG^o_{WT} is the free energy of unfolding of the wildtype, ΔG^o_A and ΔG^o_B are the free energies of unfolding of the respective single mutants, and ΔG^o_{AB} is the unfolding free energy of the double mutant. When the analysis of double mutant cycles makes the inescapable assumption that the mutations do not alter the free energy of the unfolded ensemble, double mutant cycles can be analyzed in the conceptual context of a double mutant cycle “square” or thermodynamic cycle. However, a thermodynamic cube is required if unfolded state effects are present. In this case, knowledge of

changes in the free energy of both the folded states and unfolded states of the three mutants is needed and this is generally impossible to obtain. This effect may be important for HP36 since the subdomain appears to have a partially structured denatured state ensemble in which the aromatic residues may be involved in hydrophobic clusters (*5, 155, 156*). Another important issue is that the double mutant cycle cancels desolvation effects to first order (*61*). None-the-less the approach is still powerful and in the present case should be able to deduce which pairs of aromatic residues interact most strongly. Table 2-3 summarizes the experimental interaction free energies determined using ΔG^o values from urea denaturation studies. The interaction free energies between F47 and F58 and between F51 and F58 are both small while the interaction energy between F47 and F51 is large. The same result is obtained if the interaction free energy between F47 and F51 is measured in DM F58L background. The fact that a self consistent rank order is obtained independent of the background used argues that, while unfolded state effects may play a role, the qualitative conclusion that the F47-F51 interaction is stronger than the native state F47-F58 or F51-F58 interactions is likely to be robust.

2.4 Discussion

By using a hyperstable double mutant which retains the same fold and core packing as the wildtype subdomain we have been able to define the role of the core phenylalanine residues in HP36 and to probe the role of a conserved surface aromatic residue. Previous attempts to do so have been hindered by the low stability of the wildtype subdomain. Although mutation of any one of the core phenylalanine results in a decrease in stability none are critical for folding and even a triple Phe to Leu mutant folds. The W64L mutant confirms the important role proposed for this residue, but indicates that aromatic proline interaction is not strictly required to specify

the fold (*157, 173*). A more detailed analysis of potential Trp-Pro interactions is given in chapter 3. F58 is consistently found to play a larger role than either F47 or F51 both in the content of single and double mutants. These conclusions are independent of whether thermal or urea unfolding data is used to assess stability.

It is interesting to note that mutation of F58 clearly has a larger effect than mutation of F51 even though F51 is the central residue in the triad of phenylalanines. This likely reflects the fact that F58 makes contacts with a larger number of other residues including F47, F51, M53, A57, A59, L61, Q66 and M70 while F51 is involved in long range contacts with a smaller number of residues, F47, V50, G52, F58, M70 and K73. The double mutant cycle analysis implies a significant favorable interaction between F47 and F51. Interaction free energies involving F58 and the other phenylalanines are much smaller even though mutation of F58 has the largest effect upon stability. Again, this likely reflects the fact that F58 makes more extensive contacts with other hydrophobic core residues than with the F47 or F51. The important role of F58 is broadly consistent with some MD simulations which have examined the folding of HP36 (*149*). F58 is located near the N-terminus of helix-2 and make contacts with both helix-1 and helix-3. A potential folding scenario derived from one MD study is that helix-1 folds after helix-2. In this case a correctly positioned F58 in helix-2 will help to stabilize helix-1.

The data presented here clearly demonstrates that aromatic proline, aromatic hydrophobic and aromatic aromatic interactions, while important for stability are not required to specify the overall fold of the villin headpiece helical subdomain. Many, but not all, small cooperatively folded domains exhibit aromatic aromatic or aromatic proline interactions and this has lead to some speculation that they may be critical determinants of “mini-protein” folding (*161*). Clearly this is not the case for HP36. It is interesting to speculate if unusual interactions are generally

required to stabilize minature protein domains. We believe that this is likely not the case at least for proteins the size of HP36 or larger since there are several examples, including HP36, where aromatic aromatic or aromatic proline interactions are not critical for specifying the fold (*180, 181*). Important early studies on protein thermodynamics led to the suggestion that proteins needed to exceed a minimum size, estimated to be 50 residues, to be cooperatively folded (*182*). There are obviously a number of examples of proteins that fold cooperatively yet are much smaller and this has promoted the idea that small cooperatively folded proteins require special interactions. However, it is important to remember that the early thermodynamic analysis was designed to calculate the minimum size required so that ΔG^o of folding was significantly larger than RT. In fact all minature proteins characterized to data have ΔG^o values which are only several times RT (*183*). Furthermore their thermodynamic properties are well predicted by data bases derived from larger proteins (*182, 184*). Thus there is no inherent contradiction between the early thermodynamic analysis and the fact the proteins of forty residues or less can be found which fold cooperatively. Of course we do not want to imply that aromatic aromatic or aromatic proline interactions may not be required in specific cases, but they are clearly not mandatory in all cases.

Table 2-1. Comparison of theoretically calculated and apparent molecular weights from AUC data for the three double mutants and one triple mutant. The uncertainties were standard errors to the fit.

Mutant	Theoretical MW	Observed MW
DM F47LF51L	4082	4920 ± 96
DM F51LF58L	4082	4360 ± 79
DM F47LF58L	4082	4640 ± 133
DM F47LF51LF58L	4048	4840 ± 64

Table 2-2. Thermodynamic parameters for the unfolding of DM HP36 mutants. Data were collected at 25 ^{o}C pH 5.0 in 10 mM sodium acetate, 150 mM sodium chloride.

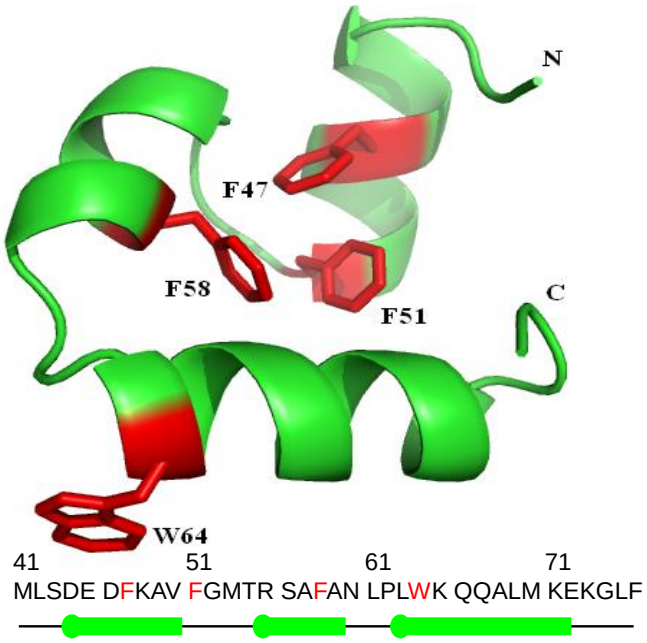
Protein	T_m (^{o}C)	$\Delta H^{o}(T_m)$ (kcal/mol)	ΔH (50^{o}C) (kcal/mol) ΔCp^{o}=0.23	ΔH (50^{o}C) (kcal/mol) ΔCp^{o}=0.38	ΔG^{o} (kcal/mol)	m-value (kcal/mol·M)	C_M [5] (M)
DM HP36	90.6 [1]	----- [2]	----- [2]	----- [2]	4.94-5.06 [3]	----- [4]	9.60
DM F47L	69.6	25.9	21.4	18.4	2.85	0.46	6.00
DM F51L	76.0	28.4	22.4	18.5	3.01	0.44	7.00
DM F58L	60.9	21.3	18.8	17.2	2.62	0.55	4.75
DM F47LF51L	68.1	23.2	19.0	16.3	2.54	0.40	6.50
DM F47LF58L	48.9	13.3	13.6	13.7	0.75	0.34	2.85
DM F51LF58L	46.4	15.6	16.4	17.0	0.72	0.29	3.30
DM F47LF51LF58L	49.8	13.3	13.3	13.4	1.02	0.34	3.50
DM W64L	74.4	36.6	31.0	27.3	3.58	0.52	6.75

(1) Determined from the derivative of the curve of the CD signal vs. temperature.
(2) Not determined owing to an incomplete thermal unfolding transition.
(3) Calculated from the C_M value determined by numerical differentiation of the curve together with the wildtype m-value. Data for DM H36 is from reference (*147*).
(4) Not determined owing to an incomplete urea unfolding transition.
(5) Determined by numerical differentiation of the plot of CD signal vs. [urea].

Table 2-3. Double mutant cycle analysis of apparent interaction free energies. ΔG^o values from the urea unfolding studies were used. Experimental data were collected at 25 ^oC pH 5.0 in 10 mM sodium acetate, 150 mM sodium chloride.

Background	$\Delta\Delta G^o$ (F47,F51) (kcal/mol)	$\Delta\Delta G^o$ (F47,F58) (kcal/mol)	$\Delta\Delta G^o$ (F51,F58) (kcal/mol)
DM HP36	1.68	0.28	0.09
DM F47L	-----	-----	0.58
DM F51L	-----	0.77	-----
DM F58L	2.17	-----	-----

a)



b)

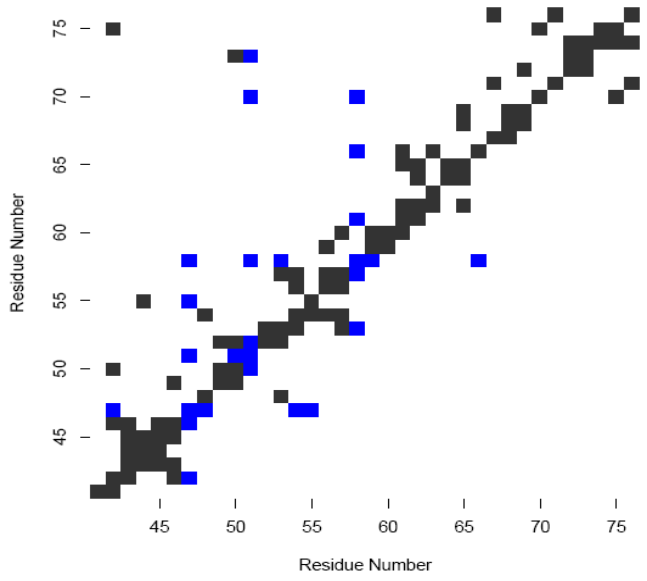


Figure 2-1. (a) A ribbon diagram of the structure of DM HP36 generated using PyMol, version 0.99. The N and C-termini are labeled and the Phe hydrophobic core residues and the surface Trp residue are shown in stick format and are labeled. The primary sequence is shown together with a diagram of secondary structures. α-helices are represented as cylinders. (b) A contact map for DM HP36 sidechain sidechain contacts are shown above the diagonal and sidechain mainchain contacts below the diagonal. Contacts involving the three Phe residues are colored blue. A 6Å cutoff was used.

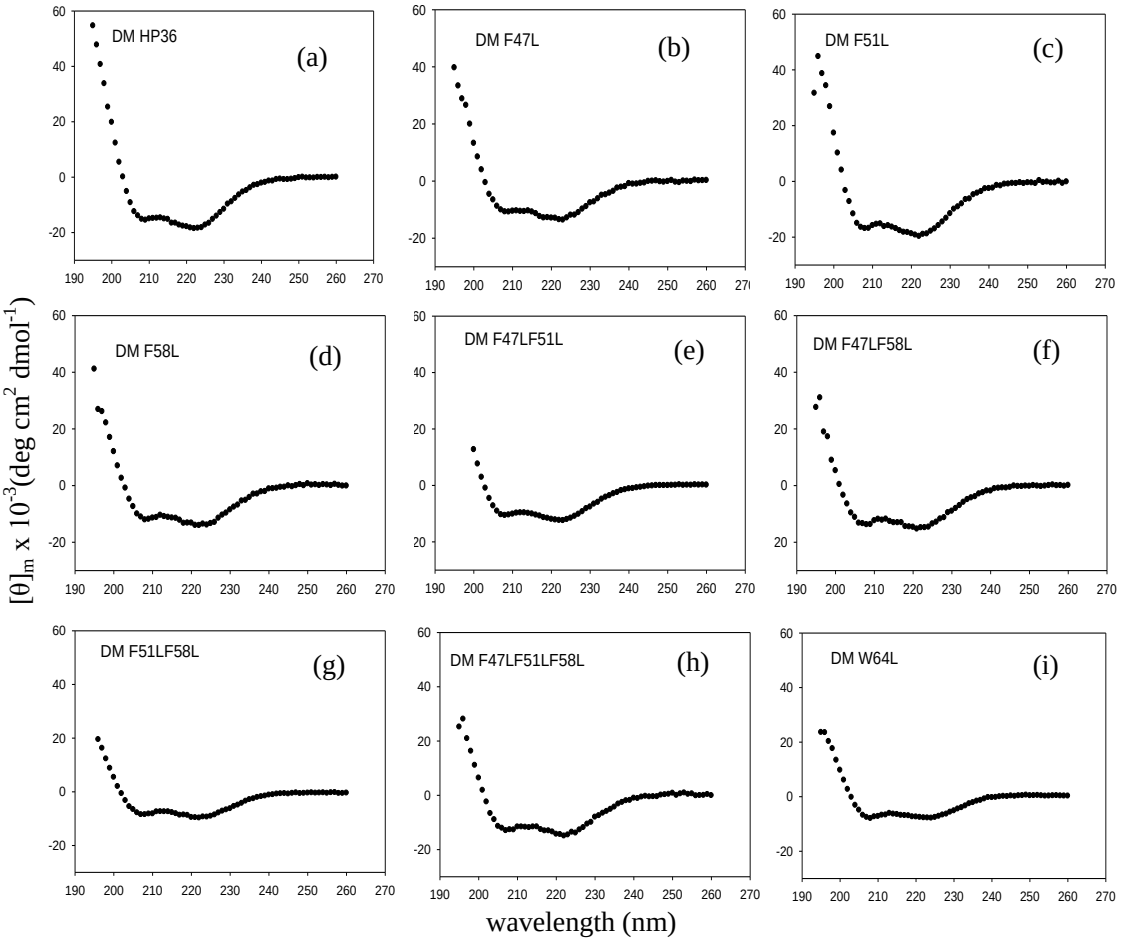
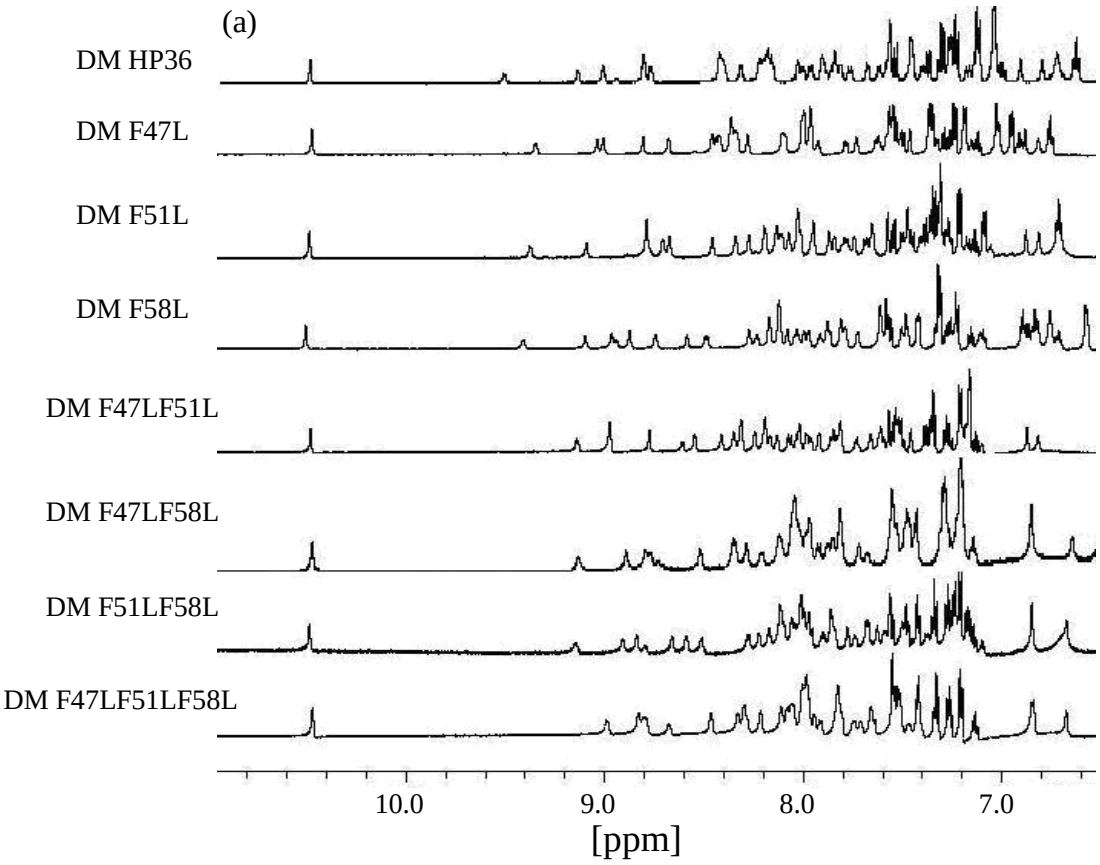


Figure 2-2. Far UV CD spectra show all mutants are folded. (a) DM HP36, (b) DM F47L, (c) DM F51L, (d) DM F58L, (e) DM F47LF51L, (f) DM F47LF58L, (g) DM F51LF58L, (h) DM F47LF51LF58L, (i) DM W64L. All spectra were collected at 25 ^{o}C in 10 mM sodium acetate 150 mM sodium chloride, pH 5.0.


(a)
DM HP36
DM F47L
DM F51L
DM F58L
DM F47LF51L
DM F47LF58L
DM F51LF58L
DM F47LF51LF58L
10.0
9.0
8.0
7.0
[ppm]

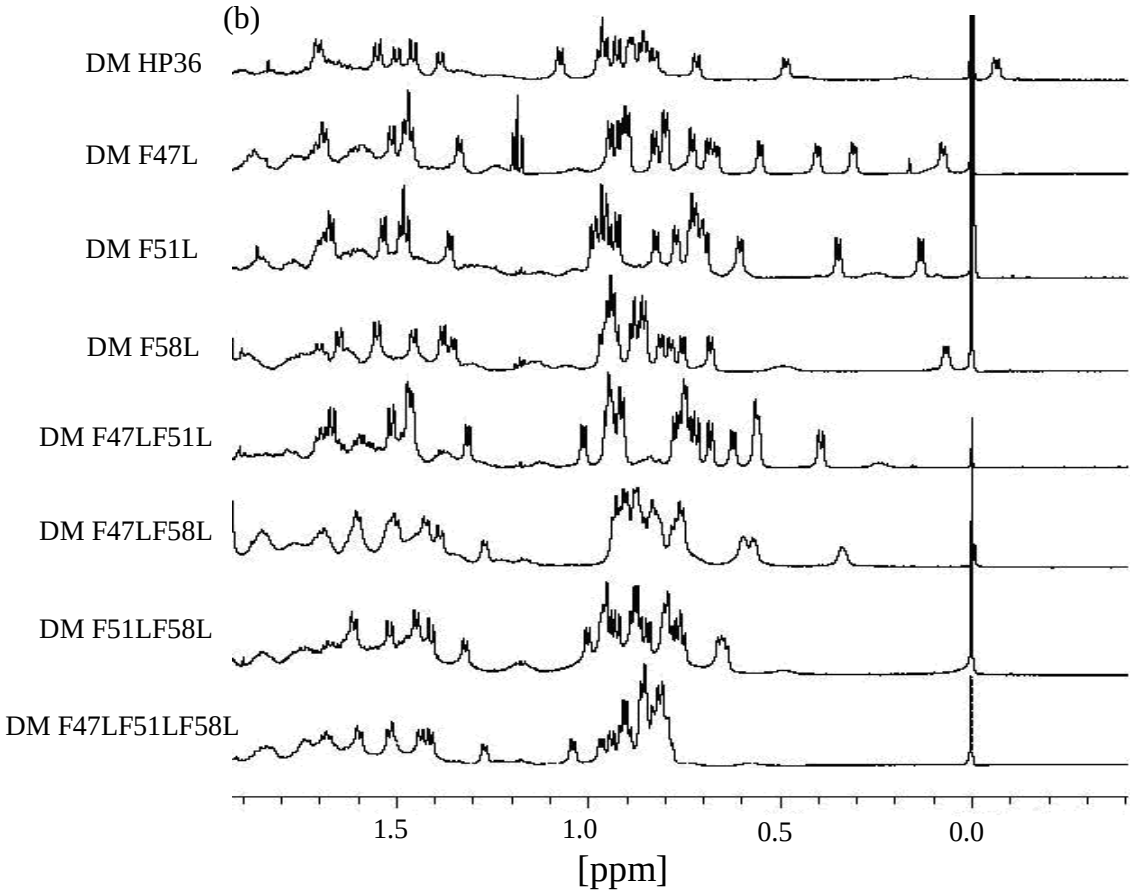


Figure 2-3. ^{1}H NMR spectra confirm that all mutants are folded. (a) Stacked plot of the downfield region of the spectra of the DM HP36 mutants. Spectra are labeled to denote which mutant they correspond to. (b) Stacked plot of the upfield region of the spectra of DM HP36 mutants. The sharp peak at 0.00 ppm is the chemical shift standard. Spectra were recorded at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate 150 mM sodium chloride, pH 5.0.

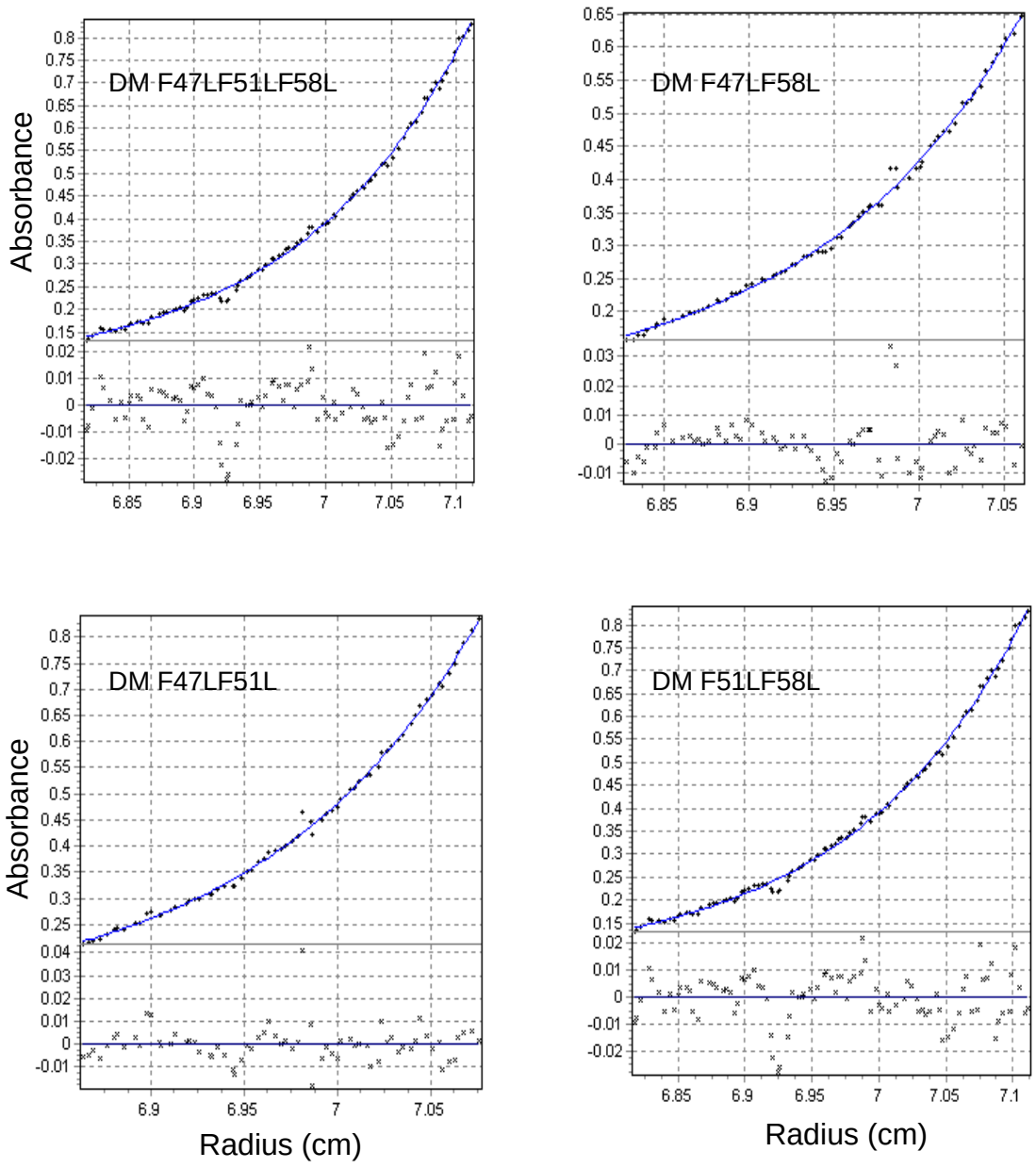


Figure 2-4. Analytical ultracentrifugation data and analysis plots of absorbance vs. radial position in the cell together with best fits to a single species model. The experiments were done at 25 ^{o}C in 10 mM sodium acetate 150 mM sodium chloride pH 5.0 buffer. The residuals are shown below each fit.

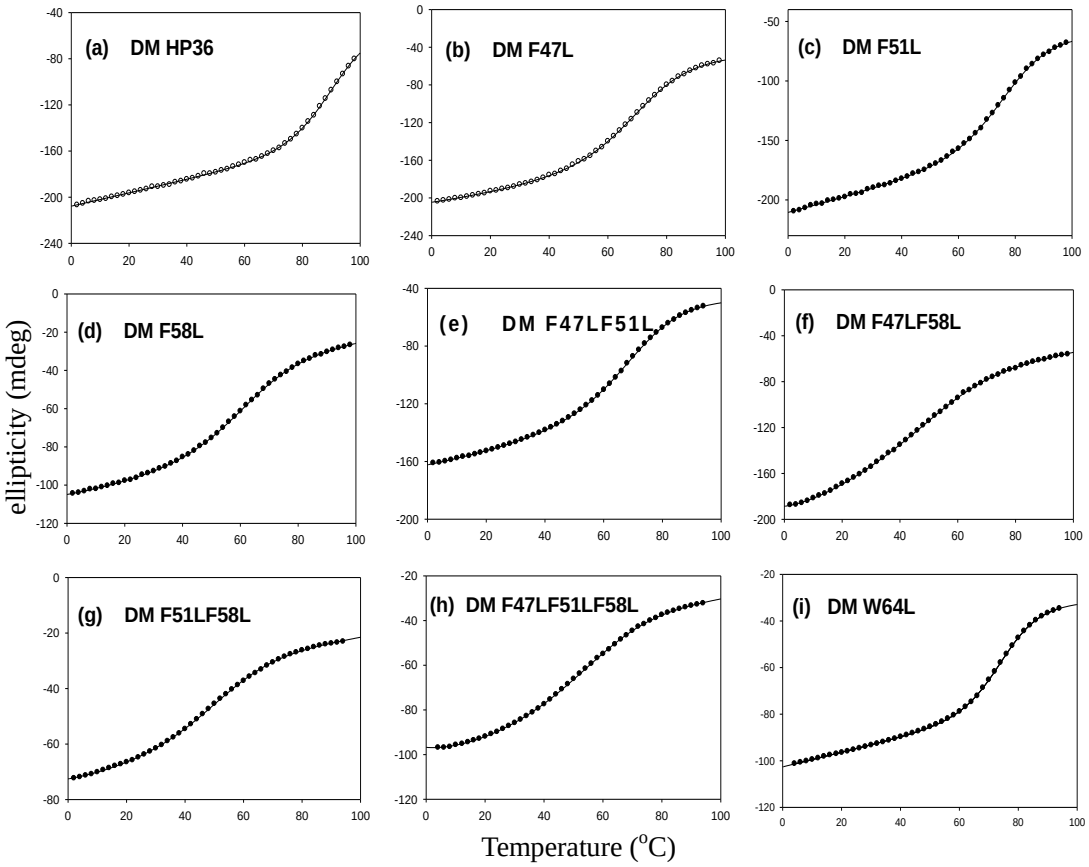


Figure 2-5. CD monitored thermal unfolding curves for DM HP36 and its mutants. (a) DM HP36, (b) DM F47L, (c) DM F51L, (d) DM F58L, (e) DM F47LF51L, (f) DM F47LF58L, (g) DM F51LF58L, (h) DM F47LF51LF58L, (i) DM W64L. Samples were between 15 μM and 30 μM protein in 10 mM sodium acetate 150 mM sodium chloride, pH 5.0 in a 1.0 cm cuvette. The CD signal at 222 nm was detected.

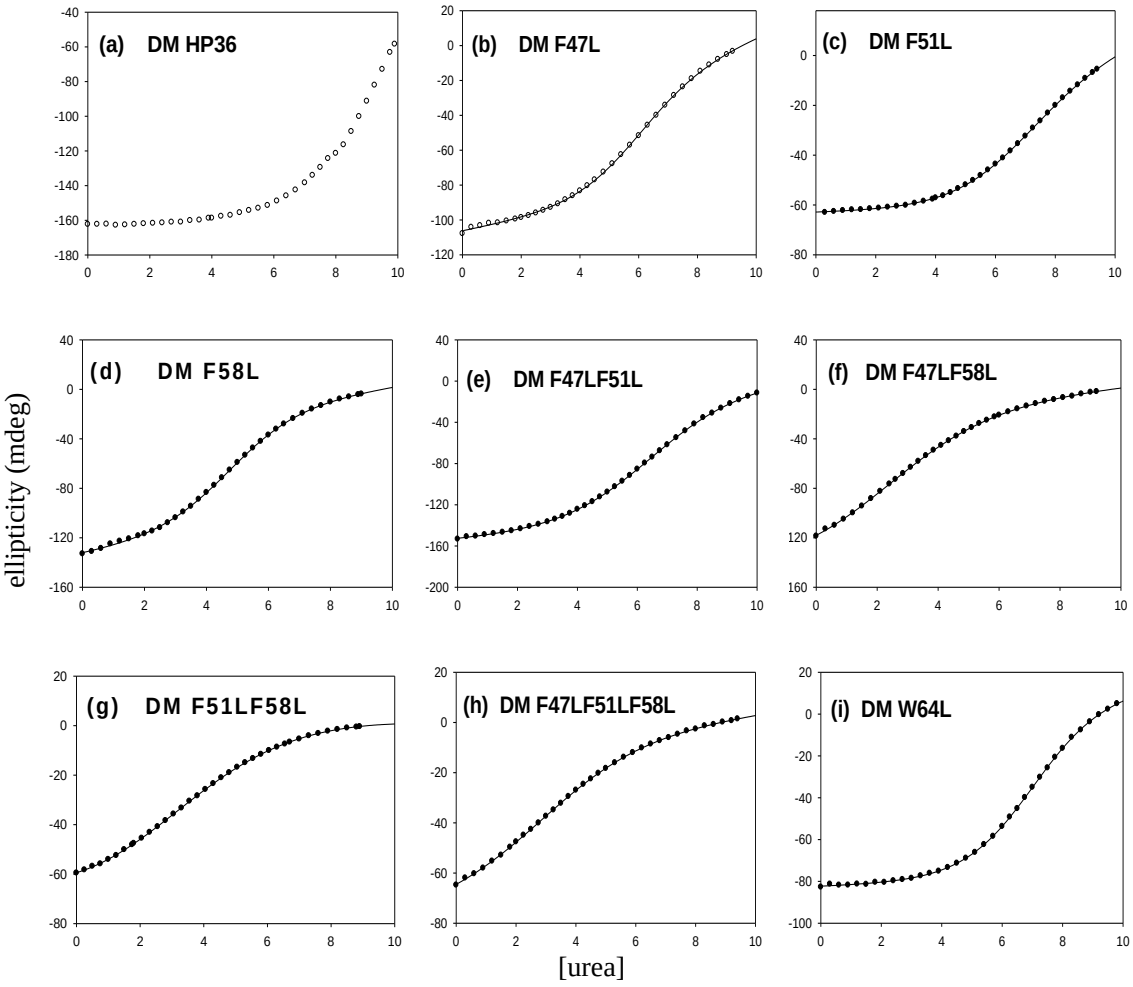


Figure 2-6. CD monitored urea induced unfolding curves for DM HP36 and its mutants. (a) DM HP36, (b) DM F47L, (c) DM F51L, (d) DM F58L, (e) DM F47LF51L, (f) DM F47LF58L, (g) DM F51LF58L, (h) DM F47LF51LF58L, (i) DM W64L. Samples were between 15 μM and 30 μM protein in 10 mM sodium acetate 150 mM sodium chloride, pH 5.0. The CD signal at 222 nm was detected.

3. A critical assessment of putative gatekeeper interactions in the villin headpiece helical subdomain

Abstract

Aromatic proline interactions involving Trp64 and Pro62 have been proposed to play a critical role in specifying the fold of the villin headpiece helical subdomain by acting as gatekeeper residues. Gatekeeper residues are defined as strategically situated residues which play a critical role in specifying the protein fold. Note that the numbering corresponds to full length headpiece. Mutation of Pro62 has been shown to lead to a protein which does not fold, but this may arise for two different reasons: The residue may make interactions which are critical for the specificity of the fold or the mutation may simply destabilize the domain. In the first case the protein cannot fold, while in the second the small fraction of molecules that fold adopt the correct structure. The modest stability of the wildtype prevents a critical analysis of these interactions because even moderately destabilizing mutations lead to a very small folded state population. Using a hyperstable variant of HP36, denoted DM HP36, as our new wildtype, we characterized a set of mutants designed to assess the role of the putative gatekeeper interactions. Four single mutants DM P62A, DM W64L, DM W64K, DM W64A and a double mutant DM P62AW64L were prepared. All mutants are less stable than DM HP36, but all are well folded as judged by CD and ^{1}H NMR. All of the mutants display sigmoidal thermal unfolding and urea induced unfolding curves. Double mutant cycle analysis shows that the interactions between Pro62 and Trp64 are weak but favorable. Interactions involving Pro62 and proline aromatic interactions are, thus, not required for specifying the subdomain fold. The implications for the design and thermodynamics of miniature proteins are discussed.

Acknowledgements

The data presented in this chapter has been published (Xiao, S., and Raleigh, D.P. (2010) *J. Mol. Biol.* 401, 274-285). This chapter contains direct excerpts from the manuscript with a few adjustments. I thank Dr. Bing Shan and Mr. Wenli Meng for helpful discussions. I also thank Dr. Yuan Bi and Dr. Jae-Hyun Cho for their development of the expression system.

3.1 Introduction

The investigation of protein folding remains an active area of research, driven in part by the realization that incorrect folding can play a key role in protein aggregation *in vitro* and *in vivo* (*2, 185-188*). In recent years much attention has been directed at small, rapid folding, globular proteins because they are attractive targets for molecular dynamics simulations and theoretical approaches. These studies offer the possibility of directly comparing simulations and experiments. Small stably folded domains are also of interest because they can be used to present recognition surfaces in a defined structural context (*189-191*). There is a great deal of interest in deducing the interactions which stabilize small domains and HP36 has received considerable attention because of its wide use as a model system for folding (*161-163, 183, 189-192*). Elucidating the interactions which stabilize "mini" proteins is also important for the rational design of stable protein scaffolds (*161, 183, 189, 190, 192-194*). The subdomain has a simple topology consisting of three α-helices which form a well packed hydrophobic core (Figure 3-1). The equilibrium folding transitions of HP36 appear to be cooperative and kinetic studies have shown that the subdomain folds on the microsecond time scale, making it one of the fastest folding proteins reported (*5, 115, 134, 195*). Unfortunately, the domain, like most small proteins, has only a modest stability in terms of its unfolding free energy, ΔG^{o}, although it has a high T_m (*172, 184, 196*). One consequence of its modest stability is that it can be difficult to probe the role of sidechains in stability and folding because even seemingly conservative mutations can destabilize the protein so that it does not fold. This leads to an inherent ambiguity in the analysis of mutational studies, namely a mutant which prevents the protein from folding may do so by preventing the protein from adopting the correct fold, or by simply destabilizing the domain enough so that the population of folded molecules is small. In the first case the protein cannot

fold because the interaction(s) affected by mutation were essential for specifying the fold and may result in the misfolding, while in the second, those molecules which do fold adopt the correct native structure. In the first case the issue is the specificity of the fold and in the second it is stability (*172*).

There have been a number of suggestions that special interactions may be required to stabilize small proteins and both aromatic aromatic and aromatic proline interactions have been proposed to play a special role. Mutational and computational studies have been conducted of a potential "gatekeeper" interaction in HP36 involving Trp64 and Pro62. Again, note that residues in HP36 are normally numbered according to their position in the complete headpiece domain. The additional Met which results from recombinant expression is denoted residue 41. Pro62 and Trp64 are part of a strongly conserved cluster of solvent exposed residues which have been proposed to be critical for stability and for Actin binding. As shown in Figure 1-4, position 62 is a Pro in 22 out of 29 known sequences and a Ala in the other 7 sequences, while Trp64 is completely conserved. Aromatic proline interactions have been suggested to play a key role in stabilizing a number of proteins and are known to be energetically favorable (*152, 161, 162, 164*).

Pro62 in HP36 has been proposed to be a critical gatekeeper residue, in part because of its interaction with Trp64. Gatekeepers are defined as strategically situated residues which play a critical role in specifying the fold and in stability (*152, 157*). Previous work has shown that mutation of Pro62 to Ala in the villin headpiece subdomain drastically destabilized the subdomain and NMR studies indicated that the mutant did not fold, while FTIR provided evidence that it had a tendency to aggregate (*157, 195*). A similar substitution in the corresponding subdomain from Advillin notably destabilized the protein, although it could still

fold. Conversely, substitution of Pro for the naturally occurring Ala in the helical subdomain from Dematin lead to a stably folded protein while the wildtype Dematin subdomain sequence, (with Ala at position 62), did not fold in isolation. Taken together, these observations led to the proposal that Pro62 acts as a critical gatekeeper and is essential for the proper folding of the headpiece helical subdomain (*157*). Some molecular dynamics simulations have also suggested an important role for this region in the folding of helix-2 and helix-3, which has been predicated to be a key step in the folding of the subdomain (*149, 197*).

Unfortunately, the modest stability of HP36 makes it extremely difficult, if not impossible, to deduce if Pro62 and the Pro62 Trp64 interaction are essential for specifying the fold or simply contribute to stability. *Ab initio* simulations have been used to predict the energetic consequences of substitutions at these sites, but the only available experimental thermodynamic data are reported midpoints of thermal denaturations, T_m, for some variants of the villin family (*152*). Changes in T_m do not have to correlate linearly with changes in ΔG^o of unfolding at a given temperature, and large changes in T_m can correspond to modest changes in ΔG^o for small proteins. Thus, it is difficult to judge the accuracy of the calculations (*184*).

Here we use a hyperstable variant of HP36, the same double mutant Asn68Ala Lys70Met described in chapter 2, to critically probe the effects of mutations of Pro62 and Trp64. We show that while Pro62 and the Pro62 Trp64 interaction contribute to stability they are not required to specify the structure, nor are they required to ensure cooperative folding.

3.2 Materials and methods

3.2.1 Protein expression and purification

General aspects of protein expression were described in chapter 2. The factor Xa cleavage

was carried out at 23 ^{o}C for 16 to 20 hours. Protein purification was also described in chapter 2. The HPLC elution time corresponded to 49% buffer B for DM P62A, 47% for DM W64L, 45% for DM W64K, 45% for DM W64A, and 43% for DM P62AW64L. All proteins were more than 95% pure as judged by analytical HPLC. The identities of all proteins were confirmed by matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI-TOF). The observed and expected molecular weights are as follows: DM P62A, observed 4121, expected 4124; DM W64L observed 4076, expected 4077; DM W64K, observed 4091, expected 4091; DM W64A, observed 4035, expected 4035; DM P62AW64L, observed 4051, expected 4051.

3.2.2 Nuclear magnetic resonance (NMR) spectroscopy

All NMR spectra were acquired on a Varian Instruments Inova 600 MHz spectrometer at 25 ^{o}C using approximate 1 mM protein samples. The DM P62AW64L mutant was also examined at 100 μM. Samples were prepared in both H_2O buffer and D_2O buffer which contained 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0 and referenced against sodium 3-(trimethylsilyl) tetradeuteriopropionate. Samples in D_2O were at pH* 5.0 where pH* indicates the uncorrected meter reading in D_2O.

3.2.3 Analytical Ultracentrifugation (AUC)

Data were collected with a Beckman Optima XL-A analytical ultracentrifuge at 298 K, using rotor speeds of 38,000 rpm for 24 hours and 48,000 rpm for another 24 hours. Six channel, 12 mm path length, charcoal-filled Epon cells with quartz windows were used. UV absorbance at 280 nm was used for DM P62A, 220 nm for DM W64A and 230 nm for DM W64L, DM W64K

and DM P62AW64L. Samples of 20, 50 and 90 μM were prepared in 10 mM sodium acetate, and 150 mM sodium chloride at pH 5.0. The partial specific volume and solution density were calculated from the SEDNTERP program. The partial specific volume is 0.7497 mL/g for DM P62A, 0.7504 for DM W64A, 0.7547 for DM W64L, 0.7527 for DM W64K, and 0.7543 for DM P62AW64L. The buffer solution density is 1.005 g/mL. Data were analyzed using the HeteroAnalysis program from the Analytical Ultracentrifugation Facility at the University of Connecticut (Storrs, CT).

3.2.4 Circular Dichroism (CD) spectroscopy

Wavelength scans were recorded on an Applied Photophysics Chirascan CD instrument over the range of 260 nm to 195 nm as the average of 5 repeats. Spectra were collected at 298 K with samples of 15 μM to 30 μM protein in 10 mM sodium acetate, and 150 mM sodium chloride at pH 5.0 using a 0.1 cm cuvette.

Thermal unfolding experiments were performed over the range of 2 ^{o}C to 94 ^{o}C with a 2 ^{o}C interval and the signal was monitored at 222 nm. The buffer and protein concentration were the same as used in the wavelength scan experiments. The reversibility was confirmed by comparing the initial CD signal at the start of the run to the signal measured after the run was completed and the sample was cooled to the starting temperature. All thermal unfolding curves were analyzed by non-linear least squares fitting using SigmaPlot. Data was fit to:

$$y(T) = \frac{a_n + b_n T + (a_d + b_d T) e^{-\Delta G^\circ_{F-U}(T)/RT}}{1 + e^{-\Delta G^\circ_{F-U}(T)/RT}} \qquad (3.1)$$

where y(T) is the observed signal, a_n and b_n represent the intercept and slope of the pre-transition. a_d and b_d represent the intercept and slope of the post-transition. $\Delta G^\circ_{F\text{-}U}(T)$ is given by the Gibbs-Helmholtz equation:

$$\Delta G^{\circ}_{F\text{-}U}(T) = \Delta H^{\circ}(T_m) - T\Delta S^{\circ}(T_m) + \Delta C_p^{\circ}\left(T - T_m - T\ln\frac{T}{T_m}\right) \quad (3.2)$$

where T_m is the midpoint of the thermal unfolding transition.

Urea unfolding data was collected at 298 K and 222 nm on an AVIV Instruments model 202SF CD instrument. The concentration of urea was increased from 0 to about 10 M in ~0.25 M steps, which were determined by measuring the reflective index. The buffer and protein concentration were the same as used in the wavelength scans. Urea unfolding curves were analyzed by a non-linear least squares fit to equation 3.3:

$$y([urea]) = \frac{a_n + b_n[urea] + (a_d + b_d[urea])e^{-\Delta G^{\circ}_{F-U}([urea])/RT}}{1 + e^{-\Delta G^{\circ}_{F-U}([urea])/RT}} \quad (3.3)$$

where y([urea]) is the observed signal, a_n and b_n represent the intercept and slope of the pre-transition. a_d and b_d represent the intercept and slope of the post-transition. $\Delta G^{\circ}_{F\text{-}U}([urea])$ is assumed to be linearly proportional to the concentration of urea:

$$\Delta G^{\circ}_{F\text{-}U}([urea]) = \Delta G^{\circ}_{F\text{-}U}(H_2O) - m \times [urea] \quad (3.4)$$

3.3 Results

3.3.1 Design of mutants to probe the Pro62 Trp64 interactions

Pro62 is located immediately before the start of the third helix of HP36 and Trp64 is within the helix. The two residues lie on the surface of the subdomain and the proline ring stacks against the indole ring of Trp64 which is partially exposed to solvent (Figure 3-1). A set of mutants were prepared in which Trp64 was changed to Leu, to Lys, and to Ala. The Leu substitution was described in chapter 2 and destabilizes the subdomain. The Trp-to-Leu mutation still allows for the possibility of hydrophobic interactions between residue 64 and Proline 62, but eliminates the aromatic proline interaction. Thus, it allows us to test the importance of aromatic proline

stacking. The Ala mutation is obviously more drastic and eliminates the possibility of sidechain sidechain interactions between the two sites. The Lys mutant contains the hydrophobic four-carbon polymethylene chain, allowing for hydrophobic interactions with Pro62, but it is a different shape than Leu and introduces an additional surface charge. Comparison of the Leu and Lys mutants provides a useful control for any potential effects due to surface hydrophobicity. Pro62 was mutated to Ala and a Pro62Ala Trp64Leu double mutant, DM P62AW64L, was prepared for use in double mutant cycle analysis.

3.3.2 All of the mutants adopt the wildtype fold

Far UV CD spectra of DM HP36 and the various mutants have very similar shapes which are consistent with helical structure (Figure 3-2). We display the data as plots of mdeg vs. wavelength instead of plots of mean residue ellipticity because there is only one Trp and no Tyr in HP36. Hence, there is a large uncertainty in the determination of the protein concentration for the Trp mutants. The NMR spectra of all of the mutants exhibit a well resolved amide ^{1}H region as well as a number of ring current shifted methyl peaks (Figure 3-3). The wide chemical shift dispersion provides additional evidence that each mutant folds to a compact globular structure. The peaks for the single point mutants are all sharp, consistent with them being monomeric. The double mutant, DM P62AW64L, exhibits slightly broader but, nonetheless, well resolved peaks. The modest increase in line widths for the P62AW64L double mutant suggests some self association at the high concentration used for the NMR study, (1 mM). Analytical ultracentrifugation (AUC) data indicates that it is monomeric at 20 μM (Figure 3-4) and NMR spectrum recorded at 100 μM has narrower peaks. The characteristic ring current shifted resonances of HP36 are observed in all of the mutants (Figure 3-3). The negative peak is from

the methyl protons of Val50 due to its interaction with the core aromatic residues. The C4 proton of Phe47 is also strongly shifted in the spectrum of the wildtype appearing at 5.5 ppm due to the interaction of Phe47 with Phe51 and Phe58. Those peaks are observed in all of the mutants, providing additional evidence that a well packed hydrophobic core is formed which is similar to that of the wildtype.

The mutants were examined using AUC to ensure that self association is not an issue (Figure 3-4 and 3-5). The apparent molecular weights determined from a single species fit are within 10 to 15% of the expected values for the P62A, W64L, W64K and W64A mutants (Table 3-1). The apparent molecular weight for the P62AW64L double mutant is 4030 Da at 20 μM, and the expected molecular weight is 4051 Da. An increase in apparent molecular weight was observed for the P62AW64L double mutant at higher concentrations, 4580 Da at 50 μM and 5120 at 90 μM, indicating that there may be some self association at higher concentrations (Table 3-1). This is consistent with the slightly broaden lines in the ^{1}H NMR spectrum recorded at 1 mM. Consequently, the stability studies were performed at protein concentrations of 30 μM or below.

3.3.3 The mutants are less stable but fold cooperatively

Having established that all of the mutants are folded, and having shown that the ability to fold is not due to self association, we moved to stability measurements. CD monitored thermal unfolding curves are shown in Figure 3-6. The curves are shown as plots of CD signal as a function of temperature, instead of as plots of fraction unfolded, because plotting fraction unfolded can camouflage problems with baselines especially for small proteins like HP36. All of the mutants exhibit sigmoidal thermal transitions. The transitions are relatively broad, as is

expected for a small domain (*183*). The reported T_m of DM HP36 is 90.6 ^oC as estimated from the derivative of the thermal unfolding curve (*147*). The mutants are less thermally stable with T_m's ranging from 68 ^oC to 79 ^oC (Table 3-2). The enthalpy of unfolding at T_m, $\Delta H^o(T_m)$, was obtained from fitting of the thermal unfolding curves. We calculated ΔH^o extrapolated to a common temperature, 75 ^oC, chosen because it is in the middle of the range of the T_m values for the mutants. We used two ΔC_p^o values in the calculation, 0.38 kcal·mol^{-1}·deg^{-1}, which is obtained from pH dependent thermal unfolding studies of wildtype HP36 and 0.23 kcal·mol^{-1}·deg^{-1}, which is derived from the analysis of the ΔH^o of surface mutants of HP36 (*198*). The high thermal stability of DM HP36 prevents an accurate estimate of $\Delta H^o(T_m)$, but $\Delta H^o(T_m)$ can be determined for wildtype HP36. The two mutations which generate DM HP36 involve solvent exposed surface residues, thus the wildtype $\Delta H^o(T_m)$ should provide a reasonable estimate for ΔH^o of DM HP36 although the K70M mutation does lead to new packing interactions as described in chapter 4. The values of ΔH^o(75 ^oC) for all mutants are within experimental uncertainty of the value for HP36 providing further evidence that mutations do not alter core packing.

We also conducted urea induced unfolding experiments. As noted in the introduction, T_m measurements may not always provide an accurate picture of protein energetics. The urea induced unfolding curves are shown in Figure 3-7 as plots of CD signal versus urea concentration. The data is plotted in this fashion to avoid disguising any potential baseline issues. All of the mutants, including the double mutant DM P62AW64L, give complete unfolding curves, highlighting the importance of using DM HP36 as our new wildtype. Several previous studies of the villin headpiece subdomain were unable to report $\Delta\Delta G^o$ values because of the low stability of the wildtype (*160, 187*). All of the mutants are less stable than the wildtype DM HP36 and the free energy of unfolding, ΔG_u^o, is decreased by ~1.3 kcal/mol for DM P62A, ~1.4 kcal/mol for

DM W64L, ~1.0 kcal/mol for DM W64A, ~1.4 kcal/mol for DM W64K, and ~2.3 kcal/mol for DM P62AW64L respectively. The *m* value of DM HP36 cannot be determined since the protein is too stable to give a complete urea induced unfolding curve but the *m*-value of wildtype HP36 has previously been determined to be 0.52 kcal·mol^{-1}·M^{-1} (*198*). The *m* values of the four single mutants are almost the same as that of HP36, while the m value of the double mutant is only slightly smaller, but is still within 8%.

Table 3-2 summarizes the measured $\Delta G_u^o(H_2O)$ values, *m*-values, and the concentration of urea at the midpoint of the transition, C_M. The thermodynamic parameters derived from the thermal unfolding curves are also included in the table. Interestingly, we find that DM W64A is slightly more stable than DM W64L and DM W64K both in terms of T_m and ΔG^o, even though the Ala sidechain can make fewer interactions with the Proline ring than the Leu sidechain. One contributing factor may be that Trp64 is located in the third α-helix and Ala has a larger helical propensity than Leu and Lys, however calculations done using the AGADIR program (*199*) show that the increase in helical propensity is small (Figure 3-8). The stability of the Lys and Leu mutants are very similar despite the fact that sidechains have different shapes. Lysine can make hydrophobic interactions with the Pro62 ring via the four carbon methylene units in its sidechain. The similar stability may reflect the fact that the sterics of the Pro interaction are less important in the present case because the groups are located on the protein surface.

3.3.4 Double mutant cycle analysis indicates that there are weak, but energetically favorable interactions between Pro62 and Trp64

Double mutant cycle analysis was used to measure the coupling free energy, $\Delta\Delta G^o_{coupling}$, between Pro62 and Trp64. As noted in chapter 2, the coupling free energy is not equal to the

contribution that the interaction makes to protein stability because desolvation effects are not included in the analysis, as they cancel to first order (*61, 179*). This is important when analyzing charged and polar interaction, but is likely to be less important here because the residues are nonpolar and on the surface. As discussed in chapter 2, an important point is that the standard interpretation of coupling free energies always makes the assumption that the mutations do not alter the free energy of the denatured state ensemble (DSE) or that they alter them in such a way that the cycle cancels unfolded state effects. When this is true the coupling free energy represents the strength of the interaction in the native state. It is very difficult to determine if mutations alter the energetics of the DSE. It is worth noting that there is experimental evidence for residual structure in the DSE of HP36 (*80, 155, 156*). In addition, a Pro-to-Ala substitution should affect local conformational propensities in the DSE owing to the steric constraints imposed by a proline and should thus modulate the chain entropy of the unfolded state. The calculated interaction free energy between Pro62 and Trp64 is a favorable 0.5 kcal/mol. The value is small, but not zero.

3.4 Discussion

The data presented here shows that all of the mutants are well folded and the spectroscopic evidence indicates that they share the same fold as wildtype, while the denaturation experiments indicate that the mutants, like the wildtype, appear to fold cooperatively. Note in particular the similarity in the ΔH^o of unfolding extrapolated to a common temperature and the close correspondence of the m-values for the mutants and for HP36. The data show that while Pro62 and Trp64 make contributions to the stability of the subdomain, they are not essential for specifying the structure of the folded state. In fact their contribution to stability is well within the range observed for mutations at other positions in HP36. For example single Phe to Leu

substitutions at position 47, 51 or 58 have larger effects on the stability than do any of the Pro62 or Trp64 mutants. Even the P62AW64L double mutant has a smaller effect on stability than a F58L single mutant (*125*). The Phe-to-Leu mutations described in chapter 2 involve residues buried in the core and one might argue that their mutation to Leu should be expected to lead to larger perturbations in ΔG^o than mutation of residues such as Pro62 and Trp64 which lie on the surface and make more localized interactions. However, the energetic effect of the Pro62 or Trp64 single mutations are also comparable to the effects observed for other surface mutations and even the P62AW64L double mutant is not significantly more destabilizing than other pairs of surface mutants (*147*). For example, replacing Ala68 and Met70 in DM HP36 with Asn and Lys to regenerate the wildtype sequence leads to a $\Delta\Delta G^o$ of ~1.8 kcal/mol while the P62AW64L double mutant destabilizes DM HP36 by ~2.3 kcal/mol. Thus, Trp64 and Pro62 do not appear to have unique roles, but rather make typical contributions to the stability of the subdomain. The fact that some mutations at these sites fail to fold in some constructs is simply due to the low stability of the wildtype.

The concept of gatekeepers is intriguing, but it would seem desirable, from the prospective of evolutionary fitness, to avoid having to rely on a small subset of critical sites to specify a fold since having a broader distribution of sites that contribute to defining the fold might be more robust to random mutation.

As previously noted, analysis of large body of protein thermodynamic data has led to scaling laws relating the magnitude of ΔG^o and T_m to size. The approximate relationship is derived based on parameterization of ΔS^o and ΔH^o per residue for globular proteins (*184, 200*). It is instructive to compare the stability of HP36 normalized by the number of residues, N, to the normalized stability of other proteins. If very small domains required exceptional interactions to

fold one would expect that their normalized stability would be larger than observed for more typical globular proteins. Figure 3-9 displays a plot of $\Delta G^o/N$ vs. N for a set of proteins listed in Table 3-3. The proteins originally analyzed by Ghosh (*200*) and Dobson (*201*) are included, but domains which contain disulfides or bind cofactors are excluded. Data for additional 5 proteins which are less than 50 residues are included in the plot (*180, 202, 203*). No detectable correlation is observed between size and normalized stability. The average value is 0.067 ± 0.023 kcal/mol per residue. The value of HP36, 0.089, is well within this range.

It is also interesting to examine the relationship between ΔG^o and T_m for the mutants of HP36. The available data for Pro62, Trp64, core aromatic, and other surface residues is plotted in Figure 3-10. Measurements reported under identical solution conditions and made using the same methodology are included (*125, 147*). The previously analyzed mutants exhibit an empirical linear relationship between T_m and ΔG^o. The Pro62 and Trp64 single mutants and the double mutant fall on the line, providing further evidence that they make typical contributions to stability. The large changes in T_m observed for mutations in HP36 do not require that the residue in question makes exceptional contributions to stability, instead they reflect the expected size dependent relationship between $\Delta\Delta G^o$ and ΔT_m (*184*).

The analysis presented here, together with other studies of small stably folded proteins, argues that special, highly optimized interactions are not required to stabilize their structures (*183, 194*). Instead the thermodynamics and cooperativity of their folding are normal in the sense that their behavior falls within the range predicted by relationships between protein size and folding thermodynamics. On a practical level the analysis presented here illustrates the utility of using stabilized domains for the thermodynamic analysis of small proteins and reinforces the importance of measuring $\Delta\Delta G^o$ as well as ΔT_m. In the present case mutation of two solvent

exposed surface residues lead to the DM HP36 variant that was sufficiently stable to allow the energetic costs of other mutations to be accurately described.

Finally, we note that a recent computational study estimated the energetic consequence, $\Delta\Delta G^{o}$, of the Pro62-to-Ala mutation to be 1.2 kcal/mol (*152*). The calculation could not be directly compared to experimental data since only ΔT_m values were available in the literature. Our data shows the calculation is remarkably accurate.

Table 3-1. Comparison of theoretically calculated and apparent molecular weights from AUC data for the four single mutants and one double mutant. The uncertainties were the standard errors to the fit.

Mutant	Protein concentration	Theoretical MW	Observed MW
DM P62A	90 μM	4124	4630 ± 49
DM W64L	90 μM	4076	4471 ± 58
DM W64K	90 μM	4091	4591 ± 123
DM W64A	20 μM	4035	4631 ± 412
DM P62AW64L	20 μM	4051	4028 ± 223
	50 μM	4051	4578 ± 124
	90 μM	4051	5115 ± 70

Table 3-2. Thermodynamic parameters for the unfolding of HP36 and DM HP36 mutants. Data were collected at 25 ^{o}C in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0.

Protein	T_m (^{o}C)	$\Delta H^{o}(T_m)$ (kcal/mol)	$\Delta H^{o}(75^{o}C)$ (kcal/mol) ΔCp^{o}=0.23	$\Delta H^{o}(75^{o}C)$ (kcal/mol) ΔCp^{o}=0.38	$\Delta G_u{}^{o}$ (kcal/mol)	m-value (kcal/mol·M)	C_M [1] (M)
HP36 [2]	73.0 ±1.5	31.8 ±0.3	32.3	32.6	3.17 ±0.06	0.52 ±0.01	6.2
DM HP36 [2]	90.6	----- [3]	-----[3]	-----[3]	4.94-5.06	----- [4]	9.6
DM P62A	71.8 ±0.3	35.5 ±0.5	36.2	36.7	3.67 ±0.12	0.52 ±0.02	7.5
DM W64A	78.7 ±0.4	37.8 ±0.4	36.9	36.4	3.97 ±0.07	0.52 ±0.01	7.85
DM W64L	74.4 ±0.3	36.6 ±0.4	36.7	36.8	3.58 ±0.05	0.52 ±0.01	6.75
DM W64K	74.9 ±0.4	36.4 ±0.5	36.4	36.4	3.59 ±0.10	0.53 ±0.02	6.4
DM P62AW64L	67.9 ±0.4	36.2 ±1.2	37.8	38.9	2.74 ±0.07	0.48 ±0.01	5.5

The numbers listed after the ± symbol are the standard errors to the fit.
1) C_M is determined by calculating the derivative of the plot of CD signal vs. [urea].
2) Date for HP36 and DM HP36 taken from Bi *et al*.(*147*)
3) The high thermal stability of DM HP36 prevented determination of $\Delta H^{o}(T_m)$.
4) The high stability of DM HP36 prevents an accurate estimate of the m-value.

Table 3-3. Thermodynamic parameters of the proteins

Protein	PDB code	Chain length	ΔG (kcal/mol)	ΔG/N (kcal/mol·res)	Temperature (°C)	Thermophilic/ Mesophilic?	Unfolding condition
Abp1 SH3 [1]	1JO8	69	3.11	0.045	25	M	Gdn·HCl
ACBP [1]	2ABD	86	7.80	0.091	5	M	Gdn·HCl
ADAh2 [1]	1PBA	81	4.09	0.050	25	M	urea
CI2 [1]	1COA	66	7.78	0.118	25	M	Gdn·HCl
CTL9 [1]	1DIV	93	6.51	0.070	25	T	urea
FKBP12 [1]	1FKB	110	5.60	0.051	25	M	urea
FRB [1]	2GAQ	100	6.34	0.063	23	M	Gdn·HCl
HPr [1]	1HDN	85	4.98	0.059	25	M	urea
Im7 [1]	1CEI	87	6.34	0.073	10	M	urea
Im9 [1]	1IMQ	86	6.24	0.072	10	M	urea
NTL9 [1]	2HBB	56	4.14	0.074	25	T	Gdn·HCl
raf RBD [1]	1RFA	81	6.22	0.077	25	M	Gdn·HCl
Spectrin SH3 [1]	1SHG	64	3.33	0.052	25	M	urea
Src SH2 [1]	1HCS	112	7.42	0.066	25	M	urea
Src SH3 [1]	1SRL	64	3.80	0.059	22	M	Gdn·HCl
U1A [1]	1URN	103	8.33	0.081	25	M	Gdn·HCl
Ubiquitin [1]	1UBQ	76	7.13	0.094	25	M	Gdn·HCl
lamba repressor G46AG48A [1]	1LMB	80	4.81	0.060	37	M	urea
CspB [1]	1CSP	66	4.81	0.073	25	M	Gdn·HCl
CspB [1]		68	6.32	0.093	25	T	Gdn·HCl
CspA [1]	1MEF	69	3.01	0.044	25	M	urea
CspA [1]	1MEF	69	2.89	0.042	10	M	urea
PI3 kinase [1]	1PKS	84	3.40	0.040	20	M	Gdn·HCl
Fyn [1]	1NYF	67	6.00	0.090	20	M	Gdn·HCl
TWIg18 [1]		93	3.90	0.042	20	M	urea
TNfn3 (short form)[1]	1TEN	90	4.81	0.053	20	M	urea
TNfn3 (long form) [1]		90	6.70	0.074	25	M	urea

Arc repressor [1]	1ARR	106	6.32	0.060	25	M	urea
IgG binding domain of protein L [1]	2PTL	62	4.59	0.074	22	M	Gdn.HCl
Hpr [1]	1HDN	85	4.59	0.054	20	M	Gdn·HCl
Muscle Acp [1]	1APS	98	5.41	0.055	28	M	urea
T4 Lysozyme [2]	1L63	164	10.67	0.065	25	M	thermal
acyl carrier protein [2]	1ACP	77	2.78	0.036	25	M	thermal
IL-1β [2]	6I1B	153	5.53	0.036	25	M	thermal
Barnase [2]	1BNI	110	10.50	0.095	25	M	urea
Barstar [2]	1BTA	89	4.81	0.054	25	M	urea
ROP [1]	1ROP	126	7.73	0.061	25	M	Gdn·HCl
B2 of protein G [2]	1PGX	56	5.55	0.102	25	M	thermal
B1 of protein G [2]	1PGB	56	5.74	0.103	25	M	urea
stefin A [2]	1CYV	98	9.35	0.095	25	M	thermal
Sac7d [2]	1AZQ	66	4.57	0.069	25	T	thermal
1-39 NTL9 [3]		39	1.76	0.045	25	T	urea
PSBD [4]	1PRB	41	2.14	0.052	25	T	Gdn·HCl
GA [4]	2PDD	47	4.71	0.100	25	M	urea
UBA [4]	1UBA	45	1.40	0.031	25	M	Gdn·HCl
hPin1 WW [5]	1I6C	34	1.78	0.052	37	M	thermal
S16 [2]	3BN0	112	6.22	0.056	25	T	thermal
CheY [2]	1TMY	120	9.47	0.079	25	T	thermal
Sso7 [2]	1SSO	62	7.92	0.128	25	T	thermal
MGMT [2]	1MGT	174	10.12	0.058	25	T	thermal
Cold Shock Tm [2]	1G6P	66	4.02	0.061	25	T	thermal
Ribosomal L30E [2]	1H7M	100	11.84	0.118	25	T	thermal
TRP repressor [2]	2OZ9	107	10.14	0.095	25	M	thermal
Parvalbumin [2]	5CPV	108	13.09	0.121	25	M	thermal
HistidineBSt [2]		88	9.31	0.106	25	T	thermal
HT c552 [2]	1AYG	80	7.44	0.093	25	T	thermal
RNaseH [2]	1RIL	166	8.85	0.053	25	T	thermal
SPCI [2]		180	13.04	0.072	25	M	thermal

Histidine Bh [2]		87	7.73	0.089	25	T	thermal
Btk [2]	1AWX	64	3.95	0.062	25	M	thermal
Iso2 Cytochrome [2]	1YEA	112	10.45	0.094	25	M	thermal
Sem5 [2]	1KFZ	60	6.48	0.108	25	M	thermal
Cold Shock Bc [2]	1C9O	66	4.81	0.073	25	M	thermal
Histidine St [2]	1Y4Y	87	8.80	0.101	25	T	thermal
Bergerac D48G [2]	1BK2	57	4.90	0.086	25	M	thermal
HistidineBS [2]	1SPH	88	6.00	0.068	25	M	thermal
Bergerac SHH mutant [2]	2OAW	65	4.86	0.075	25	M	thermal
Tec [2]	1GL5	63	3.16	0.050	25	M	thermal
Glutamate dehydrogenase domain II [2]	1B26	149	5.10	0.034	25	T	thermal
Itk [2]	2RNA	57	3.06	0.054	25	M	thermal
Abl [2]	1BBZ	63	3.61	0.057	25	M	thermal
Cytochrome b562 [2]	1QPU	106	5.14	0.049	25	M	thermal
Bergerac SHA mutant [2]		70	4.21	0.060	25	M	thermal
Bergerac SH3 [2]	1SHG	62	3.71	0.060	25	M	thermal
Lac Repressor Headpiece [2]	1LCD	51	2.56	0.050	25	M	thermal
Rnase B [2]	1RBB	124	11.24	0.091	25	M	thermal
Histidine Ec [2]	1POH	85	5.26	0.062	25	M	thermal
Arabinose Binding Protein [2]	1ABE	305	14.83	0.049	25	M	thermal

EIN [2]		258	11.39	0.044	25	M	thermal
Lamda-Repressor 6-85 [2]	1LMB	80	4.31	0.054	25	M	thermal
Cold Shock A [2]	1MJC	70	2.92	0.042	25	M	thermal
CT-Acp [2]	2VH7	98	4.21	0.043	25	M	thermal
ECAD2 [2]	1EDH	106	6.24	0.059	25	M	thermal
α-sarcin [2]		150	9.81	0.065	25	M	thermal
Stefin B [2]	1STF	95	3.83	0.040	25	M	thermal
Rnase Sa3 [2]		99	5.19	0.052	25	M	thermal
S16 [2]		117	3.21	0.028	25	M	thermal
CheY [2]	3CHY	120	3.13	0.026	25	M	thermal
HP36	1VII	36	3.22	0.089	25	M	urea

1) Data from ref. (*201*)
2) Data from ref. (*200*)
3) Data from ref. (*180*)
4) Data from ref. (*203*)
5) Data from ref. (*202*)

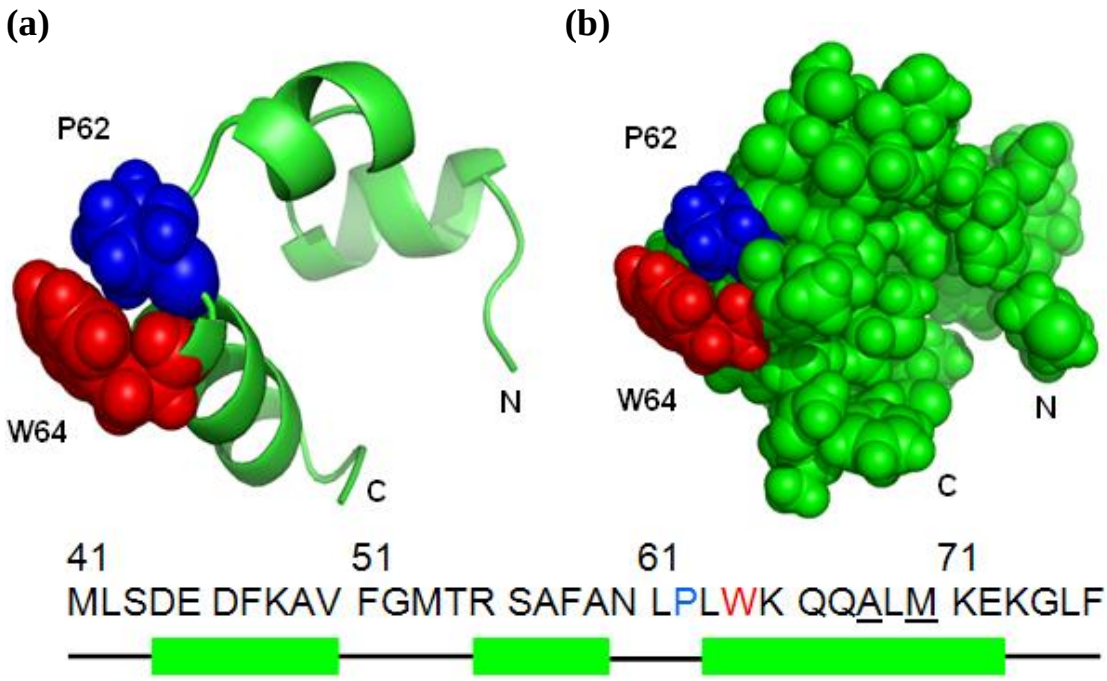


Figure 3-1. (a) A ribbon diagram of the structure of DM HP36 (*147*) generated using PyMOL, version 0.99. The N and C-termini are labeled. Pro 62 and Trp 64 are shown in space filling format and are labeled. (b) A space filling representation of DM HP36 shown in the same orientation as part (a). The primary sequence of DM HP36 is shown together with a diagram of the secondary structure at the bottom of the figure. α-helices are displayed as cylinders. Pro62 is indicated in blue, Trp64 in red and the sites which are mutated in DM HP36 relative to widltype are underlined.

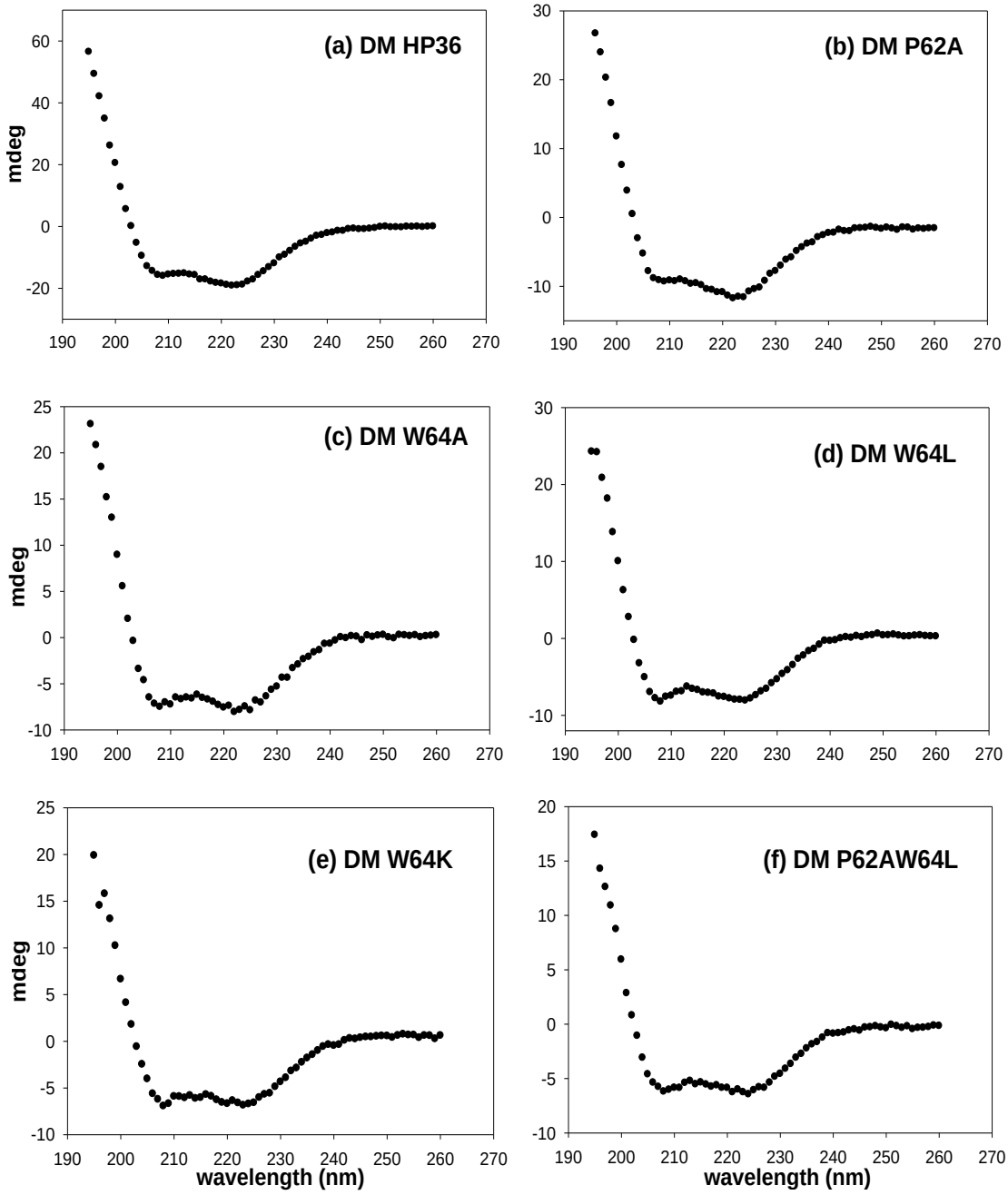
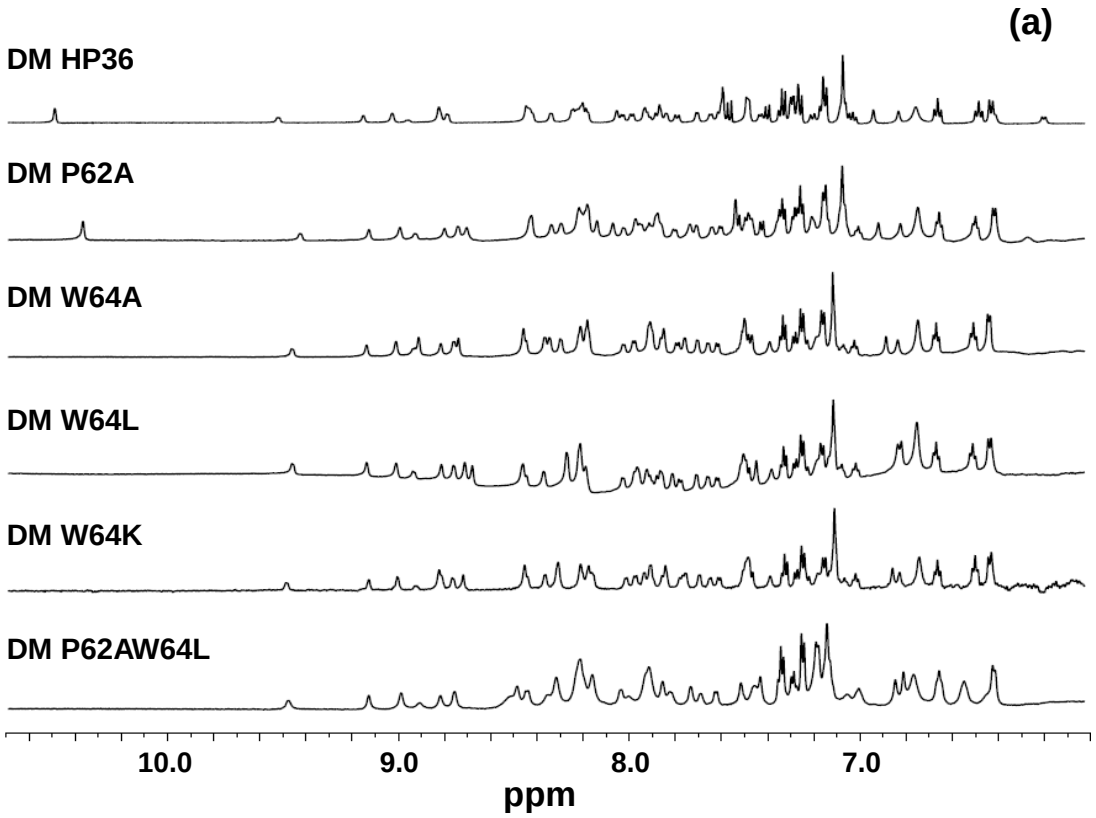


Figure 3-2. Far UV CD spectra show that all mutants are folded. (a) DM HP36, (b) DM P62A, (c) DM W64A, (d) DM W64L, (e) DM W64K, (f) DM P62AW64L. All spectra were collected at 25^{o}C in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0 using a 0.1 cm cuvette.


(a)
DM HP36
DM P62A
DM W64A
DM W64L
DM W64K
DM P62AW64L
10.0
9.0
8.0
7.0
ppm

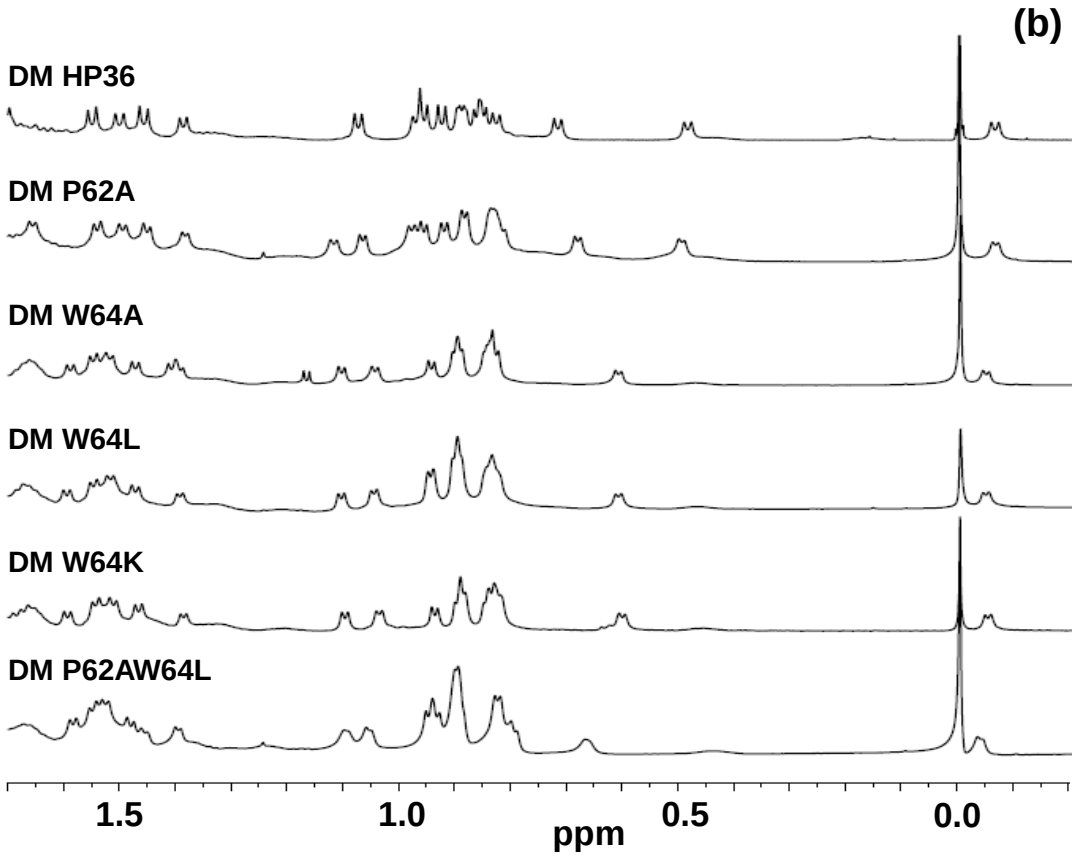


Figure 3-3. ^{1}H NMR spectra show that all mutants are folded. (a) Stacked plot of the downfield region of the spectra of the mutants. (b) Stacked plot of the upfield region of the spectra of the mutants. Spectra are labeled to denote the mutant to which they correspond. The sharp peak at 0.0 ppm is the chemical shift standard. All spectra were collected at 25^{o}C in 90% H_2O, 10% D_2O, 10 mM sodium acetate, 150 mM sodium chloride at pH 5.0.

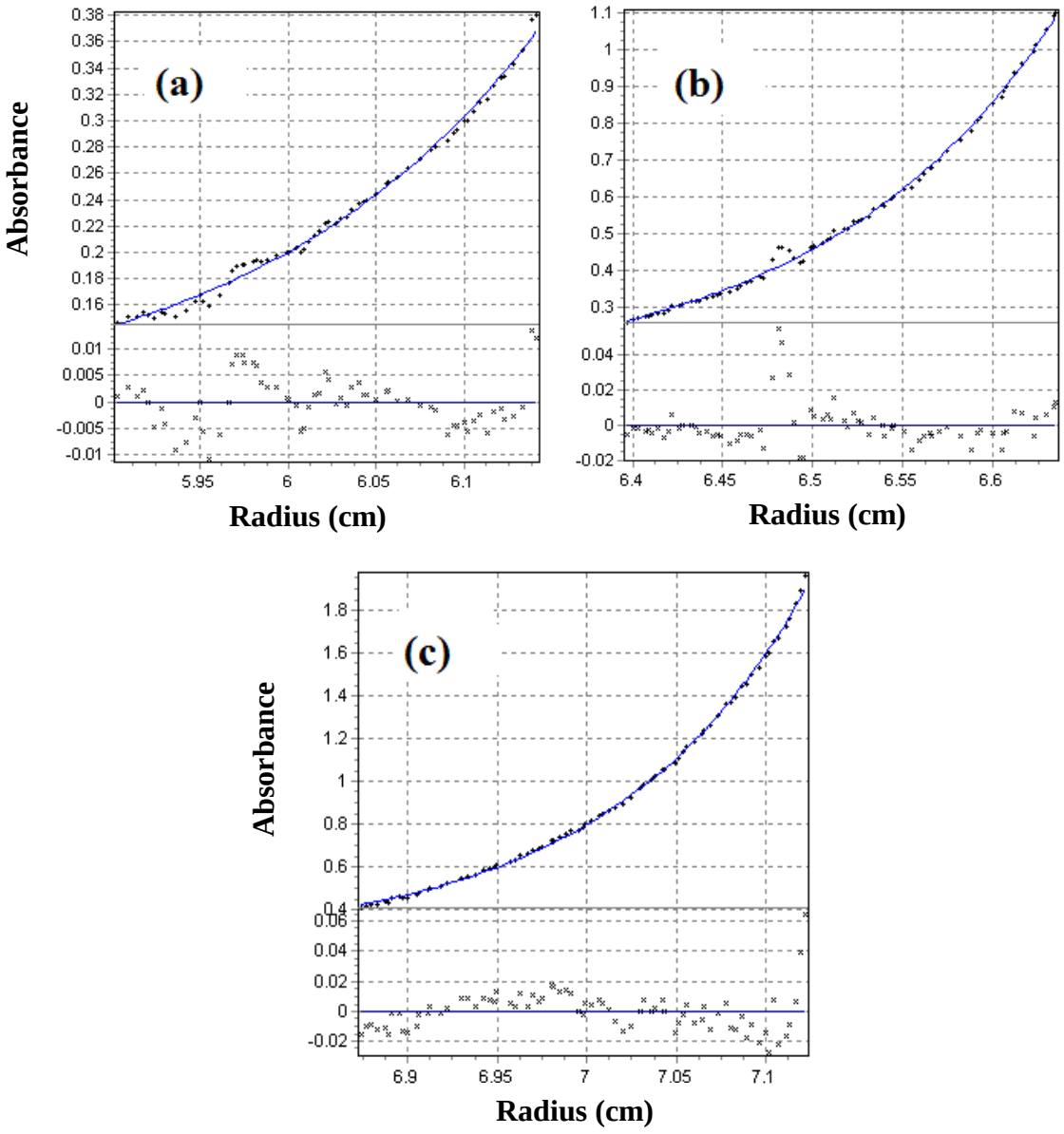


Figure 3-4. Concentration dependent analytical ultracentrifigation data for the DM P62AW64L double mutant. Sample concentrations were (a) 20μM, (b) 50μM and (c) 90μM. The apparent molecular weights were 4028, 4578 and 5115 respectively. The residuals for a fit to a single species model are shown at the bottom of each plot. Experiments were conducted at pH 5.0 in 10 mM sodium acetate 150 mM sodium chloride.

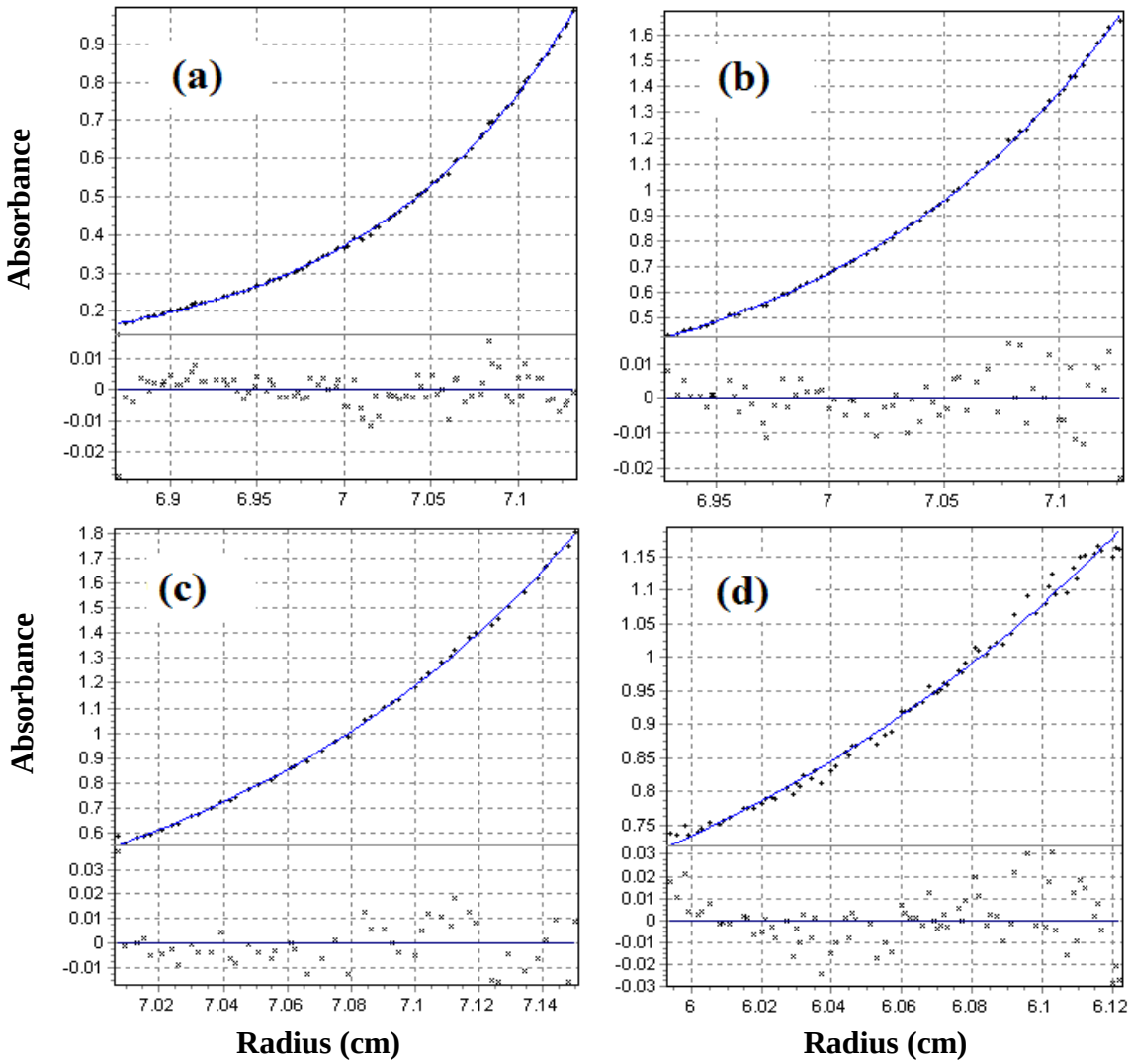


Figure 3-5. Analytical unltracentrifigation data collected for the various single mutants at 25 ^{o}C, pH 5.0, 10 mM sodium acetate 150 mM sodium chloride. (a) DM P62A at 90 μM, (b) DM W64L at 90 μM, (c) DM W64K at 90 μM, (d) DM W64A at 20 μM. The apparent molecular weights are: 4630 for DM P62A, 4471 for DM W64L, 4591 for DM W64K and 4631 for DM W64A. The residuals for a fit to a single species model are shown at the bottom of each plot. Experiments were conducted at pH 5.0 in 10 mM sodium acetate 150 mM sodium chloride.

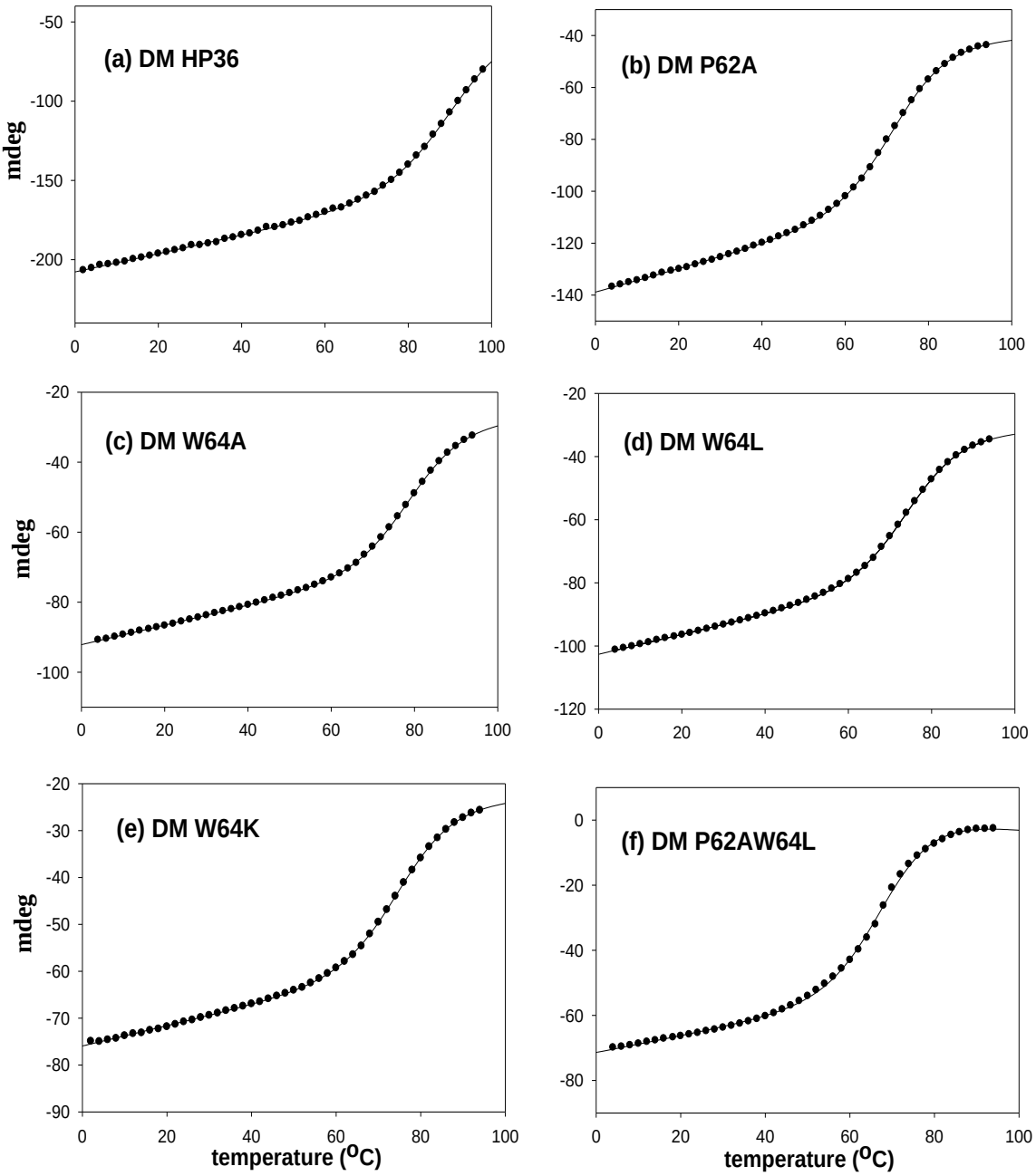


Figure 3-6. CD monitored thermal unfolding curves for DM HP36 and mutants. (a) DM HP36, (b) DM P62A, (c) DM W64A, (d) DM W64L, (e) DM W64K, (f) DM P62AW64L. Samples were between 15 μM and 30 μM protein in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0 in a 1.0 cm cuvette. The CD signal was recorded at 222 nm. The solid line represents the best fit to a two-state folding transition.

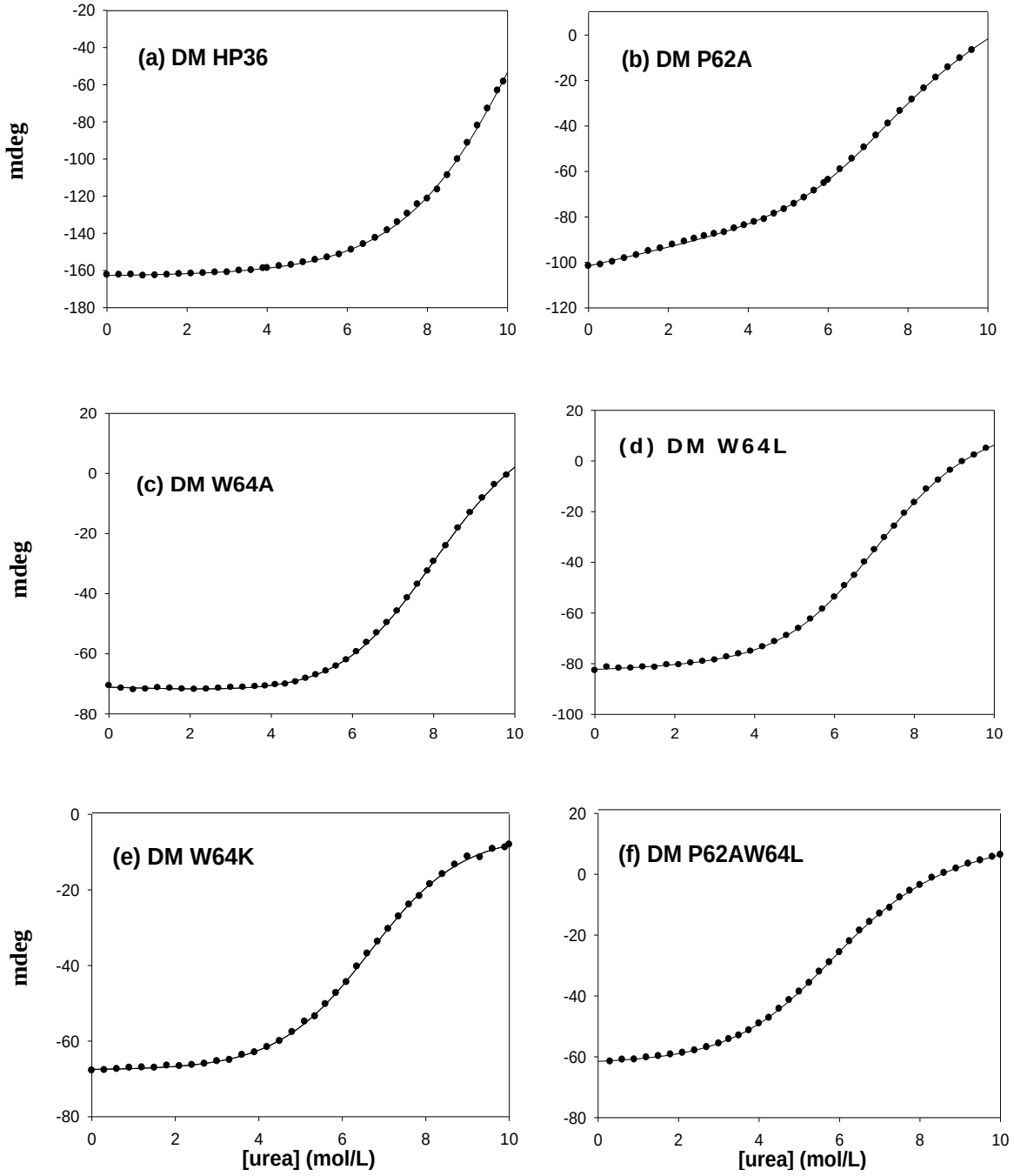


Figure 3-7. CD monitored urea induced unfolding curves for DM HP36 and mutants. (a) DM HP36, (b) DM P62A, (c) DM W64A, (d) DM W64L, (e) DM W64K, (f) DM P62AW64L. Samples were between 15 μM and 30 μM protein in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0. The CD signal was recorded at 222 nm. The solid line represents the best fit to a two-state folding transition.

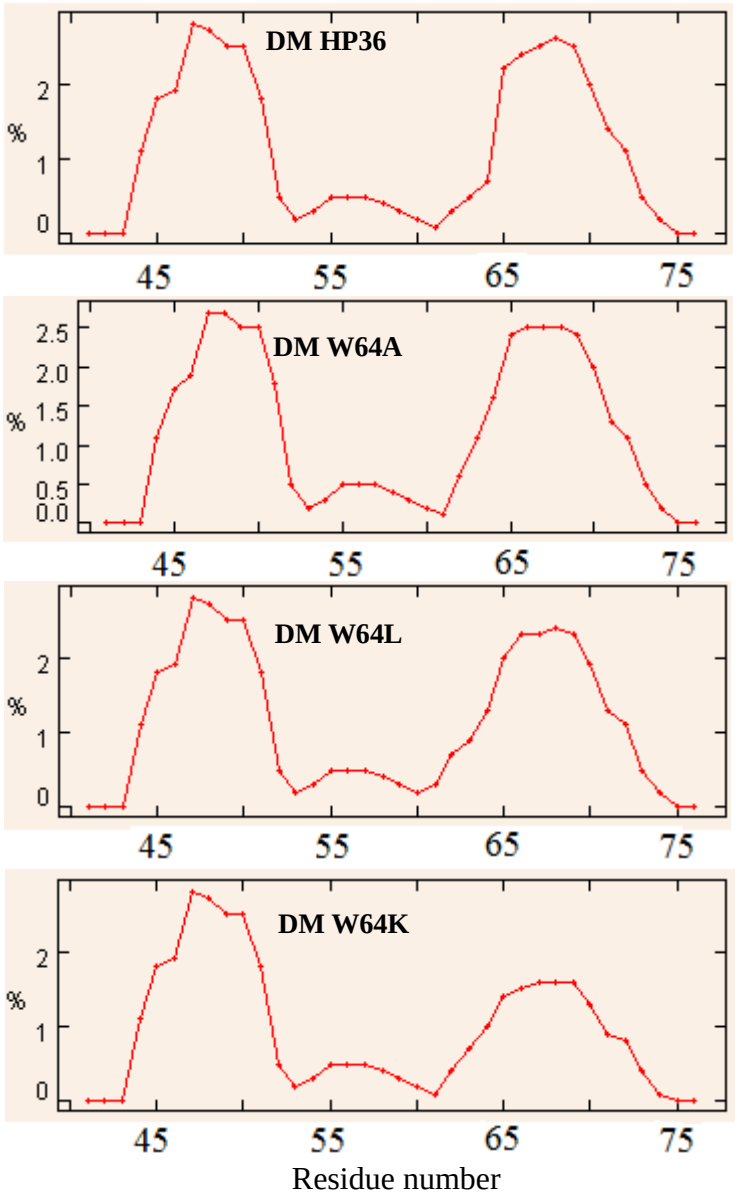


Figure 3-8. Residue level helix content calculations for DM HP36 and the three Trp mutants using AGADIR program (http://agadir.crg.es/).

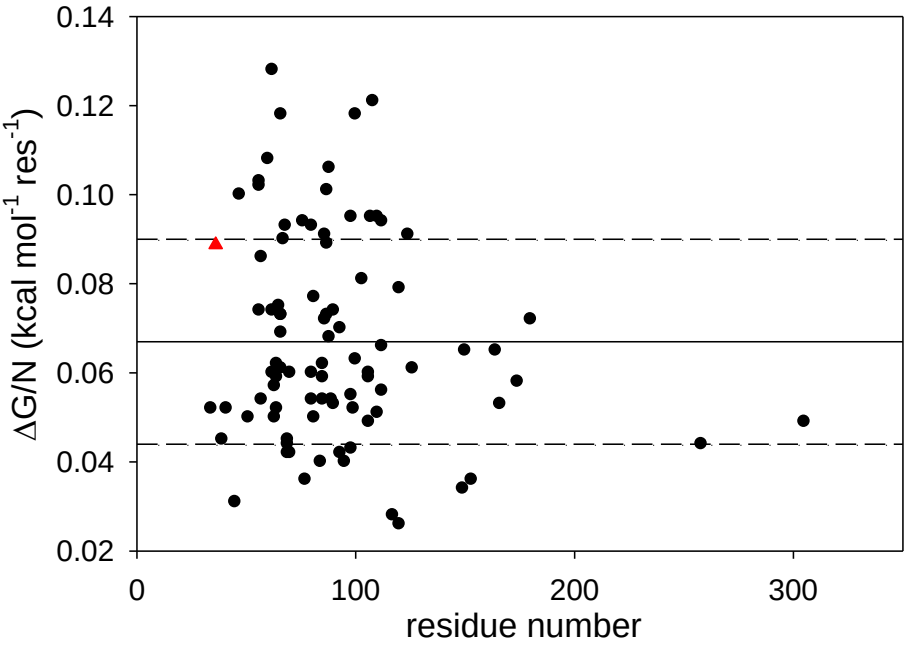


Figure 3-9. A plot of stability normalized by size, $\Delta G^{o}/N$, vs. protein size in residues, N. The solid line represents the average value, 0.067 kcal/mol per residue. The dashed lines represent ± one standard deviation. The value for HP36 is indicated as a red triangle.

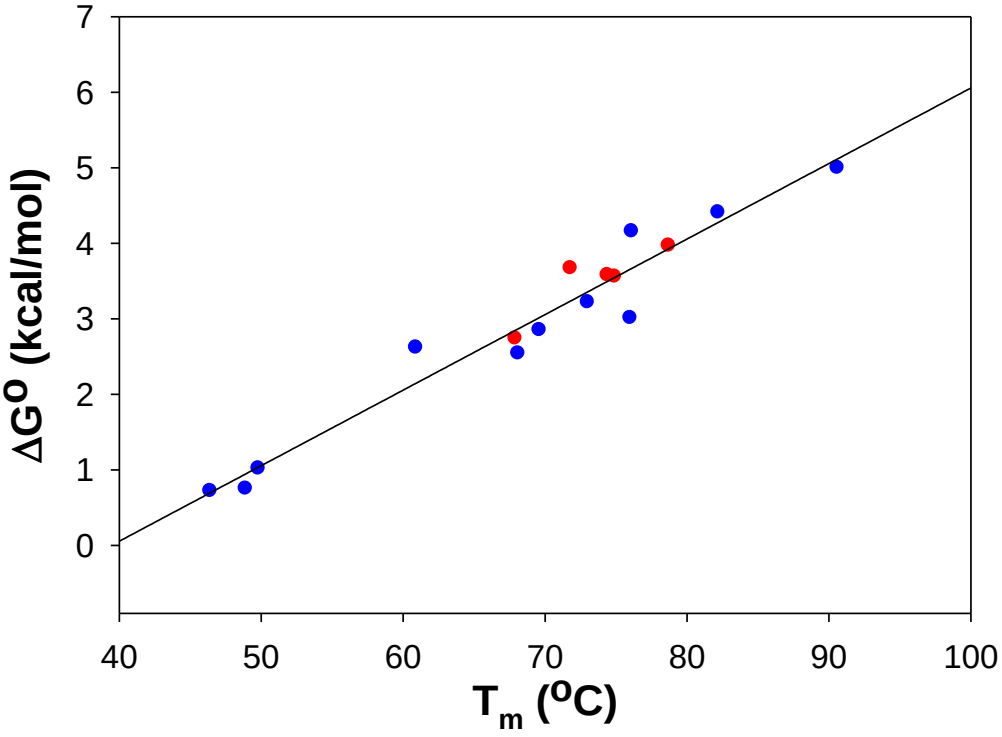


Figure 3-10. A plot of ΔG^o vs T_m for DM HP36 mutants. The values for the Pro62, Trp64 mutants are indicated as red circles and non Pro62, Trp64 mutants as blue circles. The solid line represents a linear fit to all of the points, with a slope of 0.100 $\pm$ 0.006 kcal mol^{-1} $^oC^{-1}$. R^2 = 0.95.

4. Rational modification of protein stability by targeting surface sites: simple approaches, complicated results.

Abstract

The rational modification of proteins to improve their stability or solubility is an important goal in biotechnology and protein design. The mutation of charged surface residues is an attractive approach since surface mutations run little risk of perturbing the structure. It is generally thought that surface exposed charged residues are not involved in critical packing interactions, are not subject to large desolvation penalties upon folding, and are not involved in unfolded state electrostatic interactions, thus residues to target can be selected using estimates of the surface electrostatic potential. Any increase in stability is usually assumed to result from modification of native state electrostatic interactions and the approach is believed to avoid generating a significant number of false negatives. We use a combined computational and experimental approach to show that the situation is much more complicated, even for very simple proteins. Individual mutation of three surface exposed lysines, K48, K65 and K70, in the villin headpiece helical domain increases ΔG^o of unfolding by 0.7 to 1.2 kcal mol^{-1} and increases the melting temperature by 4 to 9 ℃. The residues are on the surface and exposed, but the mechanism of stabilization is very different in each case. The K48M mutation actually destabilizes native state electrostatic interactions, but has a more pronounced effect on unfolded state interactions leading to a net increase in stability. The K65M mutation alters the desolvation penalty paid by this residue in the native state, while the K70M mutation introduces new packing interactions, but does not alter electrostatic interactions significantly in the folded or unfolded states. Only one of these sites would be identified by currently available algorithms designed to target surface

electrostatics.

Acknowledgements

I thank Dr. Jae-Hyun Cho for helpful discussions and Dr. Yuan Bi and Dr. Bing Shan for experimental assistance. The work described in this chapter is a collaborative experimental computational study. The computational work was performed by Mr. Vadim Patsalo.

4.1 Introduction

Protein stability is an important consideration in protein engineering, and there is a great interest in developing methods to increase the stability of a protein. Major forces in folding of globular proteins are the hydrophobic interaction and hydrogen bonding, and they are just able to overcome the major force favoring unfolding, that is conformational entropy. Most globular proteins have a low conformational stability of just 2-10 kcal/mol (*204*). It is known that many proteins are unfolded under physiological conditions, but fold when it is required for their function (*205*). Thus, net interactions in the 1-3 kcal/mol range such as electrostatic interaction become important. Surface electrostatic interactions in proteins are potentially attractive targets for the rational redesign of proteins because it is difficult to design a protein by burying hydrophobic surface or by adding hydrogen bonds. This approach relies upon identifying residues which are involved in unfavorable surface electrostatic interactions, typically interactions with residues of the same charge, or by identifying sites where new favorable electrostatic interactions can be introduced. Several studies have shown that charge reversal mutations on the protein surface can be used to increase stability by > 1 kcal/mol (*206-211*). Some of the stability increases are less than predicted by summing the charge-charge interactions using Coulomb's law and it is suggested that the charge-charge interactions that stabilize the folded state of proteins also stabilize the unfolded state so that the net contribution to protein stability is small (*212*).

Sites to be modified are selected based upon electrostatic calculations which treat the electrostatic interactions as a sum of pairwise additive terms. Methods include simple inspection of the protein surface, modified Tanford-Kirkwood approaches, and Poisson Boltzmann (PB) calculations. PB calculations can be combined with molecular dynamics simulations to treat protein flexibility and possible relaxation of the protein matrix. Irrespective of the details, the

strategy is based upon the assumption that surface electrostatic interactions can be modified without altering other native state interactions or without altering unfolded state energetics. An attractive aspect of these approaches is that they are computationally inexpensive and, unlike selection based methods or directed evolution, involve the generation of a limited number of mutants. The methods are thought to delineate all sites which are targets for rational redesign, provided that the energetics of charged surface residues are dominated by electrostatic and desolvation effects, and are thus believed to avoid false negatives.

The villin headpiece helical subdomain is a small, cooperatively folded three helix protein. Despite its small size, the domain is a typical cooperatively folded globular protein and is stabilized by a typical balance of interactions. There are ten charged residues in the protein including two Asp, two Glu, one Arg and five Lys residues (Figure 4-1). The lysine residues are distributed across the surface of the protein and their sidechain solvent accessibility ranges from 39% to 72% (Table 4-1). However, the solvent accessibility of their NH_3^+ group ranges from 76% to 89% (Table 4-1). The solvent accessibility is calculated by the NACCESS program (*213*) based on the 200 snapshots of the MD simulation trajectory from the crystal structure of HP36 reported in ref. (*147*). The percentage sidechain exposure was given by the ratio of side-chain surface area to "random coil" value, that is the average solvent-accessible surface area of X in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations (*214*). The percentage of NH_3^+ exposure was given by the ratio of NH_3^+ surface area to the maximal area observed for the NH_3^+ group in the calculation. Here we individually mutated each lysine to a methionine, chosen because methionine is approximately isosteric for lysine, and studied both native state and unfolded state electrostatic interactions for the three stabilized mutants. We show that the situation is much more complicated, even for simple proteins, and demonstrate that surface charge mutants can stabilize proteins by a

myriad of effects, including ones which are not predicted by electrostatic based methods, even when only solvent exposed sites are targeted.

4.2 Materials and methods

4.2.1 Protein expression and purification

All proteins were expressed as a C-terminal fusion linked to the N-terminal domain of L9 by a factor Xa cleavage site as described in chapter 2. Uniformly labeled ^{13}C, ^{15}N labeled HP36 was produced by using M9 minimal media with $^{15}NH_4Cl$ and ^{13}C glucose. The factor Xa cleavage was carried out at 23 ^{o}C for 16-20 h. All proteins were purified by ion exchange chromatography and reverse-phase HPLC using a gradient of 30-65% buffer B in 70 minutes. Buffer A was H_2O (0.1% TFA) and buffer B was 90% acetonitrile and 10% H_2O (0.1% TFA). The elution time for all proteins is around 47% buffer B. The identities of all proteins were confirmed by matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI). The observed and expected molecular masses were as follows: K48M, observed 4193.9, expected 4192.9; K65M, observed 4192.3, expected 4192.9; K70M, observed 4193.3, expected 4192.9; K70E, observed 4190.4, expected 4190.8; K70A, observed 4131.3, expected 4132.8; K71M, observed 4194.3, expected 4192.9; K73M, observed 4194.4, expected 4192.9.

4.2.2 Peptide synthesis and purification

Peptides were synthesized by standard solid-phase Fmoc chemistry on an Applied Biosystems 433A automated peptide synthesizer. Peptides were cleaved from the resin with a 91% trifluoroacetic acid (TFA)/ 3% anisole/ 3% thioanisole/ 3% ethanedithiol mixture. Purification of the peptides was achieved by reverse-phase HPLC. Buffer A was 100% H_2O + 0.1% TFA and

buffer B was 90% acetonitrile + 10% H_2O + 0.1% TFA. The identities of peptides were confirmed by MALDI mass spectrometry. The observed and expected molecular weights were as follows: peptide MLSDEDFMA, observed 1056.5, expected 1057.1; peptide MLSDEDFKAVFGM, observed 1490.4, expected 1489.7; peptide MLSDQDFKAVFGM, observed 1489.4, expected 1489.0; peptide KEKGLF, observed 764.0, expected 762.8; peptide KKEKGLF, observed 891.9, expected 891.0.

4.2.3 Protein stability measurements

Protein stability was measured by CD monitored urea and thermal denaturation experiments. Urea induced unfolding was performed at 298 K and 222 nm with samples of 15-30 μM protein in 10 mM sodium acetate and 150 mM sodium chloride at pH's ranging from 3.5 to 7 on an AVIV 202SF spectrophotometer. The concentration of urea was increased from 0 to about 10 M in ~0.25 M steps. Urea concentrations were determined by measuring the refractive index. Urea unfolding curves were analyzed by a non-linear least squares fit to equation 1.12. Thermal unfolding experiments were performed on an Applied Photophysics Chirascan CD instrument over the range of 2 ^{o}C to 94 ^{o}C with a 2 ^{o}C interval and the signal was monitored at 222 nm. The buffer and protein concentration were the same as used in the urea denaturation experiments, but the pH range was from 2 to 3.5. Thermal unfoldingwas used at low pH because achieving low pH values in urea solutions requires the addition of significant amounts of acid owing to protonation of the urea. The reversibility was confirmed by comparing the initial CD signal at the start of the run to the signal measured after the run was completed and the sample was cooled to the starting temperature. All thermal unfolding curves were analyzed by non-linear least squares fitting using SigmaPlot. Data was fit to equation 1.10.

4.2.4 Protein pKa measurements

The pKa values of the Asp and Glu in wildtype were measured by following the chemical shifts of the side-chain carboxyl carbons over the pH range of 2 to 7. All chemical shifts were determined using 2D version of the HCACO experiments at various pH values, after adjusting the pH using HCl and NaOH. The pKa values of the Asp and Glu in K48M, K65M and K70M mutants were measured by observing the chemical shifts of Hβ and Hγ respectively over the same pH range as above. The pKa of the C-terminal residue, F76, was determined by following the Hα chemical shifts of as a function of pH. TOCSY experiments with 75 ms mixing time were conducted at various pH values in order to determine the proton chemical shifts. Data were fit into Henderson-Hasselbalch equation to determine the pKa:

$$\delta(pH) = \frac{\delta_{base} + \delta_{acid} 10^{(pK-pH)}}{1+10^{(pK-pH)}} \qquad (4.1)$$

where δ is the chemical shift, δ_{base} is the chemical shift associated with the unprotonated residue, δ_{acid} is the chemical shift associated with the protonated residue, and pK is the pKa value for the residue. δ_{base} and δ_{acid} were described as a linear function of pH: a + b × pH. Data were fit using the nonlinear least-squares regression routine in the program SigmaPlot.

4.2.5 Peptide pKa measurements

The pKa values of D44, E45 and D46 in the peptides were measured by following the chemical shifts of Hβ, Hγ and Hβ respectively using TOCSY experiments over the pH range of 2-7. The pKa values of E72 in the peptides were measured by monitoring the Hγ chemical shifts through 1D NMR experiments over the pH range of 2-7. The pKa values of F76 in the peptides were measured by the Hα chemical shifts through 1D NMR experiments over the pH range of 3-7 and through TOCSY experiments over the pH range of 2-3. Data were fit to the

Henderson-Hasselbalch equation (equation 4.1) to determine the pKa values.

4.2.6 Molecular dynamics (MD) simulation

The MD simulation was performed by Mr. Vadim Patsalo using CHARMM22-CMAP molecular mechanics force field with TIP3P water model. The simulation time was 200 nanoseconds.

4.2.7 Poisson Boltzmann (PB) calculations

The PB calculations were performed by Mr. Vadim Patsalo with the Integrated Continuum Electrostatics (ICE) software package. 200 snapshots were collected over the 200 ns simulation. Electrostatic contributions of side chains to protein stability were obtained using standard methods by solving the linearized Poisson−Boltzmann equation using a multigrid finite-difference solver.

4.3 Results

Lysine to methionine mutations at positions 71 or 73 have no significant effect upon the stability of HP36, but were stabilizing at positions 48, 65 and 70 (Table 4-2 and Figure 4-2), increasing ΔG^o of unfolding by 0.7 to 1.2 kcal mol^{-1} depending upon the site and increasing the melting temperature (T_m) by 4 to 9 ℃ (Table 4-2). In the following studies, we will focus on the three stabilizing mutants: K48M, K65M and K70M. The far UV CD spectra of all the mutants are very similar to that of HP36 wildtype indicating that all mutants are folded, and adopt the wildtype helical structure (Figure 4-3). ^{1}H NMR spectra provide further evidence that each mutant is well folded and the core packing is not disrupted (Figure 4-4). As discussed in chapter 2, HP36 has a characteristic chemical shift of 10.5 ppm for the indole N-H resonance of Trp64 in the native state.

A single peak for the protein at 10.5 ppm was observed in the spectrum of each mutant. HP36 also contains several ring-current-shifted methyl resonances which are diagnostic of the fold. One example is the peak below 0 ppm that corresponds to the methyl protons of Val50. Another one is the C4 proton of Phe47 which is strongly shifted in the spectrum of the wildtype appearing at 5.5 ppm due to its interaction with the other two core aromatic residues, Phe51 and Phe58. All of the mutants display those two peaks (Figure 4-4). Moreover, the values of ΔH^o of unfolding are very similar, consistent with no significant perturbation of core packing (Table 4-2).

K48, K65 and K70 are surface exposed or partially exposed charged residues (Table 4-1) and the significant increase in stability would normally be assumed to be due to alteration of native state electrostatics in accord with protein design principles. In order to determine the molecular origins of the significant stabilization, we first measured the pKa's for the acidic residues in the folded states of wildtype HP36 (Figure 4-5) and each mutant (Figure 4-6, 4-7, and 4-8). pKa's are strongly influenced by electrostatic interactions and are sensitive reporters of changes in the electrostatic environment (*132*). The folded state pKa values in wildtype HP36 and the three mutants are summarized in Table 4-3. D44, E45 and D46 in wildtype HP36 have native state pKa's which are lower than model compound values by 0.4 to 1 pKa units indicating that they make favorable native state electrostatic interactions. The effect of the K48M mutant is striking; it significantly perturbs the pKa of E45 increasing it by more than 0.7 pKa units. Its effect on the other Asp and Glu pKa's is much smaller ranging from 0.04 to 0.27 pKa units. The mutation thus results in loss of a favorable native state electrostatic interaction, yet it still stabilizes the protein. The pKa values for the acidic residues in K65M and K70M are very similar to those in wildtype indicating that these two mutations do not significantly alter the native state electrostatic environment sensed by the acidic residues and that these mutations do not introduce new favorable

electrostatic interactions even though they significantly stabilize the domain. The mutants must exert their stabilizing effects by other means; possibilities include alteration of repulsive interactions with other positively charged sites, the introduction of new native state non-electrostatic interactions or the alteration of unfolded state energetics. These possibilities are systematically examined below.

In order to gain more insight we turned to a computational analysis of the domain. A 200 nanosecond long molecular dynamics trajectory in explicit water was generated by Mr. Vadim Patsalo for wild type HP36 using the CHARMM22-CMAP force field and starting from the 1.4 Å resolution crystal structure reported in ref. (*147*). 200 snapshots collected over the 200 nanoseconds were sampled and PB calculations performed on each and the results averaged. Use of an ensemble of structures helps to take into account the effects of protein dynamics. These calculations generate large amounts of data which can be challenging to represent in a simple fashion. We use a "bar code" representation developed by Mr. Patsalo in which the strength of an interaction in a particular snapshot is represented as a colored bar, the shading representing its strength (Figure 4-9). As shown in the figure, D44 and R55 form a very strong favorable interaction, consistent with our observation that D44 has a 1 pKa unit depressed native state pKa value. The calculation also shows that E45 interacts favorably with K48, consistent with the conclusion obtained from the pKa analysis. For the three Lys residues, only K48 is involved in a favorable interaction with E45, the others make no interactions or weak interactions. Taking the desolvation penalty into account, we calculated the mutational free energy for each residue, the effect of going from an uncharged, hydrophobic isostere to a charged sidechain. The free energy is the sum of the desolvation penalty paid by the residue and the interaction free energy. The sign of the mutation free energy indicates whether the transformation is favorable or unfavorable. A

negative mutation free energy means that it is favorable to turn on the charges, while groups which are more favorable as uncharged (hydrophobic) isosteres will have a positive mutation free energy. Table 4-4 summarizes the calculated mutation free energies for K48, K65 and K70 in wildtype HP36. The calculations predict that K65 is not favorable, K70 neutral, and K48 favorable in the native state. It appears that the increased stability of K65M results from reduction of the desolvation penalty at this site upon folding, but the desolvation effects do not account for the effects of K48M or K70M.

To investigate the possibility that the mutations decrease electrostatic repulsion between the positively charged residues, we measured the ΔG^o (mutant) – ΔG^o (wild type) as a function of pH for each mutant (Figure 4-10). If the increased stability is due to the charge repulsion, the stability difference should be maintained at low pH. The difference in stability between K65M and wildtype is decreased by lowering the pH, from 0.6 ± 0.1 kcal mol^{-1} at pH 6 to 0.1 ± 0.1 kcal mol^{-1} at pH 2. The difference between K48M and wildtype also decreases, in this case from 0.8 ± 0.1 kcal mol^{-1} at pH 6 to 0.3 ± 0.1 kcal mol^{-1} at pH 2. This argues that relief of charge repulsion between Lys residues in the native state is not the reason for the increased stability of the K65M and K48M mutations. However, the stability difference between K70M and wildtype is slightly changed as the pH decreases, from 1.0 ± 0.1 kcal mol^{-1} at pH 6 to 0.7 ± 0.1 kcal mol^{-1} at pH 2. This may argue that the increased stability of K70M is due to the removal of charge repulsion in the native state. To further test this possibility, we prepared another mutant, K70E. If elimination of charge repulsions caused the stability increase, one would expect a even larger stability increase for the Glu mutation than for the Met mutation. However, the stability of the K70E mutant is similar to that of wildtype (Table 4-2). Thus, the native state charge repulsions cannot be the reason for the increased stability of K70M. The PB calculations are consistent with the

conclusion that charge repulsion is not the reason for the increased stability of the three mutants (Figure 4-9). No charge repulsions are observed for the three Lys residues.

We next examined potential unfolded state effects. The overwhelming majority of studies of protein stability and folding make the assumption that mutations do not perturb the energetics of the unfolded state. It is extremely difficult to test this assumption and it often goes unchallenged. Fortunately, electrostatic interactions in the unfolded state ensemble can be probed by combining pH-dependent stability measurements with pKa measurements.

We chemically synthesized four peptide fragments from the wildtype sequence and measured the pKa values for the acidic residues that were used to provide a reference state for unfolded state pKa values. If the calculated stability from native state pKa's and peptide model pKa's deviates from the experimental measured stability, there must be unfolded electrostatic interactions that perturb the unfolded state pKa's from the peptide model values. The titration curves for the peptide fragments are shown in Figure 4-11, 4-12, 4-13, and 4-14. The pKa values are summarized in Table 4-5. Since D44, E45 and D46 are close in sequence, it is important to check whether they titrate independently before we perform the pH-dependent stability analysis. The interaction free energy calculation of wildtype shows that these three residues make minimal interactions with each other in the native state (Figure 4-9). In the native structure, the measured distances between the sidechain carbonyl group of two acidic residues are as follows: 6 Å for D44-E45, 7 Å for E45-D46, and 9 Å for D44-D46. The pKa values of Asp and Glu in the peptide fragments (Table 4-5) are very close to the random coil pKa values, 3.9 for Asp and 4.4 for Glu, indicating that D44, E45 and D46 do not interact with each other in this system. To further confirm that, we synthesized another peptide in which the Glu was replaced by Gln (Figure 4-15). The pKa values of the two Asp residues have little change compared to the peptide with the Glu residue

(Table 4-5). Thus, it is safe to perform the pH-dependent stability analysis using the assumption that the three residues titrate independently. Figure 4-16 shows the plots of $\Delta\Delta G^{o}$ vs. pH for HP36 wildtype and the three mutants. No difference between the calculated and measured stability was observed for the wildtype. There are two possibilities to account for this phenomenon. One is that there is no unfolded state electrostatic interaction, and the other is that there are both favorable and unfavorable interactions in the unfolded state which may cancel out each other's effect. No deviation was observed for the K70M and K65M mutants revealing that these two mutations have little, if any, effect on unfolded state electrostatic interactions. In contrast, the K48M mutation has a pronounced effect. The mutation destabilizes the unfolded state ensemble and this effect compensates for its impact on the native state electrostatics.

The K70M mutation has the largest effect upon stability, but appears to have no significant effect upon unfolded or folded state electrostatic interactions. Analysis of the high resolution crystal structure of the mutant reveals an unexpected result; the mutation leads to the introduction of new packing interactions with the phenylalanine core (Figure 4-17). The hydrophobic Met sidechain folds back so that it makes extensive contacts with the phenylalanine core. The calculated van der Waals forces between Met70 and the three core aromatic residues in the K70M mutant are ~ 1 kcal mol^{-1} larger than those between Lys70 and the core aromatic residues in wildtype. Our structural and computational analysis predicts that mutation to a smaller non-polar side chain will have less effect at this potion. Consequently, we examined a K70A mutation. The mutant folds cooperatively (Figure 4-18), but it has a free energy of folding and a T_m which are very similar to wild type and which are noticeably less than the K70M (Table 4-2).

4.4 Discussion

In this chapter, we examined the origin of the stability increases induced by three surface mutations in HP36. Contrary to expectations, each mutant exerts its effect by very different mechanisms. We considered the possibility of altering native state electrostatics as well as 1) reduction of the desolvation penalty at the mutation site upon folding; 2) introduction of new favorable core packing interactions in the native state; 3) removal of favorable electrostatic interactions in both the native and unfolded state, with the magnitude of unfolded state effect being larger. The stabilization of K65M mutation is due to the reduction of the desolvation penalty paid by Lys65 in the native state. The stabilization of K70M mutation is from the introduction of new hydrophobic interactions between Met70 and the hydrophobic core in the native state. The stabilization of K48M mutation is caused by the removal of the salt bridge between Lys48 and Glu45 in the native state and the favorable electrostatic interaction(s) in the unfolded state. The potential partner is Asp44 which will be demonstrated in detail in chapter 5. Thus this mutation stabilizes the protein by destabilizing both the folded and unfolded state, a mechanism which would be impossible to detect using structure based design algorithms.

More and more attention is being paid to the protein unfolded state since it is not random coil and because of the interests in intrinsic disordered proteins. We have shown that electrostatic interactions can play an important role in the energetics of unfolded state in HP36. But this is not a unique case. Unfolded state electrostatic interactions are found to be formed in Barnase (*215*), Staphylococcal nuclease (SNase) (*100*), mutants of the B1 domain of protein G (*104*), turkey ovomucoid third domain (OMTKY3) (*216*), CI2 (*217*), some leucine zippers (*218*), the drkN SH3 domain (*219*), the BBL domain (*105*), ribonuclease Sa (RNase Sa) (*220*), ribonuclease H (RNase H) (*103*), thioredoxin (*221*), N-terminal domain of L9 (NTL9) (*109*) *etc*. It has been suggested that

when the hydrophobic and hydrogen bonding interactions that stabilize the folded state are disrupted, the unfolded polypeptide chain rearranges to compact conformations with favorable long-range electrostatic interactions (*60*). These charge-charge interactions might even be more favorable in the unfolded state than they are in the native state and will help determine the properties of the unfolded state. For example, a charge-reversal mutation (D17K) was made to stabilize the native surface electrostatics of RNase Sa but the mutant was less stable than the wildtype by 1 kcal mol^{-1} (*60*), and a lysine to methionine surface mutant has been shown to stabilize NTL9 by 1.7 kcal mol^{-1} due to unfolded state effects (*101*). HP36 is another example. K48 forms favorable electrostatic interactions in both the native and the unfolded state, and the magnitude of the unfolded state interaction strength is larger than that of the native state interaction.

A peptide containing only the first 13 resides of HP36 cannot fold in isolation (*155*) and our pKa measurements of the peptide fragments display only slight changes between the peptides with and without K48. Thus the unfolded state interactions involving K48 are not due to just local sequence effects. The specific interactions presented here need to be understood in the context of the overall structure propensity of the unfolded state.

Table 4-1. Solvent accessible surface area calculations for the five lysine residues of HP36.

Residue	Sidechain exposure (Å^2)	% sidechain exposure[1]	Exposure of NH_3^+ (Å^2)	% NH_3^+ exposure[2]
K48	95.2 ±15.0	58.3 ±9.2	38.1 ±8.5	76.4 ±17.0
K65	87.1 ±6.9	53.3 ±4.2	44.4 ±3.8	89.1 ±7.7
K70	63.5 ±11.9	38.9 ±7.3	40.2 ±7.6	80.5 ±15.2
K71	117.4 ±18.8	71.9 ±11.5	42.6 ±4.9	85.4 ±9.8
K73	72.4 ±12.7	44.3 ±7.8	37.9 ±6.9	75.9 ±13.8

The numbers before the ± symbol are the average values of the 200 snapshot structures, and those after the ± symbol are the standard deviations.

1) It is given by the ratio of side-chain surface area to "random coil" value, that is the average solvent-accessible surface area of X in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations (*214*).

2) It is given by the ratio of NH_3^+ surface area to the maximal value observed for the NH_3^+ group in the calculation.

Table 4-2. Thermodynamic parameters for the unfolding of HP36 wildtype and the mutants

Protein	T_m (^oC)	$\Delta H^o(T_m)$ (kcal mol^{-1})	ΔG^o (kcal mol^{-1})	m-value (kcal mol^{-1} M^{-1})	C_M[1] (M)
HP36 WT	73.0 ±0.2	31.82 ±0.33	3.17 ±0.06	0.52 ±0.01	6.2
K48M	78.0 ±0.4	34.07 ±0.38	4.11 ±0.08[3]	0.52 ±0.01[2]	7.9
K65M	77.2 ±0.6	33.20 ±0.65	3.87 ±0.08[3]	0.52 ±0.01[2]	7.2
K70M	82.2 ±0.7	33.44 ±0.44	4.37 ±0.08[3]	0.52 ±0.01[2]	8.4
K71M	71.2 ±0.2	31.86 ±0.36	3.00 ±0.06	0.49 ±0.01	6.1
K73M	69.5 ±0.2	30.42 ±0.36	3.05 ±0.05	0.49 ±0.01	6.1
K70E	66.7 ±0.2	31.96 ±0.49	3.20 ±0.05	0.49 ±0.01	6.5
K70A	67.6 ±0.4	31.37 ±0.73	3.07 ±0.06	0.51 ±0.01	6.0

The numbers listed after the ± symbol are the standard errors to the fit.
1) The C_M value is determined by calculating the derivative of the plot of CD signals vs. [urea].
2) The high thermal stabilities prevent an accurate estimate of the m-value, so the value of HP36 wildtype is used.
3) The ΔG^o values are calculated using: $\Delta G^o = C_M \times$ m-value.

Table 4-3. Native state pKa values for the acidic residues in HP36 wildtype and the mutants. The uncertainties are the standard error to the fits. Experiments were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

Residue	Native state pKa's for HP36	Native state pKa's for K70M	Native state pKa's for K65M	Native state pKa's for K48M
D44	3.04 ±0.12	3.03 ±0.07	3.03 ±0.06	3.23 ±0.07
E45	3.95 ±0.02	4.13 ±0.10	4.15 ±0.09	4.68 ±0.09
D46	3.44 ±0.11	3.34 ±0.05	3.33 ±0.05	3.38 ±0.05
E72	4.37 ±0.03	4.43 ±0.02	4.36 ±0.14	4.41 ±0.03
C-terminus	3.17 ±0.14	3.13 ±0.08	2.91 ±0.10	2.90 ±0.06

Table 4-4. Calculated mutation free energy of the three Lys residues in HP36 wildtype. The values were determined via PB calculations and represent the average over 200 snapshots collected during the 200 nanoseconds MD trajectory.

Residue	Desolvation penalty (kcal/mol)	Interaction free energy (kcal/mol)	Mutation free energy (kcal/mol)
K65	1.10	-0.03	1.07
K70	2.31	-2.19	0.12
K48	1.70	-4.70	-3.00

Table 4-5. pKa values for the acidic residues of the peptide fragments. The uncertainties represent the standard error to the fits. Experiments were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

Residue number	Peptide sequence	pKa (D44)	pKa (E45)	pKa (D46)	pKa (E72)	pKa (C-terminus)
41-53		4.00 ±0.08	4.50 ±0.04	3.86 ±0.04		
41-49 K48M		3.97 ±0.06	4.49 ±0.05	4.06 ±0.09		
41-53 E45Q		3.86 ±0.05		4.02 ±0.07		
70-76					4.15 ±0.03	3.22 ±0.05
71-76					4.28 ±0.02	3.29 ±0.04

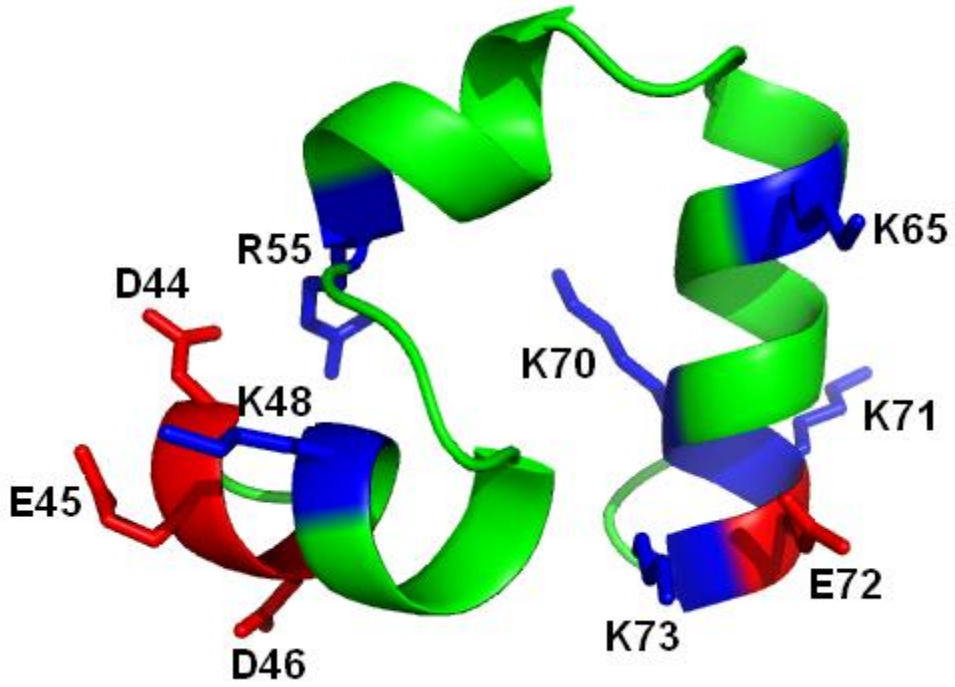


Figure 4-1. A ribbon diagram of the structure of HP36 generated using PyMol (PDB code: 1YRF). The four acidic residues are shown in red and six basic residues in blue.

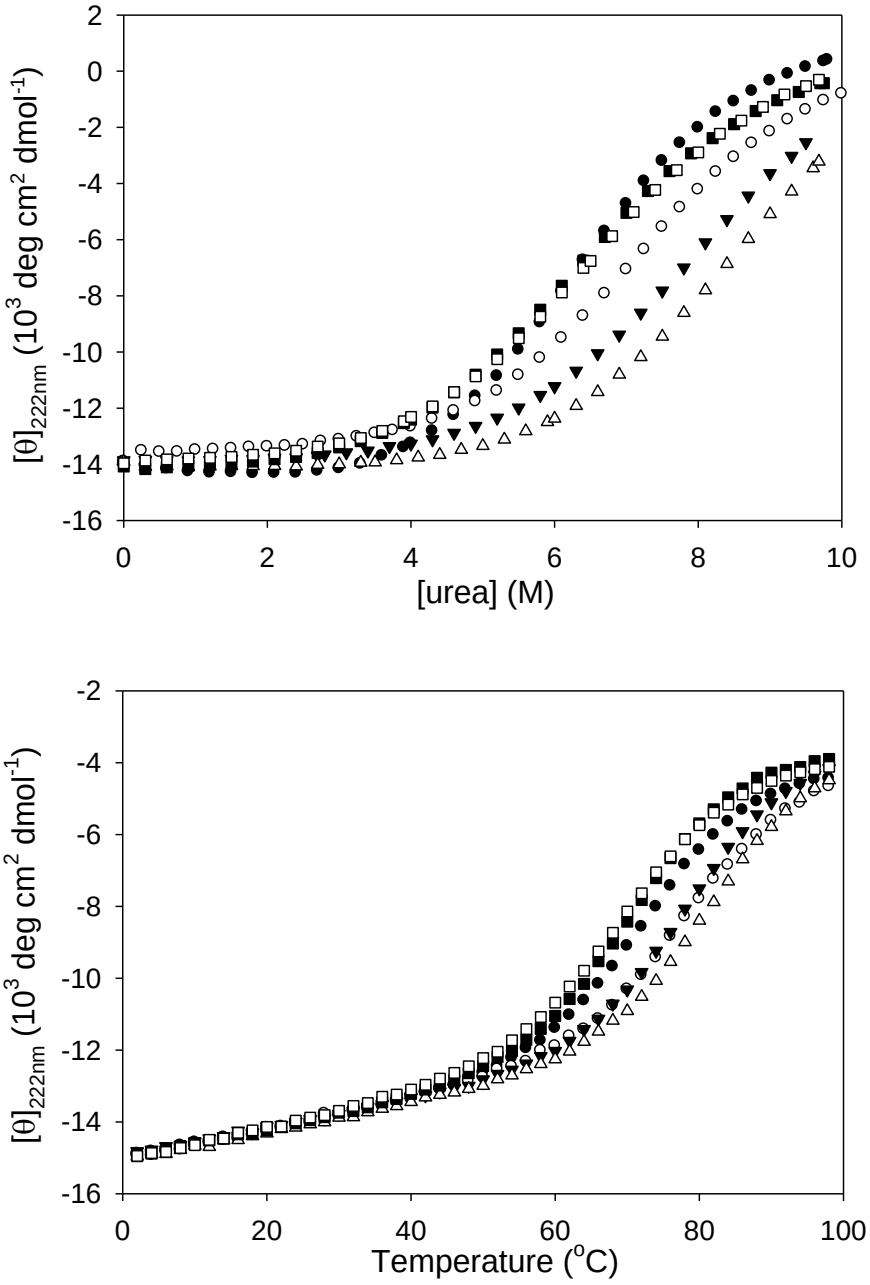


Figure 4-2. Urea and temperature induced unfolding transitions of wildtype HP36 (●), K65M (○), K48M (▼), K70M (△), K71M (■), and K73M (□). All spectra were collected in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0 buffer using a 1.0 cm cuvette. Urea unfolding experiments were conducted at 25 °C.

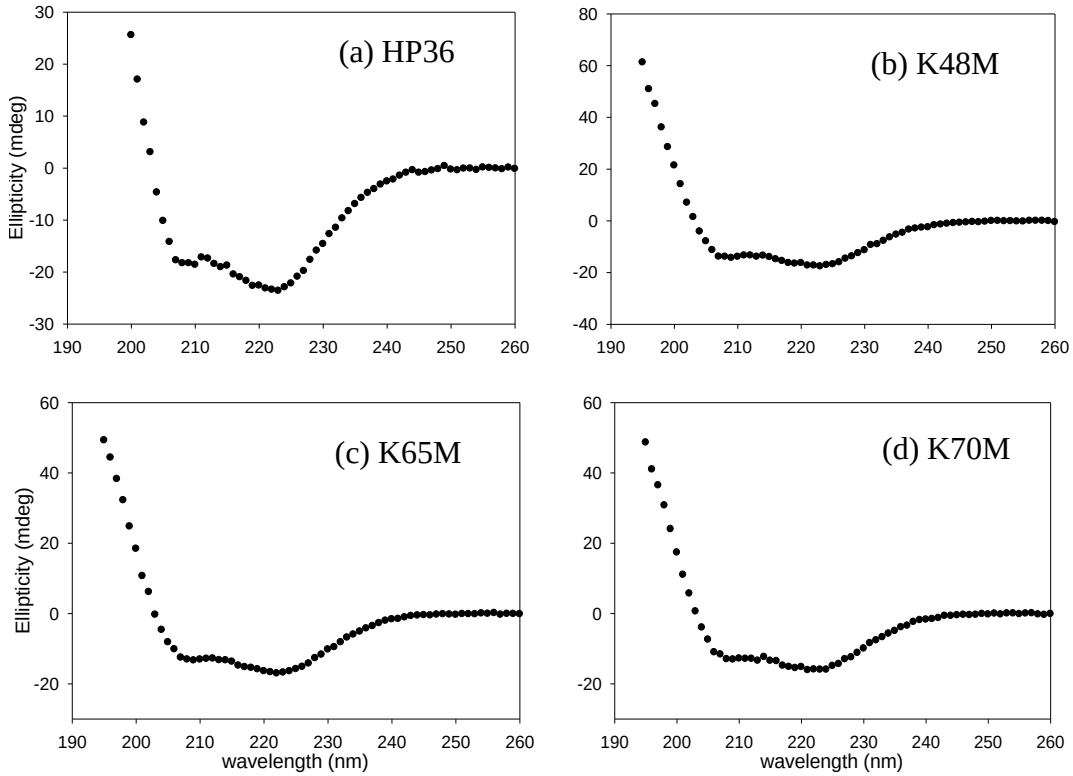


Figure 4-3. CD monitored far-UV wavelength spectra of (a) HP36, (b) K48M, (c) K65M and (d) K70M. All spectra were collected at 25 ^{o}C in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0 buffer using a 0.1 cm cuvette.

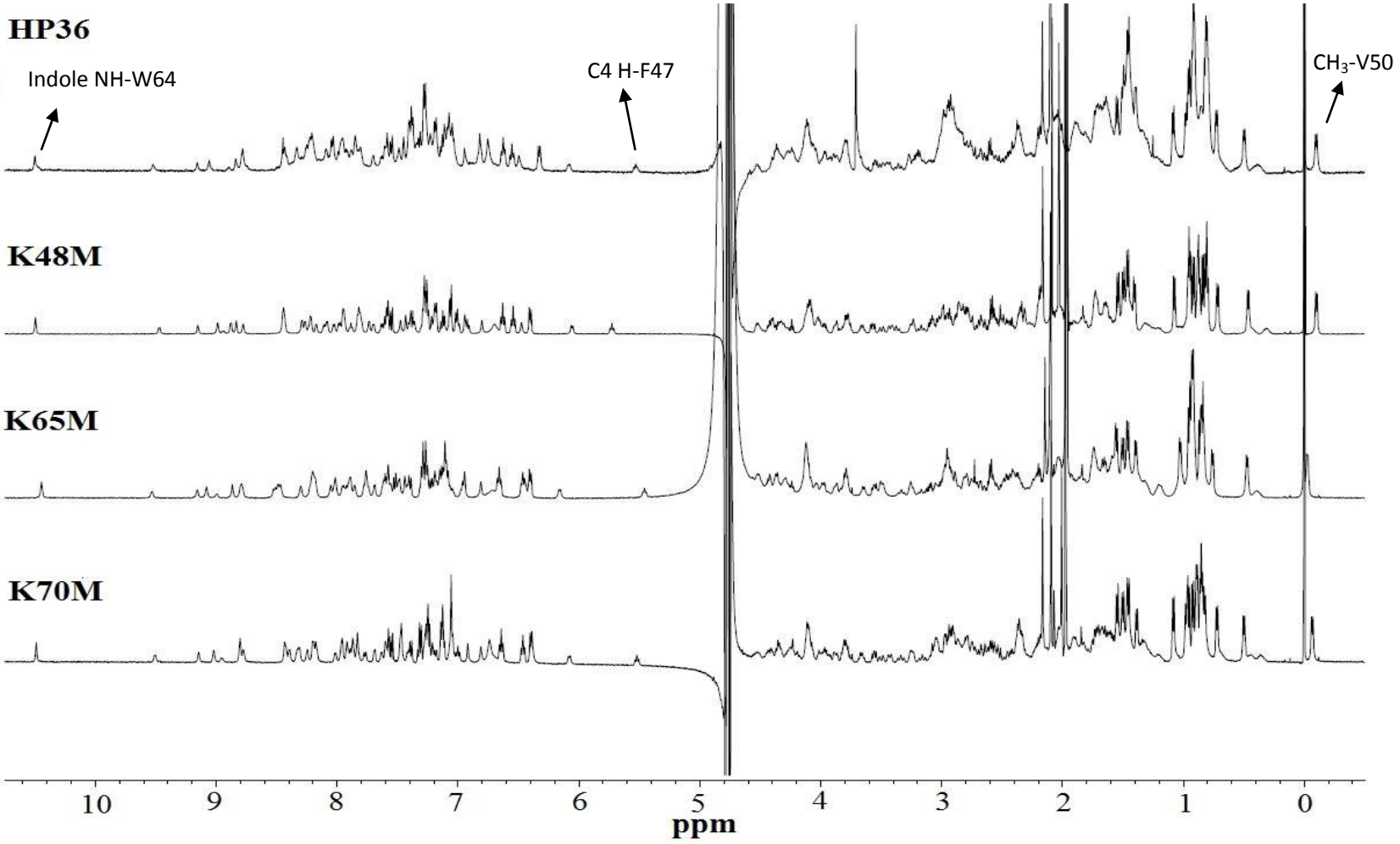


Figure 4-4. Stacked plot of ^{1}H NMR spectra of HP36, K48M, K65M and K70M. The sharp peak at 0 ppm is the chemical shift standard, DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). The three characteristic peaks are labeled. Spectra were recorded at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0 buffer.

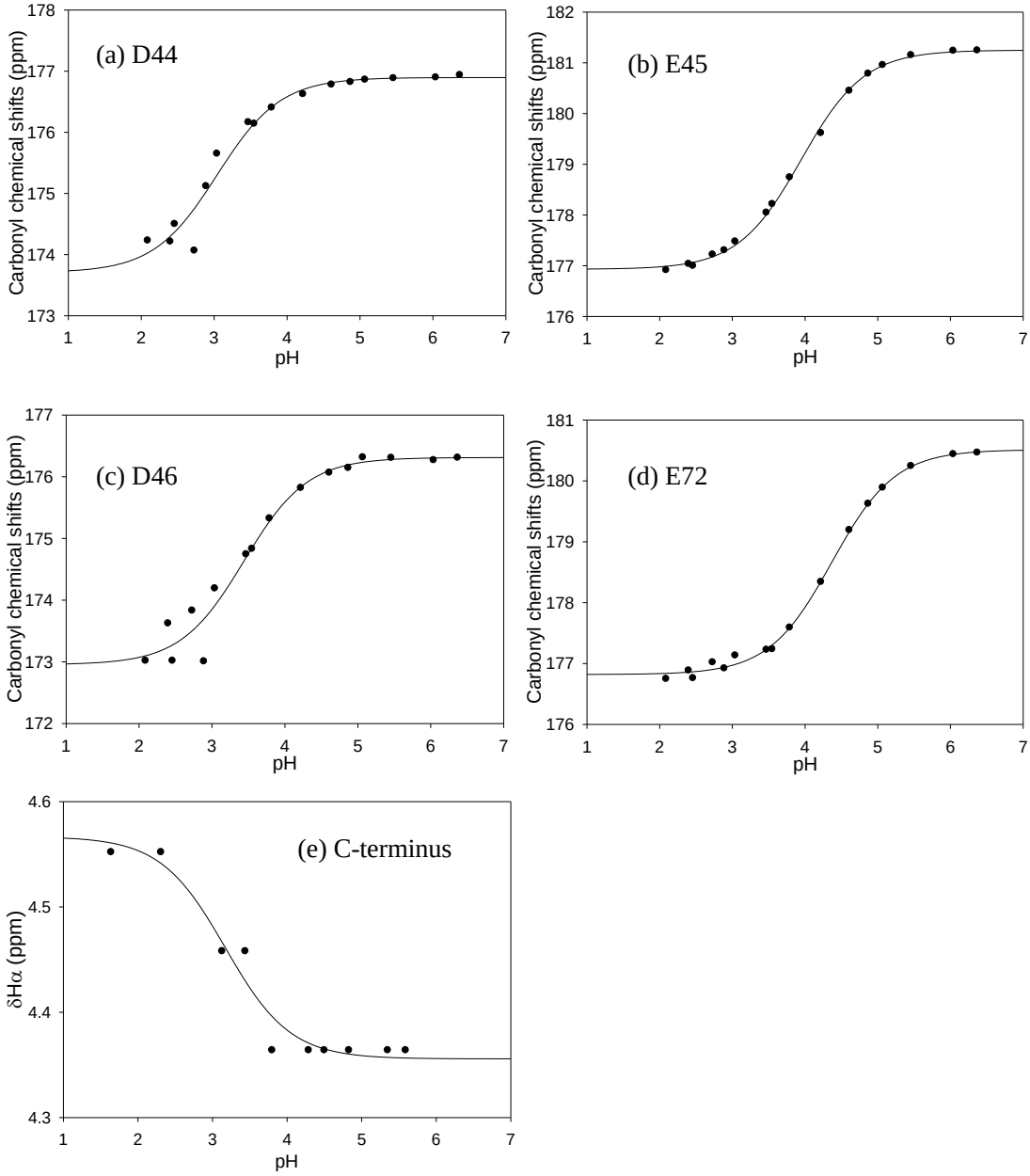


Figure 4-5. pH titration for the native state of HP36 wildtype. Carbonyl carbon chemical shifts were recorded as a function of pH for D44, E45, D46 and E72, while α-proton chemical shifts were recorded for the C-terminus. The solid curve in each graph represents the fit to equation 4.1. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

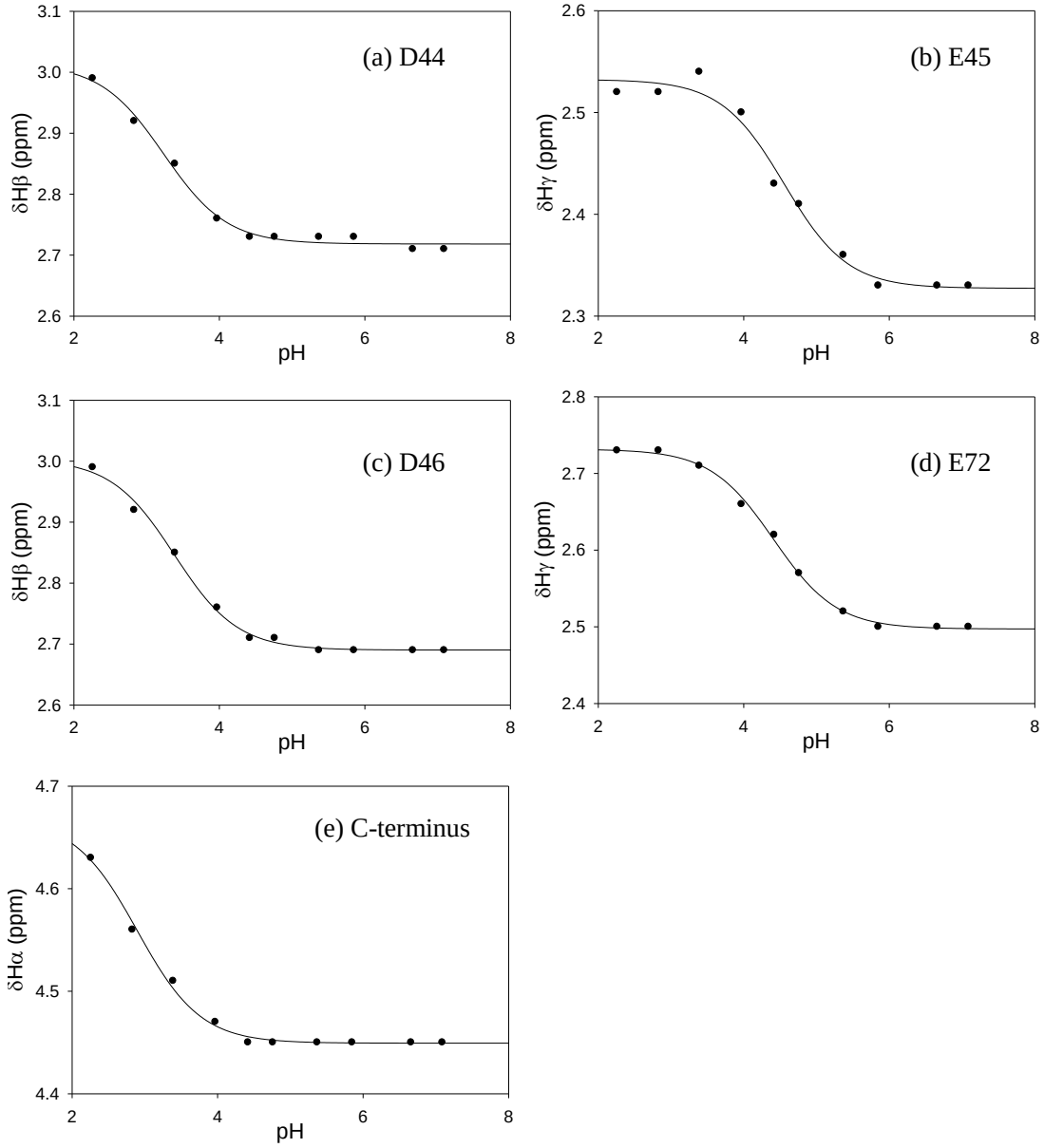


Figure 4-6. pH titration for the native state of the K48M mutant. β-proton chemical shifts were recorded as a function of pH for D44 and D46, γ-proton shifts for E45 and E72, and α-proton shifts for the C-terminus. The solid curve in each graph represents the fit to equation 4.1. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

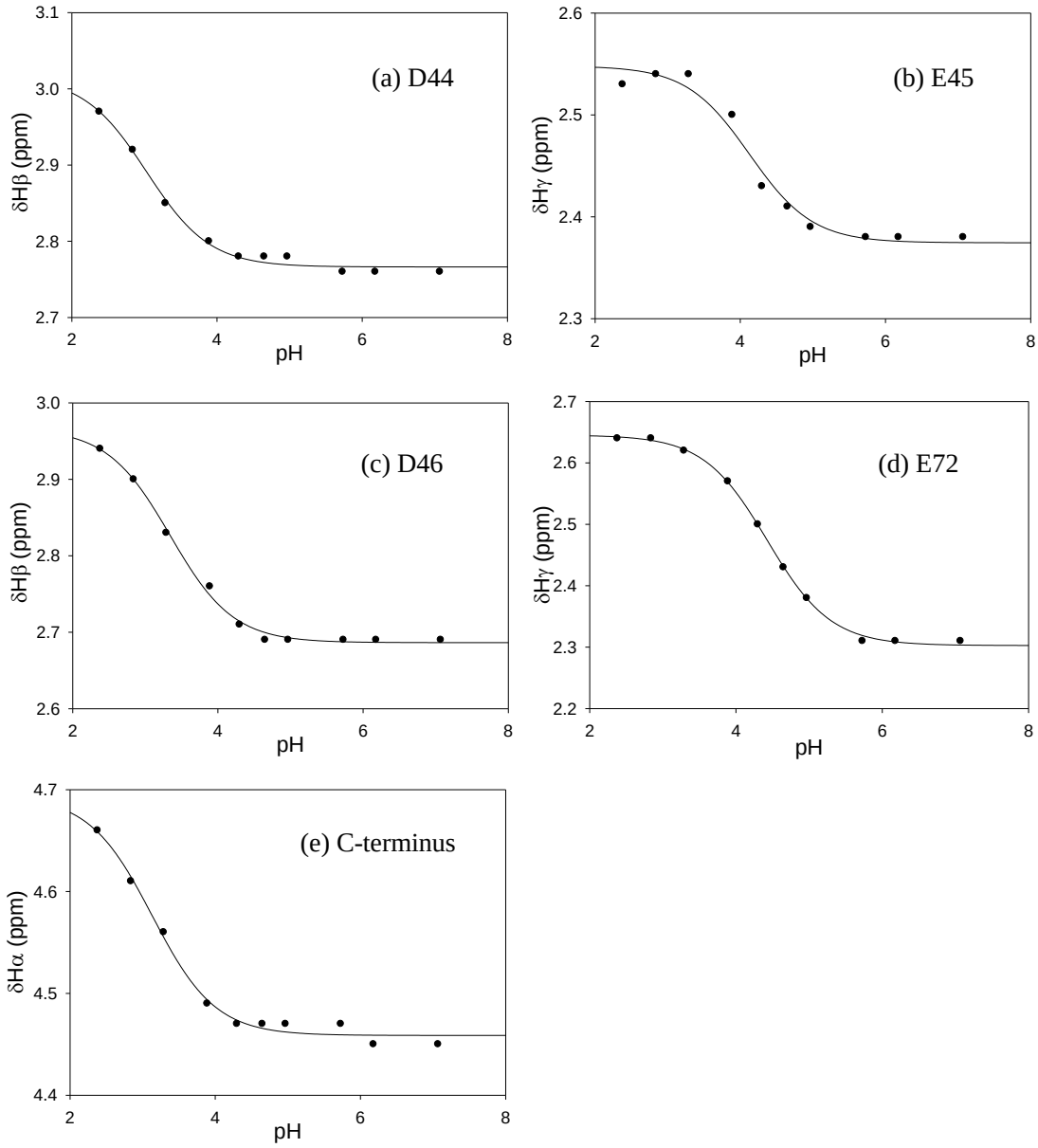


Figure 4-7. pH titration for the native state of the K70M mutant. β-proton chemical shifts were recorded as a function of pH for D44 and D46, γ-proton shifts for E45 and E72, and α-proton shifts for the C-terminus. The solid curve in each graph represents the fit to equation 4.1. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

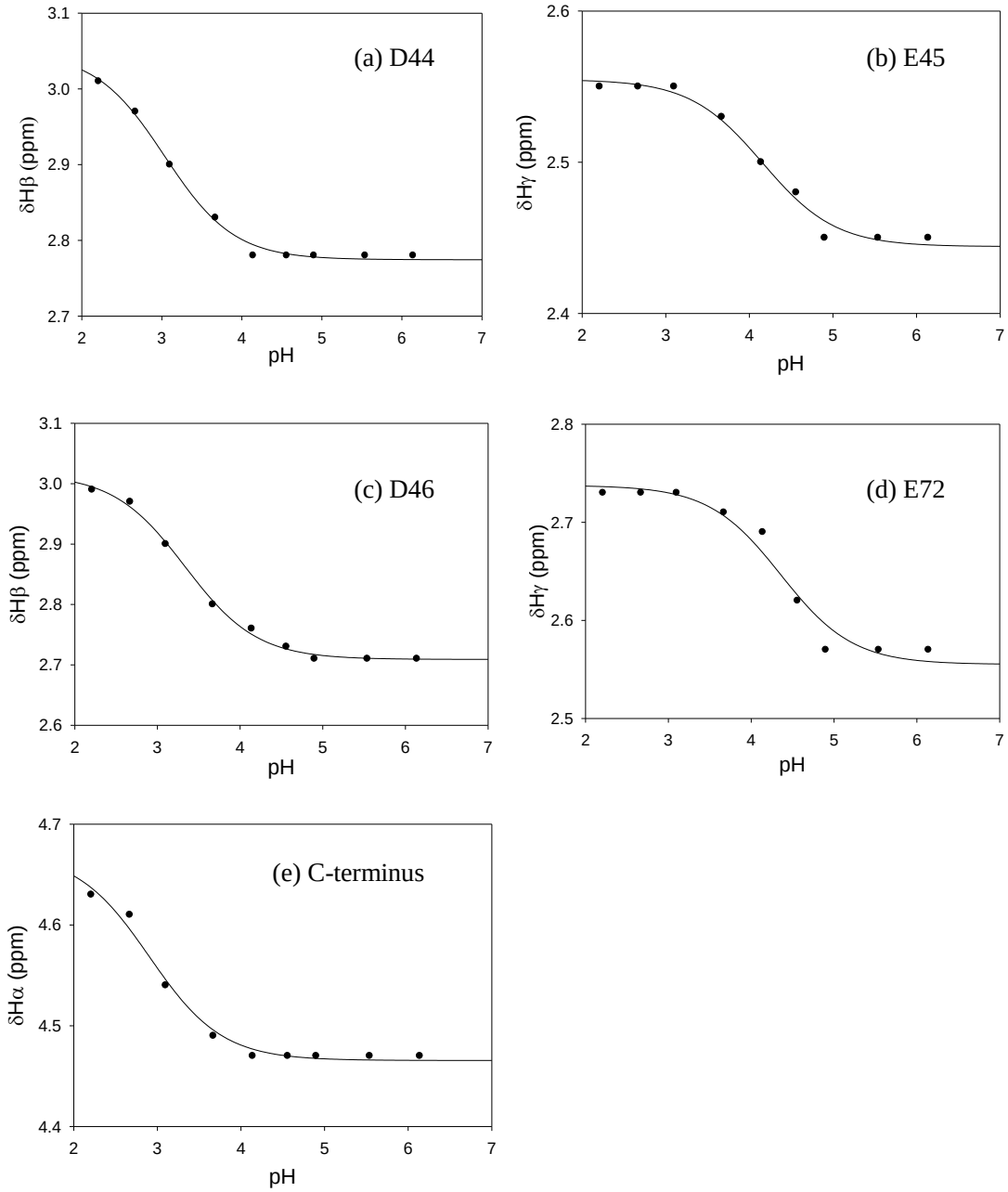
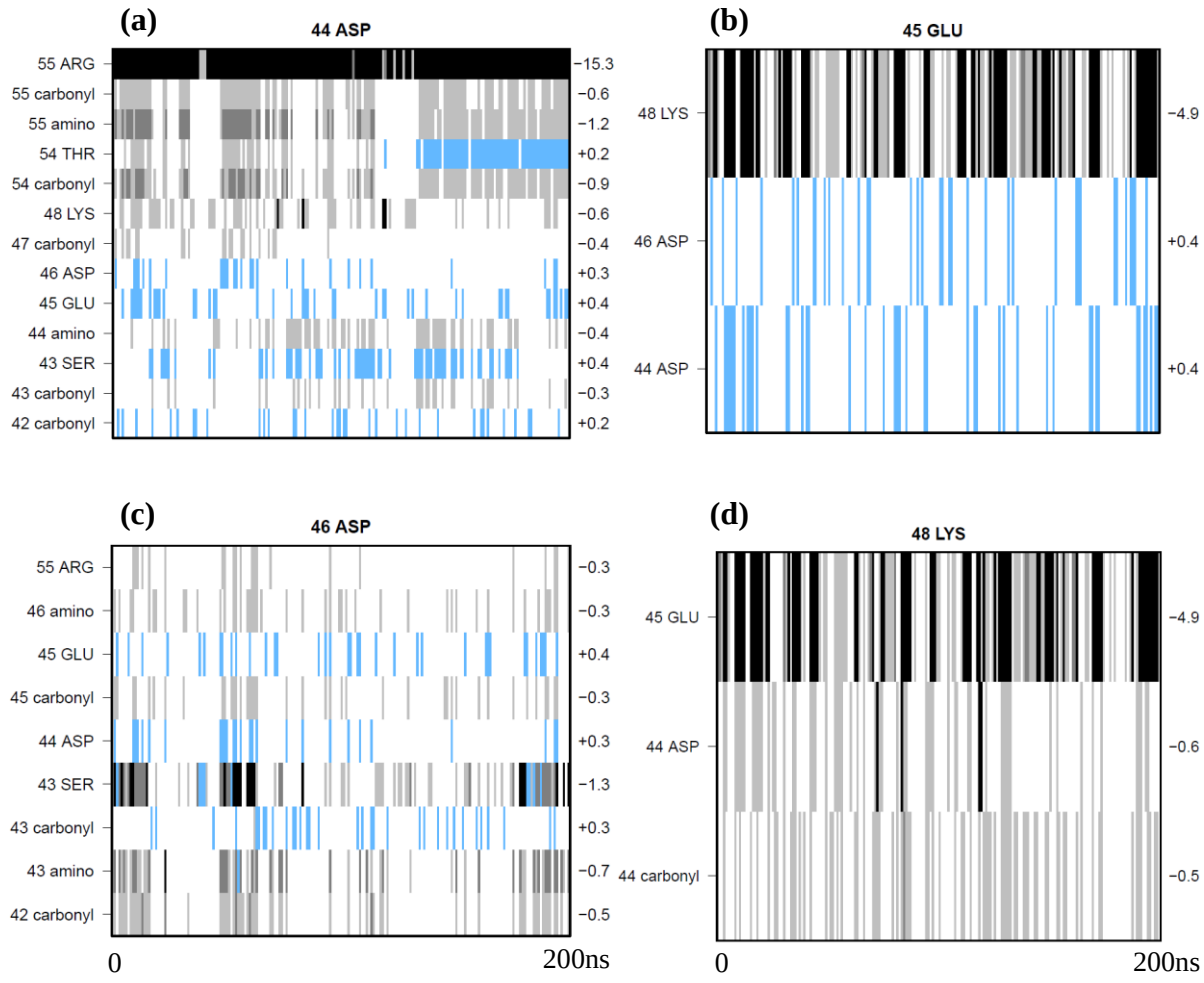


Figure 4-8. pH titration for the native state of the K65M mutant. β-proton chemical shifts were recorded as a function of pH for D44 and D46, γ-proton shifts for E45 and E72, and α-proton shifts for the C-terminus. The solid curve in each graph represents the fit to equation 4.1. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.


(a)
44 ASP
55 ARG
55 carbonyl
55 amino
54 THR
54 carbonyl
48 LYS
47 carbonyl
46 ASP
45 GLU
44 amino
43 SER
43 carbonyl
42 carbonyl
−15.3
−0.6
−1.2
+0.2
−0.9
−0.6
−0.4
+0.3
+0.4
−0.4
+0.4
−0.3
+0.2
(b)
45 GLU
48 LYS
46 ASP
44 ASP
−4.9
+0.4
+0.4
(c)
46 ASP
55 ARG
46 amino
45 GLU
45 carbonyl
44 ASP
43 SER
43 carbonyl
43 amino
42 carbonyl
−0.3
−0.3
+0.4
−0.3
+0.3
−1.3
+0.3
−0.7
−0.5
0
200ns
(d)
48 LYS
45 GLU
44 ASP
44 carbonyl
−4.9
−0.6
−0.5
0
200ns

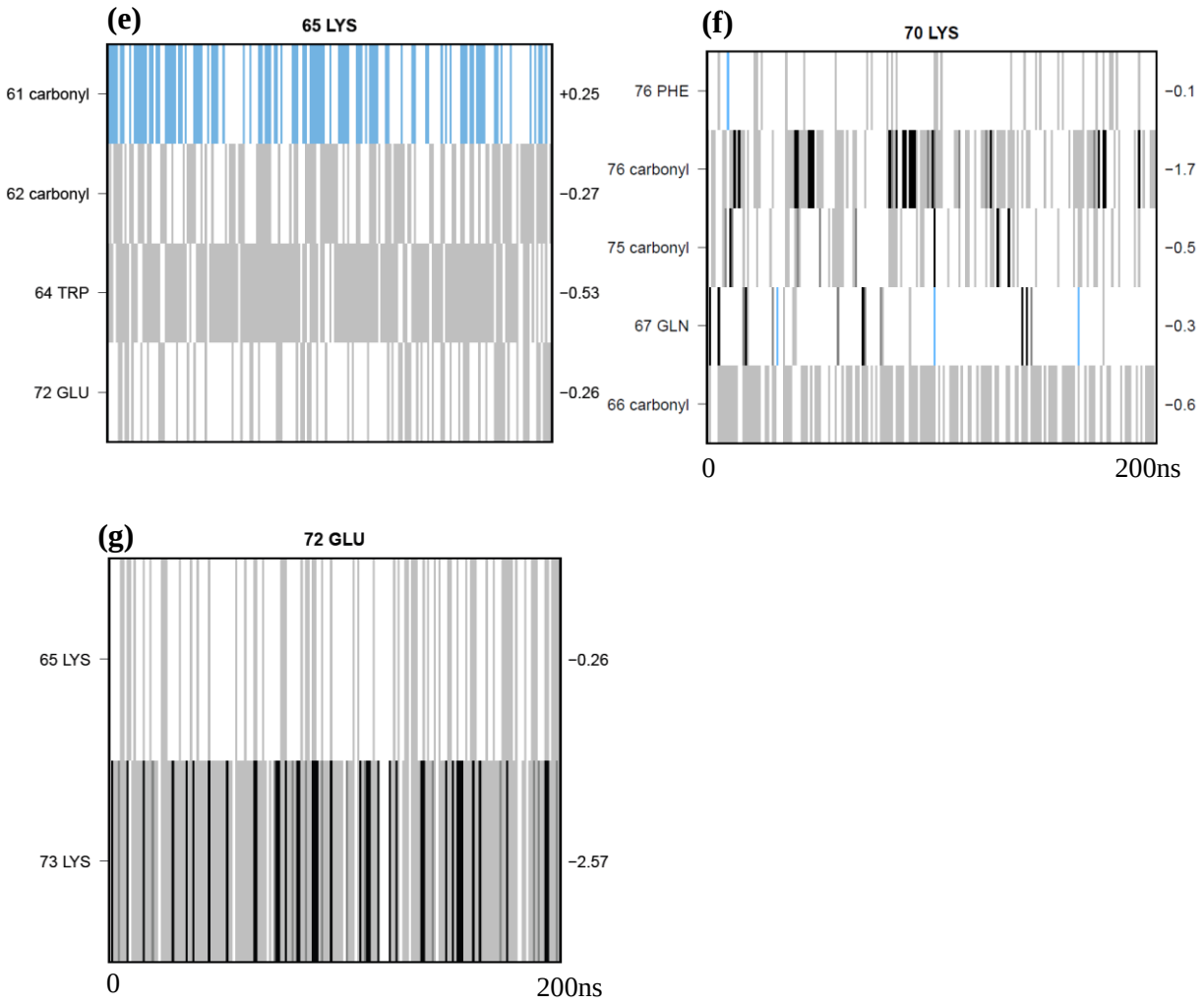


Figure 4-9. Color coded bar graphs showing the native state interactions of each snapshot in the 200 ns simulation trajectory involving (a) D44, (b) E45, (c) D46, (d) K48, (e) K65, (f) K70, and (g) E72. The left vertical axis represents the components involved in the interactions and right axis the average interaction free energy in kcal mol^{-1}. A negative number means favorable interaction and positive number unfavorable interaction. The black color indicates strong favorable interaction, grey weak favorable interaction, and cyan unfavorable interaction. Calculations were kindly conducted by Mr. Vadim Patsalo and reproduced here with his permission.

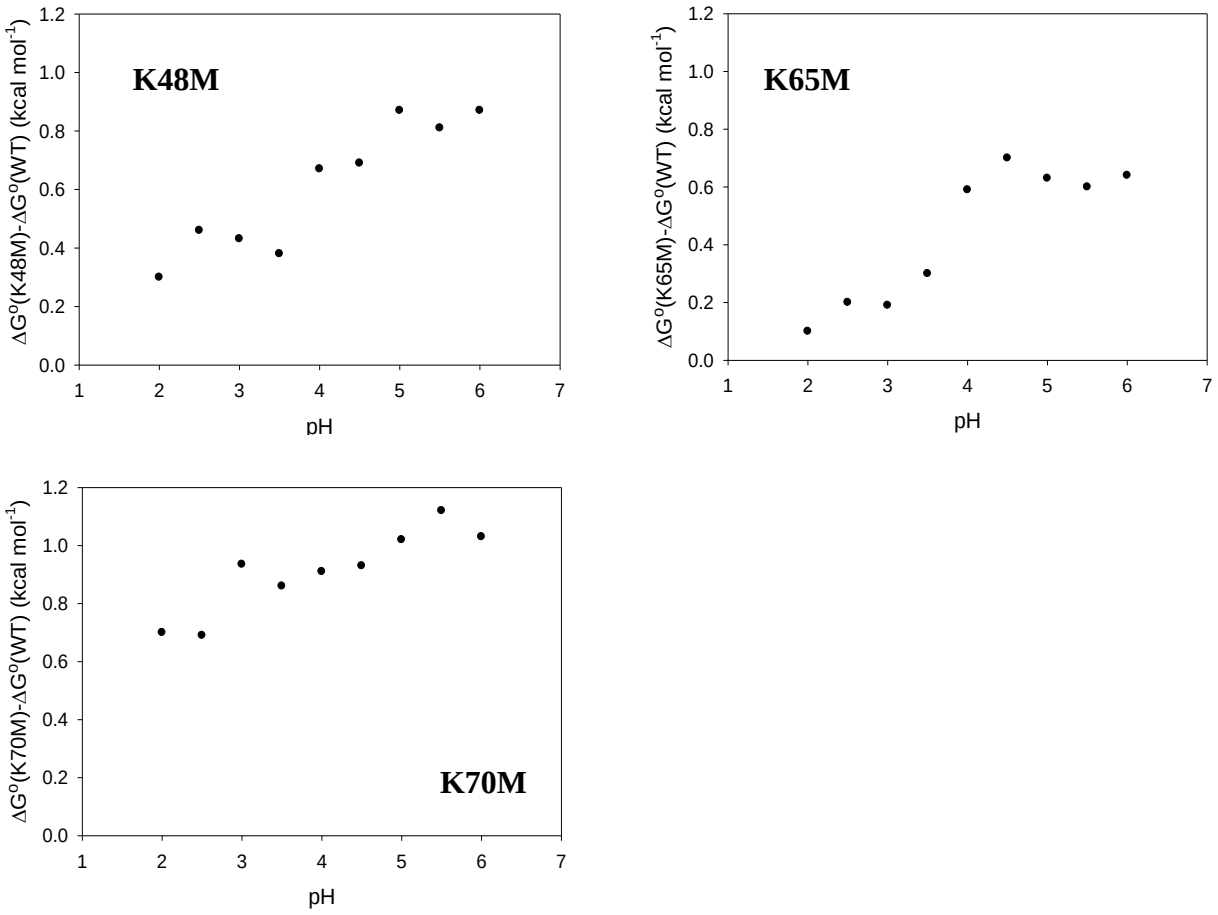


Figure 4-10. Protein stability difference between the mutants and the wildtype as a function of pH. ΔG^o values at pH 3.5 and higher were determined by urea unfolding studies at 25 ^oC. Values at pH 3 and below were determined by thermal unfolding experiments using the Gibbs-Helmholtz equation to extrapolate to 25 ^oC. A value of ΔC_p^o = 0.38 kcal mol^{-1} K^{-1} was used. All samples contained 10 mM sodium acetate and 150 mM sodium chloride.

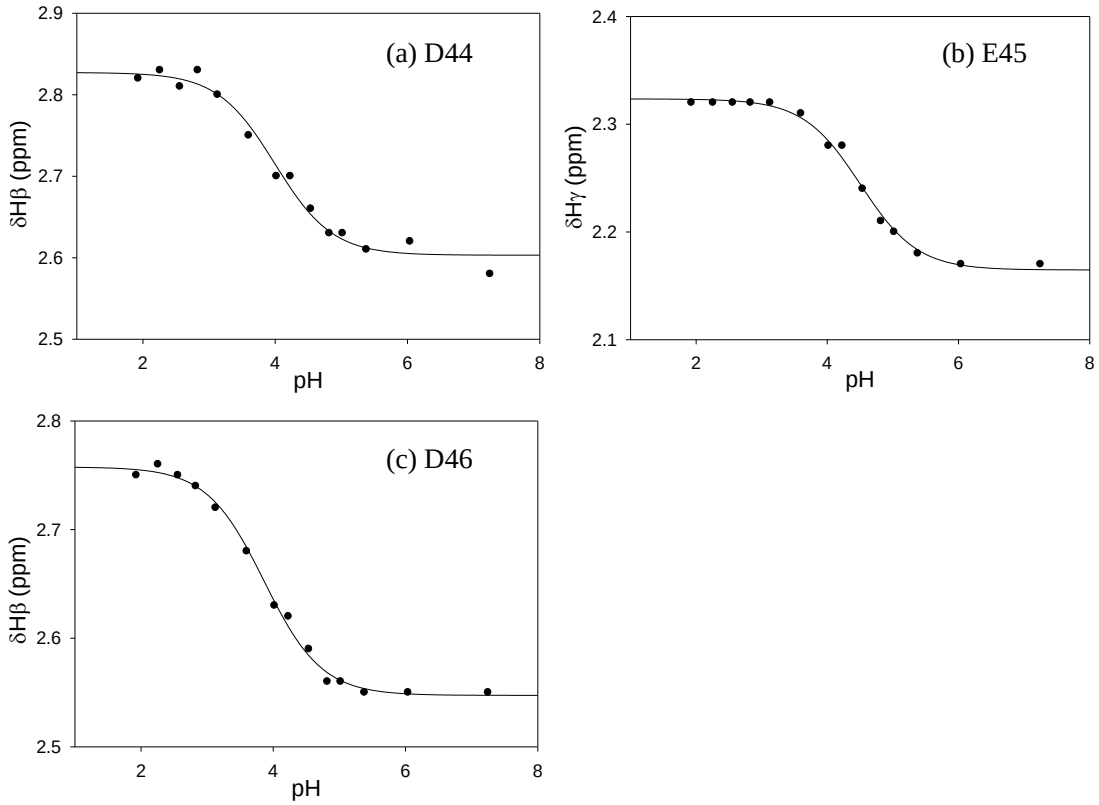


Figure 4-11. pH titration for the MLSDEDFKAVFGM peptide. β-proton chemical shifts were recorded as a function of pH for the two Asp residues, and γ-proton shifts for Glu. The solid curve in each graph represents the fit to equation 4.1. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has a free N-terminus and an amidated C-terminus.

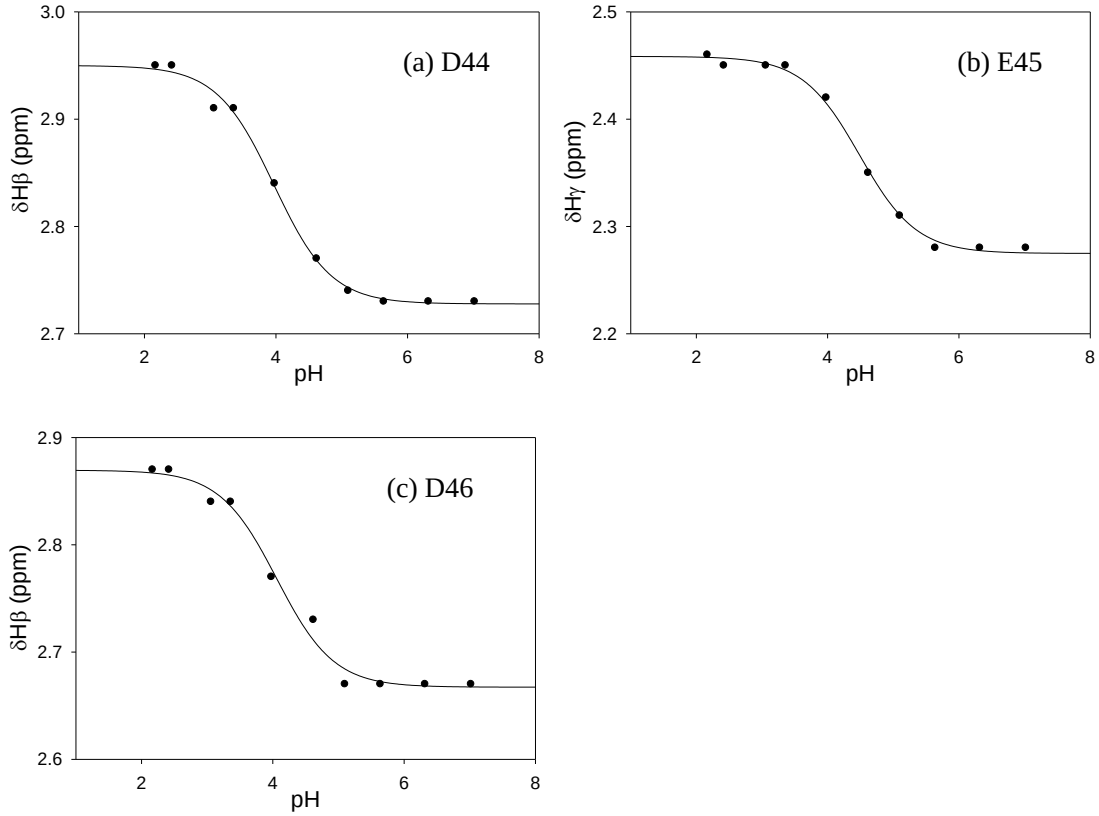


Figure 4-12. pH titration for the MLSDEDFMA peptide. β-proton chemical shifts were recorded as a function of pH for each Asp, and γ-proton shifts for Glu. The solid curve in each graph represents the fit to equation 4.1. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has a free N-terminus and an amidated C-terminus.

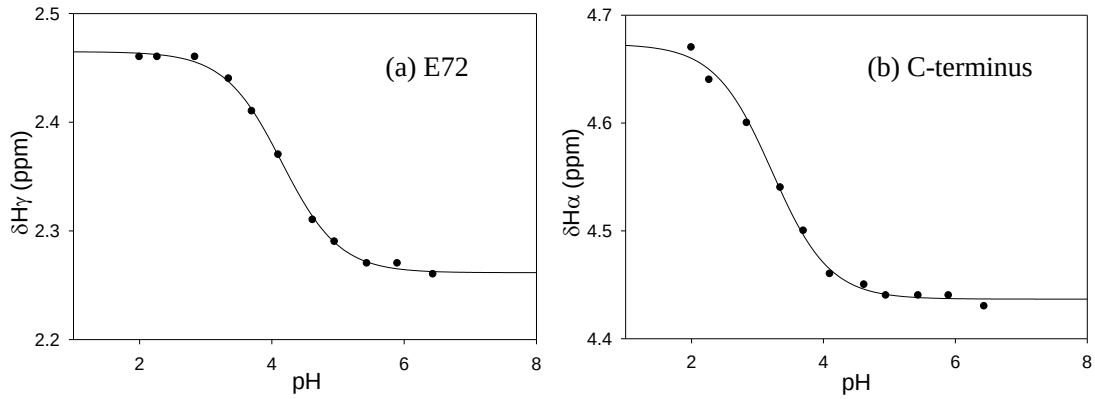


Figure 4-13. pH titration for the KKEKGLF peptide. γ -proton chemical shifts were recorded as a function of pH for Glu, and α-proton shifts for the C-terminus. The solid curve in each graph represents the fit to equation 4.1. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has an acetylated N-terminus and a free C-terminus.

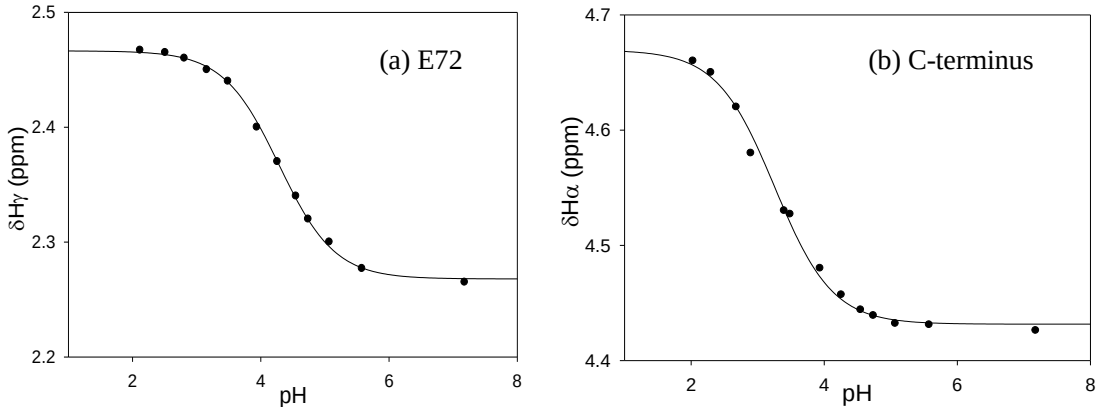


Figure 4-14. pH titration for the KEKGLF peptide**.** γ -proton chemical shifts were recorded as a function of pH for Glu, and α-proton shifts for the C-terminus. The solid curve in each graph represents the fit to equation 4.1. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has an acetylated N-terminus and a free C-terminus.

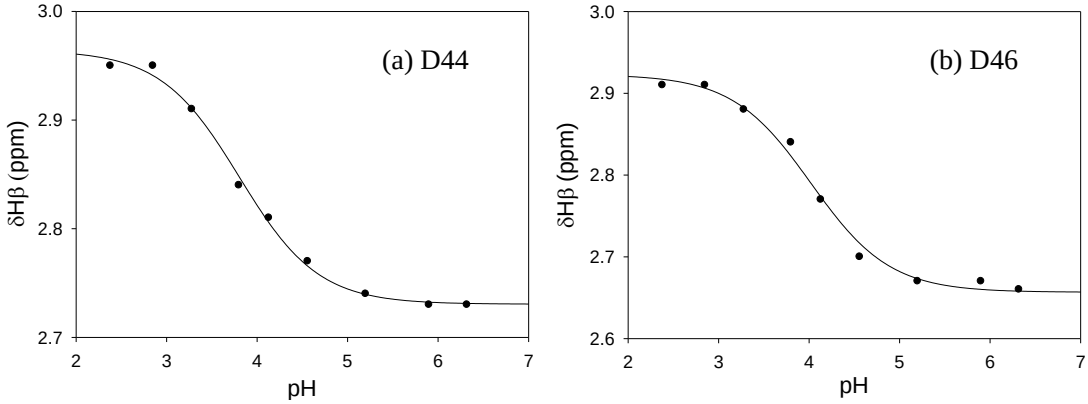


Figure 4-15. pH titration for the MLSDQDFKAVFGM peptide . β-proton chemical shifts were recorded as a function of pH for each Asp. The solid curve in each graph represents the fit to equation 4.1. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has a free N-terminus and an amidated C-terminus.

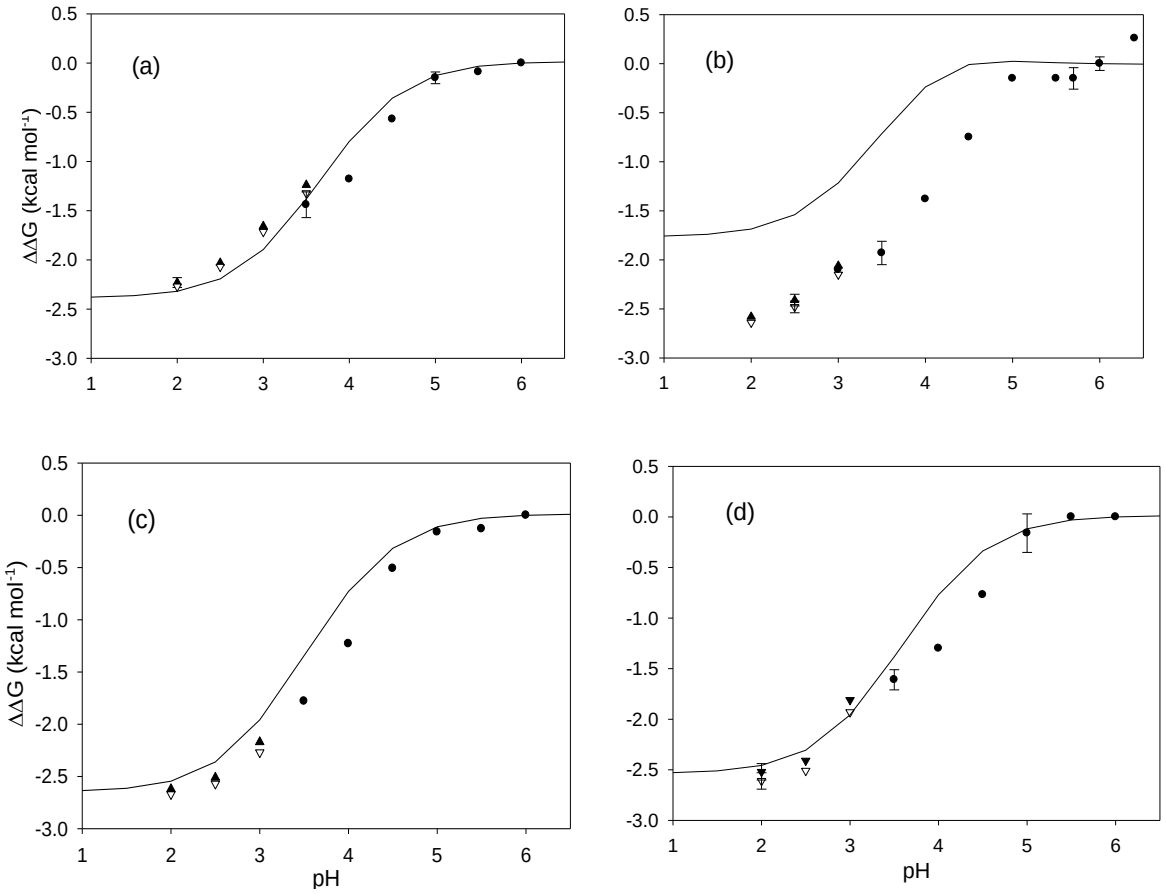


Figure 4-16. pH-dependent stability analysis of (a) HP36 wt, (b) K48M, (c) K65M and (d) K70M. The continuous lines represent the calculated stability change using known native state pKa measurements and values for peptide fragments. The filled circles represent the measured stability change from urea induced unfolding experiments. The filled triangles represent the calculated stability change from thermal unfolding experiments using Gibbs-Helmholtz equation with a ΔC_p^o value 0.38 kcal mol^{-1} K^{-1}. The open triangles represent the calculated stability change from thermal unfolding experiments using Gibbs-Helmholtz equation with a ΔC_p^o value 0.43 kcal mol^{-1} K^{-1} (*198*). The error bars represent the standard deviations of three repeated experiments. Stabilities are values at 25 ^{o}C. Samples contained 10 mM sodium acetate and 150 mM sodium chloride.

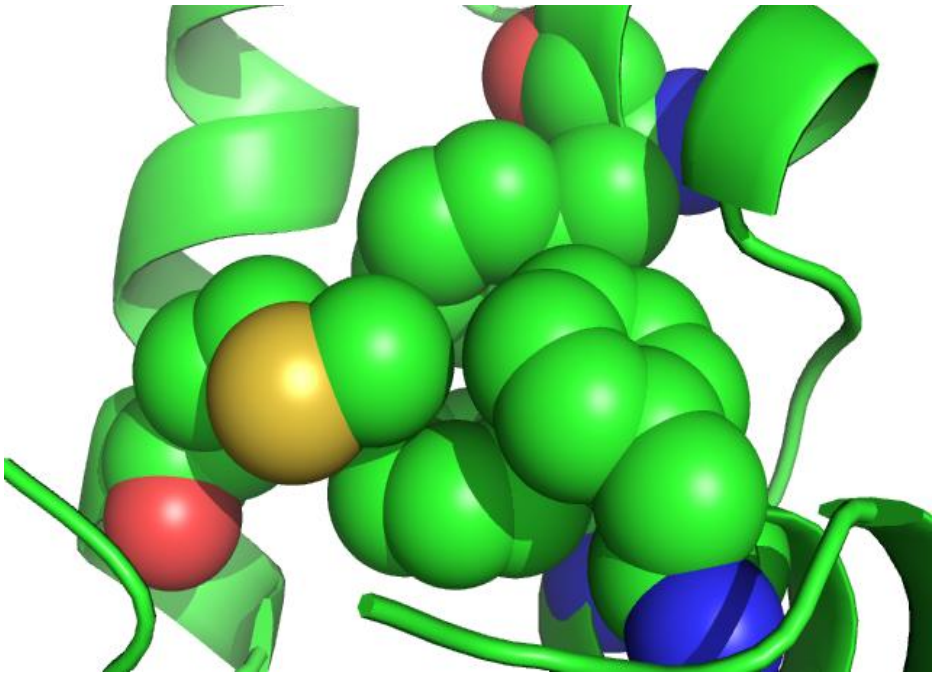


Figure 4-17. A ribbon diagram of the crystal structure of the K70M mutant showing the new packing interaction made by the Met sidechain. The figure was generated using PyMol from the X-ray structure determined by Bi *et al* (*147*). The Met70 and the three Phe residues in the hydrophobic core are shown in space-filling format.

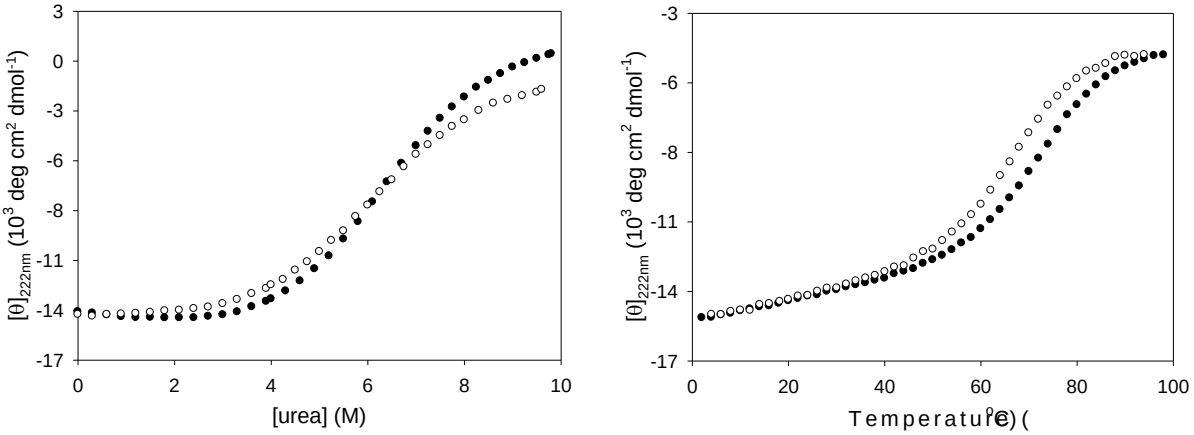


Figure 4-18. Urea induced and thermal unfolding transitions of wildtype HP36 (●) and K70A (○). All spectra were collected in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0 buffer using a 1.0 cm cuvette. Urea unfolding was conducted at 25 ^{o}C.

5. Uncovering electrostatic interactions in the unfolded state of the villin headpiece helical subdomain

Abstract

pH-dependent stability analysis of the K48M mutant of the villin headpiece subdomain indicates that the mutation disrupts a favorable interaction in the unfolded state of HP36, but little is known about the partner for this group. Simulations conducted by another lab led to the proposal that a salt bridge between D44 and K48 is formed in the unfolded and transition state, but not in the native state (*222*). In this chapter, we first estimated the unfolded state pKa's for each acidic residue based on the native state pKa values and stability measurements of each Asp-to-Asn and Glu-to-Gln mutant. The calculation suggests that D44 is the only residue with an unfolded state pKa value lower than its value in a peptide fragment, indicating that D44 is involved in favorable interactions in the unfolded state. We then measured the apparent interaction free energy between D44 and K48 by double mutant thermodynamic cycle analysis. The measurement reveals that D44 interacts favorably with K48 in the unfolded state and that this contribution to the protein stability is unfavorable by 0.75 to 1.02 kcal mol^{-1}. The two independent approaches both provide evidence that there is an interaction between D44 and K48 in the unfolded state of HP36.

Acknowledgements

I thank Dr. Yuan Bi for experimental assistance. I also thank Mr. Vadim Patsalo for helpful discussions.

5.1 Introduction

The unfolded state is the starting state for protein folding and the reference state for the stability of a protein. Thus, unfolded state interactions may perturb the unfolded state energy thereby affecting the thermodynamic stability and the kinetic folding barrier. The unfolded state was initially believed to act as random coil, but considerable lines of evidences have shown that it can contain hydrophobic clusters as well as native or nonnative secondary structure (*223*). Electrostatic interactions are often assumed to play a minor role in the unfolded state because the charge residues are thought to be well solvated, but several recent studies have suggested electrostatic interactions can be formed in the unfolded state and can make a significant contribution to the energetics of the unfolded state (*60, 100-105, 215, 217-219*). For example, the contribution of unfolded state electrostatic interactions is on the order of 4 kcal/mol for SNase (*100*), and 3 kcal/mol for chymotrypsin inhibitor 2 (*217*), while a Lys-to-Met surface mutant has been shown to stabilize NTL9 by ~ 2 kcal/mol due to modulation of unfolded state electrostatic effects (*101*).

The chicken villin headpiece helical subdomain is an extremely popular model system for experimental and computational studies. Most of these efforts have focused on folding trajectories, but some studies of the unfolded state have been reported. The Raleigh group has demonstrated that HP21, a peptide fragment containing the first and second helices of HP36, is a good model for the unfolded state structure and shows significant native-like structure (*80, 156*). Solid-state NMR experiments done by Havlin and Tycko have been interpreted to indicate that V50 in helix-1 adopts a relatively well ordered but nonhelical conformation in the unfolded state; A57 in helix-2 is more conformationally disordered, but retains significant helix content; and L69 in helix-3 is the most disordered of the three labeled residues in unfolded HP35 (*224*). This

study involved freeze quench experiments and it is not clear how accurately they report on the solution ensemble. Several computational groups have also reported on the unfolded state structure of HP36. The Shaw group observed both native and non-native secondary structure elements in the unfolded state of HP36 in a molecular dynamics simulation (*225*). Zagrovic and Pande showed that the simulated unfolded ensemble is as compact as the native structure in terms of its average radius of gyration and solvent-accessible surface area (*226*). Wickstrom *et al.* found that helix-1 can adopt a native-like conformation in the unfolded state, but the population is low if the helix is not able to interact with other elements of structure (*227*). Freddolino and Schulten found that the unfolded state of HP35 is a compact conformation with helices 1 and 3 forming in their simulations, but the solvent-exposed surface area does not reach a native-like value (*222*). They also mentioned that the interaction between D44 and K48 was consistently formed in the flipped structure (correct secondary structure but incorrect relative orientations of the helices) and other pre-folded conformations, but not the folded state.

As described in chapter 4, the K48M mutation removes favorable electrostatic interactions in both the native and unfolded state. The native state electrostatic interaction was proven to be that between K48 and E45, but the unfolded state electrostatic interaction partner is unknown. A wide range of indirect methods has been developed to study the unfolded state under native-like conditions including studies of peptide fragments, destabilizing point mutants, changes in m-values upon mutation, amide $^{1}H/^{2}H$ exchange, and pH-dependent stability (*32, 108, 109, 219, 228*). However, these studies have provided little information about the energetics of the unfolded state interactions. In this chapter, I will show that D44 is the interaction partner of K48 in the unfolded state and obtain the interaction energetics by double mutant thermodynamic cycle analysis and by a method developed by Shen (*229*) to determine residue-specific unfolded state

pKa values. The method of Shen requires certain assumptions, (discussed below), which may not be met here, but the preliminary analysis is included to illustrate the approach.

5.2 Materials and methods

5.2.1 Protein expression and purification

The four single acidic residue mutants were expressed and purified by Dr. Bi in the Raleigh lab as described in her Ph.D thesis (*198*). The expression and purification of wildtype and K48M were described in chapter 4. The D44NK48M double mutant was expressed as a fusion protein with NTL9 as described in chapter 2. The factor Xa cleavage was carried out at 23 ^{o}C for 16-20 h. The protein was purified by ion exchange chromatography and reverse-phase HPLC using a gradient of 30-65% buffer B in 70 minutes. Buffer A was H_2O + 0.1% TFA and buffer B was 90% acetonitrile + 10% H_2O + 0.1% TFA. The elution time is 49% buffer B. The identity was confirmed by matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI). The observed molecular weight was 4193.2 and expected one was 4191.9.

5.2.2 Protein stability measurements

Protein stability was measured by CD monitored urea and thermal denaturation experiments. Urea induced unfolding was performed at 25 ^{o}C and 222 nm with samples of 15-30 μM protein in 10 mM sodium acetate and 150 mM sodium chloride at pH 6.0 on an AVIV 202SF spectrophotometer. The concentration of urea was increased from 0 to about 10 M in ~0.25 M steps. Urea concentrations were determined by measuring the refractive index. Urea unfolding curves were analyzed by a non-linear least squares fit. Thermal unfolding experiments were performed on an Applied Photophysics Chirascan CD instrument at 222 nm over the range of

2 ^{o}C to 94 ^{o}C in 2 ^{o}C intervals. The buffer and protein concentration were the same as used in the urea denaturation experiments. Measurements were made at pH 3.0 and 6.0. The reversibility of unfolding was confirmed by comparing the initial CD signal at the start of the run to the signal measured after the run was completed and the sample was cooled to the starting temperature. All thermal unfolding curves were analyzed by a non-linear least squares fit to equation 1.10.

5.2.3 Unfolded state pKa calculations

pH-dependent protein stability is related to the charge of the folded and unfolded state by the Wyman-Tanford linkage relationship:

$$\frac{\partial \Delta G^{o}}{\partial pH} = 2.303RT\Delta Q^{F\to U} = 2.303RT(Q^{U} - Q^{F}) \qquad (5.1)$$

where $\Delta Q^{F\to U}$ is the difference in the net charge between the folded (F) and unfolded (U) states, Q^{U} is the net charge in the unfolded state and Q^{F} is the net charge in the folded state. Integrating equation 5.1 and applying Henderson-Hasselbach equation (eq. 1.16) will lead to:

$$\Delta\Delta G^{\circ} = \Delta G^{\circ}(pH) - \Delta G^{\circ}(pH^{ref}) = RT\sum_{i=1}^{j}\ln\left[\frac{(1+10^{(pK_{i,F}-pH)})(1+10^{(pK_{i,U}-pH^{ref})})}{(1+10^{(pK_{i,F}-pH^{ref})})(1+10^{(pK_{i,U}-pH)})}\right] \qquad (5.2)$$

where pH^{ref} is a reference pH, $pK_{i,F}$ and $pK_{i,U}$ are the folded and unfolded state pKa's of residue *i*. Since pH^{ref} is arbitrary, it can be chosen high enough such that $1+10^{(pK_{i,U}-pH^{ref})} \approx 1$ and $1+10^{(pK_{i,F}-pH^{ref})} \approx 1$. Data collected at pH 6.0 were used. Table 5-1 lists the estimated values of $1+10^{(pK_{i,U}-pH^{ref})}$ and $1+10^{(pK_{i,F}-pH^{ref})}$, where we use the peptide fragment values as a first approximation for unfolded state pKa's, bearing mind that the purpose of the exercise is to determine the unfolded state pKa's and these may differ from the peptide values. In this case, the electrostatic contribution from residue *i* is given by:

$$\Delta\Delta G_i^o = RT \ln \frac{1+10^{(pK_{i,F}-pH)}}{1+10^{(pK_{i,U}-pH)}} \quad (5.3)$$

The unfolded state pKa can be calculated if $\Delta\Delta G_i^o$ and folded state pKa is known. $\Delta\Delta G_i^o$ is difficult to measure directly, but the substitution of an ionizable residue by a charge-neutral one without altering structure gives the closest result. Assuming that the mutation does not affect pKa's of other residues, equation 5.2 shows that the electrostatic contribution can be given as the difference between the pH-dependent stability of wildtype (WT) and mutant (MT).

$$\Delta\Delta G_i^o = \Delta\Delta G_{WT}^o - \Delta\Delta G_{MT}^o \quad (5.4)$$

A thermodynamic cycle can be constructed by performing stability measurements for WT and MT at two pH values. This is the approach developed by Shen to estimate unfolded state pKa's (*229*).

5.3 Results

5.3.1 D44 has a depressed unfolded state pKa

Stability measurements were conducted at pH 3.0 and pH 6.0 respectively. The high pH was chosen such that the acidic residues are deprotonated while the low pH was chosen such that the residues are at least partially protonated. Table 5-1 shows that pH 6.0 is a reasonable reference pH. At pH 3.0, thermal unfolding experiments were conducted to measure stability because achieving low pH value in urea solutions requires the addition of large amounts of acid due to the protonation of the urea. The free energy was calculated from T_m and $\Delta H^o(T_m)$ using Gibbs-Helmholtz equation with two ΔC_p^o values, 0.38 and 0.43 kcal mol^{-1} K^{-1} (*198*). The curves are shown as plots of CD signal recorded at 222 nm as a function of temperature in Figure 5-1. At pH 6.0, urea unfolding experiments were performed to measure the stability. The urea induced

unfolding curves are shown as plots as CD signal as a function of urea concentration in Figure 5-2. The thermodynamic data collected at both pH 3.0 and pH 6.0 are summarized in Table 5-2. The pKa calculation using stabilities calculated from both ΔC_p^o values gave very similar results. To make things simple, only pKa calculations using stabilities from the ΔC_p^o value 0.38 kcal mol^{-1} K^{-1} will be listed in the following.

Since native state pKa values of HP36 WT are known (Table 4-3 in chapter 4), we can now calculate the unfolded state pKa's of each acidic residue by the approch described in the Materials and Methods section. The results are summarized in Table 5-3 and 5-4. E45 has unfolded state pKa similar to the peptide fragment value. The pKa value of D44 in the unfolded state is 0.4 pK units lower than that of the peptide fragment, while the unfolded state pKa values of D46 and E72 are 0.2-0.3 pK units higher than the peptide model values. We can check the self-consistency of this approach by recalculating the values of $(1+10^{(pK_{i,U}-pH^{ref})})$ in order to see if the assumption that the quantity in brackets is approximately equal to 1.0 is valid. The values are also tabulated in Table 5-4. In chapter 4 we concluded that there were some favorable and unfavorable interactions in the unfolded state of HP36 based upon the pH-dependent stability analysis. The unfolded state pKa calculations here suggest that D44 is involved in favorable interaction in the unfolded state, and that D46 and E72 form unfavorable interaction in the unfolded state.

To further validate the unfolded state pKa estimations, we re-plot the pH-dependent stability graph as shown in Figure 5-3. The calculated stability curve from native state pKa's and calculated unfolded state pKa's is in good agreement with the experimentally measured stability data. This indicates that the calculated unfolded state pKa's might represent the real unfolded state values. Note that the pKa value from peptide fragment was used for the C-terminus in the

calculation, because no natural amino acid mutation can remove the C-terminal carboxyl group. A test shows that the using of pKa(C-terminus) ± 0.2 will shift the calculated stability curve up or down by about 0.2 kcal mol^{-1}, still agreeing with the measured stability data.

5.3.2 Double mutant cycle analysis indicates that there are favorable interactions between D44 and K48 in the unfolded state

Double mutant thermodynamic cycles can be used to measure the difference in the interaction free energy between the folded and unfolded state as given by the following equation (*179*):

$$\Delta G_{int}(A,B)_F - \Delta G_{int}(A,B)_U = \Delta G^o_{WT} - \Delta G^o_A - \Delta G^o_B + \Delta G^o_{AB} \quad (5.5)$$

where $\Delta G_{int}(A,B)_F$ is the folded state interaction free energy between residue A and B, $\Delta G_{int}(A,B)_U$ is the unfolded state interaction free energy between those two, ΔG^o_{WT}, ΔG^o_A, ΔG^o_B, and ΔG^o_{AB} are the unfolding free energy of wildtype, the two single mutants, and the double mutant. To construct the double mutant cycle, two single mutants, D44N (mentioned previously in this chapter) and K48M (mentioned in chapter 4), and one double mutant, D44NK48M were characterized. The urea induced unfolding transition of D44NK48M is shown in Figure 5-4, and the stability data for the wildtype and the mutants are summarized in Table 5-5. The difference in interaction free energy between the folded and unfolded state calculated from equation 5.5 is -0.75 ± 0.17 kcal mol^{-1}. The uncertainty is given by:

$$\Delta = \sqrt{(error\ of\ \Delta G^o_{WT})^2 + (error\ of\ \Delta G^o_A)^2 + (error\ of\ \Delta G^o_B)^2 + (error\ of\ \Delta G^o_{AB})^2} \quad (5.6)$$

The negative sign means that the strength of unfolded state interaction is larger than that of folded state interaction, if it exists. The PB calculations discussed in chapter 4 (Figure 4-9) indicated that D44 forms a very weak interaction with K48 in the native state. The Lys48-to-Met

mutation increases the native state pKa value of D44 by only 0.2 pKa units (Table 4-3). Since the error for the pKa value is about 0.1, it is hard to say if the 0.2 units increase is significant or just within the experimental uncertainty. If the D44-K48 salt bridge does not exist in the folded state, then the analysis indicates that the unfolded state interaction between the two groups will be energetically favorable by 0.75 kcal mol^{-1}. If the native state interaction between D44 and K48 exists, although weak, its contribution to the native state stability is 0.27 kcal mol^{-1} calculated from the native state pKa shift given by:

$$\Delta\Delta G^{o} = 2.303RT\,\Delta pKa \qquad (5.7)$$

In this case, the unfolded state interaction is favorable, with magnitude of 1.02 kcal mol^{-1}.

5.4 Discussion

The work described in this chapter provides evidence for the presence of unfolded state interactions in HP36 using two different approaches. First, through the method of Shen (*229*) to estimate residue-specific unfolded state pKa values, D44 is found to be the only one among the four acidic residues which has a depressed unfolded state pKa, about 0.4 pK units lower than the peptide model value. As discussed in chapter 4, K48 is involved in favorable interaction in the unfolded state. Thus, D44 likely forms a favorable unfolded state interaction with K48, as observed by Freddolino and Schulten in their molecular dynamic simulations of HP35 (*222*). D44 and K48 are located in helix-1 and peptide fragment studies (*80*) have shown that there is residual helical structure in the unfolded state of the protein involving this helix. Formation of helical structure in this region would potentially bring the sidechains of D44 and K48 into proximity. The D44-K48 interaction free energy difference between the folded and unfolded states is estimated to be – 0.75 kcal mol^{-1} based on the double mutant thermodynamic cycle

analysis. Taking into account that the native state interaction between D44 and K48 could range from 0 to 0.27 kcal mol^{-1}, the unfolded state interaction free energy is thus 0.75 to 1.02 kcal mol^{-1}. The stability difference between K48M mutant and wildtype at pH 6.0 is 0.87 kcal mol^{-1}. So the unfolded state interaction energetics is reasonable.

The residue-specific unfolded state pKa calculation proposed by Prof. Shen was first used on NTL9 (*229*). That work showed that the unfolded state pKa of D8 in NTL9 is 3.05, about 0.8 pK units lower than the peptide fragment value, which is consistent with the conclusion that D8 and K12 form an interaction in the unfolded state of NTL9 (*110*). HP36 is the second system this method has been applied to and our analysis suggests that unfolded state contains specific interaction. Our analysis predicts that mutation of K48 to methionine should perturb the interaction made by D44 in the unfolded state. Stability measurements were performed for D44NK48M at pH 3.0 and 6.0 (Figure 5-4) and the unfolded state pKa of D44 estimated. The calculated unfolded state pKa of D44 in the K48M background is 3.99, close to the peptide fragment value, 3.97. This confirms our prediction.

Fersht and co-workers proposed that the denatured state effects are due to a moderate depression of all of the carboxylate pKa values for some proteins (*215*). Our unfolded state pKa calculations show that they can vary from site to site. The knowledge of unfolded state pKa's allows quantitative estimation of the unfolded state contribution to the protein stability. Furthermore, it provides useful criteria to improve force fields in simulations and validate microscopic electrostatics from pH-coupled protein folding simulations (*230*). Figure 5-3 shows the plot of $\Delta\Delta G^{o}$ vs. pH calculated using native state pKa's and estimated unfolded state pKa's. The good agreement observed suggests, but not prove that the unfolded state pKa values are reasonable.

Double mutant thermodynamic cycle analysis is a powerful method to calculate the coupling free energy between two residues. However, people usually ignore unfolded state interactions when performing the double mutant cycle analysis. However, more and more studies provide evidence that there is structure in the unfolded state, thus unfolded state interactions should be taken into consideration in the double mutant cycle analysis. Several, but not many studies have applied double mutant cycle analysis to observe the unfolded state interactions (*231, 232*). For example, the ΔG^{o}_{int} between R40 and N53 in the Fyn SH3 domain was estimated to be – 0.4 kcal mol^{-1} and this interaction is believed to form in the unfolded state (*232*). Our results show that the ΔG^{o}_{int} between D44 and K48 in HP36 is – 0.75 kcal mol^{-1}.

Because the interaction between D44 and K48 is formed in the unfolded state, but likely not in the native state, monitoring the breakage of this interaction via isotope-labeled IR spectroscopy could be an alternate probe of HP36 folding kinetics.

Future Studies

We could extend this study by analyzing data at even higher pH to check for consistency and to ensure that the conditions required for the method of Shen are met.

Table 5-1. Estimated values of $1+10^{(pK_{i,F}-pH^{ref})}$ and $1+10^{(pK_{i,U}-pH^{ref})}$. The pH^{ref} here is 6.0.

residue	Folded state pKa	Peptide fragment pKa	Folded state $1+10^{(pK_{i,F}-pH^{ref})}$	Peptide $1+10^{(pK_{i,U}-pH^{ref})}$
D44	3.04	4.00	1.001	1.010
E45	3.95	4.50	1.009	1.032
D46	3.44	3.86	1.003	1.007
E72	4.37	4.15	1.023	1.014

Table 5-2. Thermodynamic parameters for the unfolding of HP36 wildtype and the mutants.

Protein	pH 6.0	pH 3.0			
	ΔG^o (urea) (kcal mol^{-1})	T_m (^oC)	$\Delta H^o(T_m)$ (kcal mol^{-1})	ΔG^o (ΔCp=0.38) (kcal mol^{-1})	ΔG^o (ΔCp=0.43) (kcal mol^{-1})
HP36 WT	3.39 ±0.06	52.8 ±0.2	25.4 ±0.5	1.70	1.64
D44N	2.93 ±0.12	51.3 ±0.1	26.5 ±0.4	1.73	1.68
E45Q	2.99 ±0.11	52.7 ±0.1	27.8 ±0.3	1.90	1.84
D46N	2.51 ±0.09	49.7 ±0.1	25.9 ±0.4	1.61	1.56
E72Q	3.28 ±0.10	50.4 ±0.1	25.4 ±0.3	1.60	1.55

The uncertainties are the standard errors to the fit.

Table 5-3. Data used to calculate unfolded state pKa's for wildtype HP36 via the analysis of Shen method (*229*). Experiments were conducted at 25 ^{o}C in 10 mM sodium acetate 150 mM sodium chloride.

mutant	$\Delta\Delta G^{o}_{MT}$ = ΔG^{o}_{MT} (pH 3) – ΔG^{o}_{MT} (pH 6) (kcal mol^{-1})	$\Delta\Delta G^{o}_{i}$ = (ΔG^{o}_{WT} (pH 3) – ΔG^{o}_{WT} (pH 6)) – (ΔG^{o}_{MT} (pH 3) – ΔG^{o}_{MT} (pH 6)) (kcal mol^{-1})
D44N	-1.20	-0.49
E45Q	-1.09	-0.60
D46N	-0.90	-0.79
E72Q	-1.68	-0.01

Table 5-4. Calculated unfolded state pKa's for wildtype HP36. Calculations made use of the data in Table 5-3 and the native state pKa's for wildtype HP36. The pH^{ref} here is 6.0.

residue	wildtype native state $pK_{i,F}$	estimated wildtype unfolded state $pK_{i,U}$	$pK_{i,\,peptide}$	$1+10^{(pK_{i,U}-pH^{ref})}$
D44	3.04	3.58	4.00	1.004 [1]
E45	3.95	4.42	4.50	1.026 [1]
D46	3.44	4.12	3.86	1.013 [1]
E72	4.37	4.38	4.15	1.024 [1]

1) calculated using estimated $pK_{i,U}$ values from column 5 of this table.

Table 5-5. Thermodynamic parameters for the unfolding of HP36 WT and the mutants. The uncertainties are the standard error to the fits. Experiments were performed at 25 ^{o}C with 10 mM sodium acetate and 150 mM sodium chloride, pH 6.0.

Protein	ΔG^{o} (kcal mol^{-1})	m-value (kcal mol^{-1} K^{-1})	C_M (M)
HP36 WT	3.39 ± 0.06	0.52 ± 0.01	6.5
D44N	2.93 ± 0.12	0.63 ± 0.02	4.7
K48M	4.26 ± 0.07	0.52 ± 0.01	8.2
D44NK48M	3.05 ± 0.07	0.49 ± 0.02	6.3

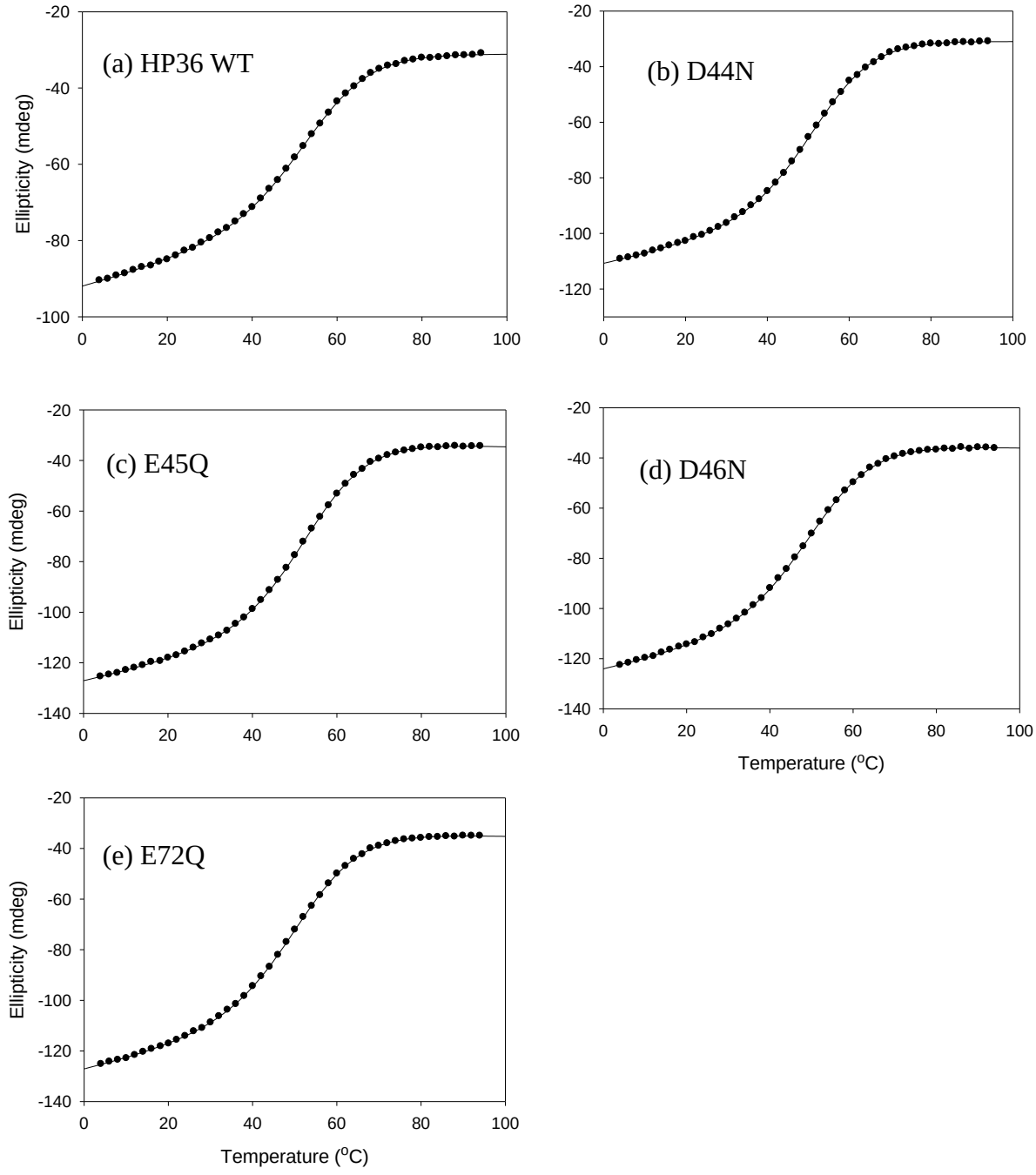


Figure 5-1. Temperature induced unfolding transitions of HP36 wildtype and the mutants. (a)HP36 WT, (b)D44N, (c)E45Q, (d)D46N, (e)E72Q. The solid line represents the best fit to a two-state folding transition. Signals were recorded at 222 nm. All spectra were collected in 10 mM sodium acetate, 150 mM sodium chloride, pH 3.0 using a 1.0 cm cuvette.

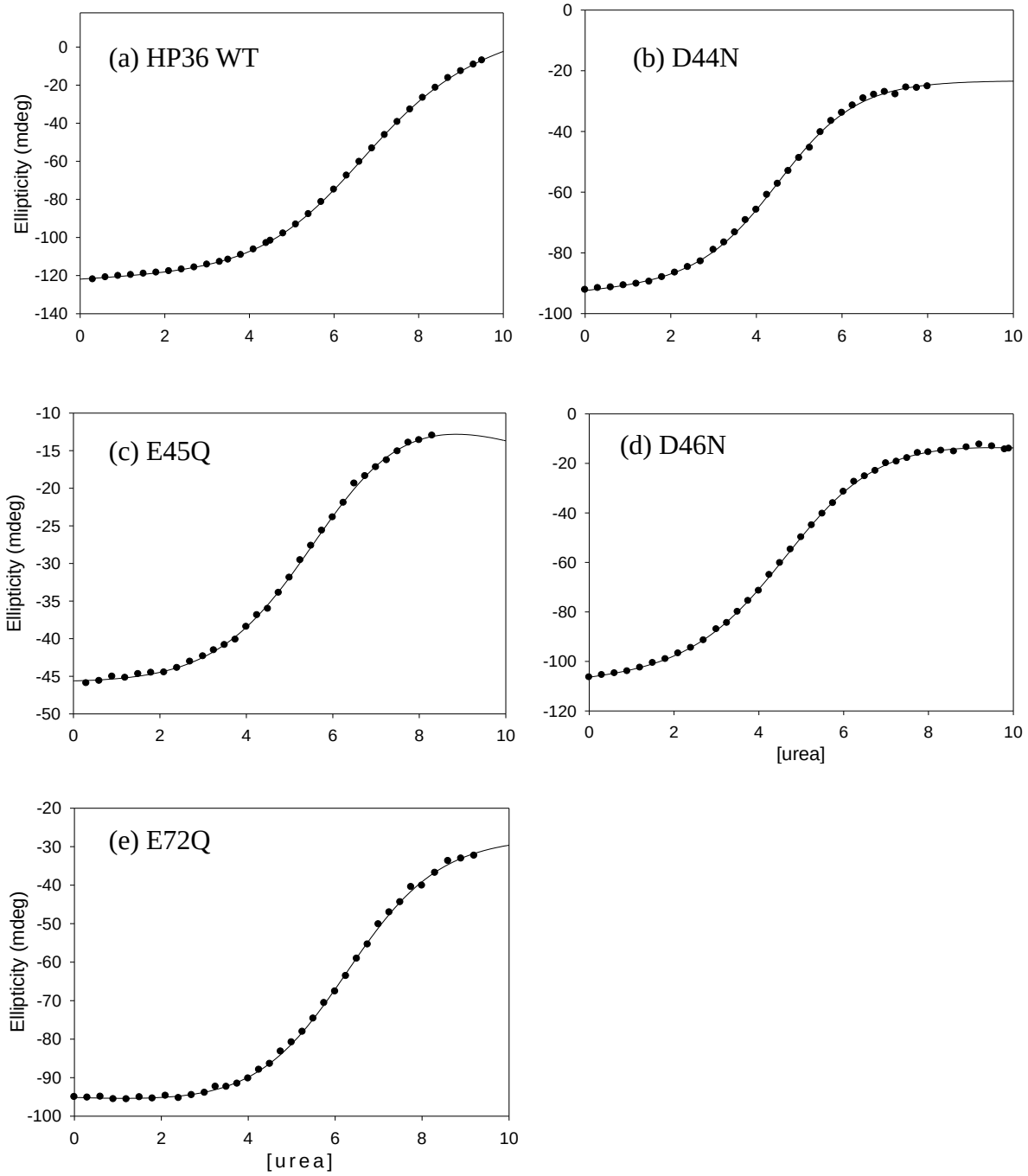


Figure 5-2. Urea induced unfolding transitions of HP36 wildtype and the mutants. (a)HP36 WT, (b)D44N, (c)E45Q, (d)D46N, (e)E72Q. The solid line represents the best fit to a two-state folding transition. Signals were recorded at 222 nm. All spectra were collected at 25 ^{o}C in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0 using a 1.0 cm cuvette.

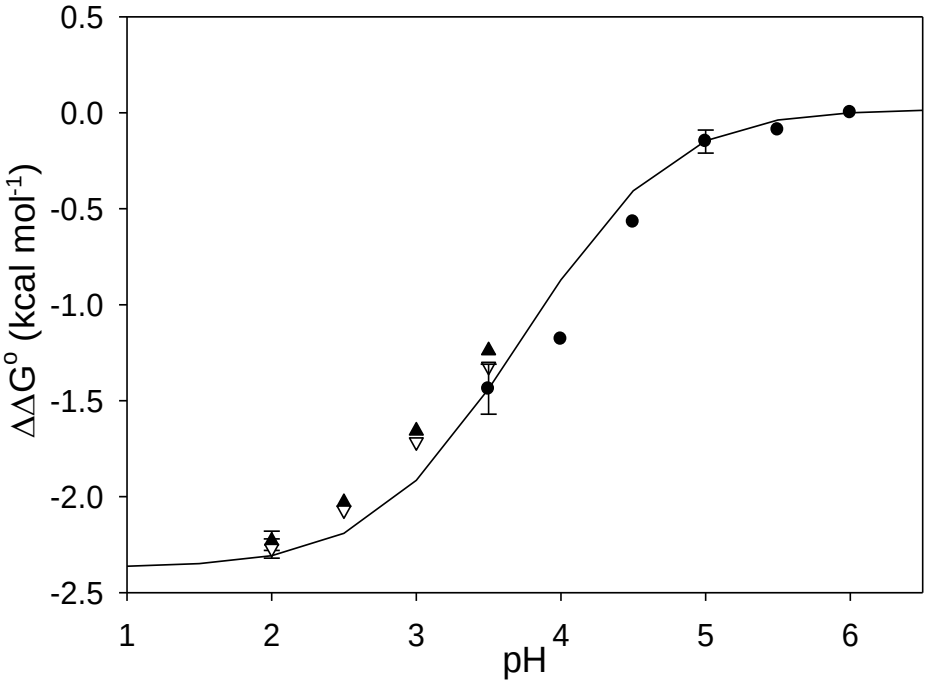


Figure 5-3. pH-dependent stability analysis of HP36 wildtype. The solid line represents the calculated stability change using known native state pKa measurements and calculated unfolded state pKa's. The filled circles represent the measured stability change from urea induced unfolding experiments. The filled triangles represent the calculated stability change from thermal unfolding experiments using the Gibbs-Helmholtz equation. A ΔC_p^o of 0.38 kcal mol^{-1} K^{-1} was used. The open triangles represent the calculated stability change from thermal unfolding experiments using Gibbs-Helmholtz equation. A ΔC_p^o of 0.43 kcal mol^{-1} K^{-1} was used. The error bars represent the standard deviations of three repeated experiments. Stabilities are values at 25 ^oC. Samples contained 10 mM sodium acetate and 150 mM sodium chloride.

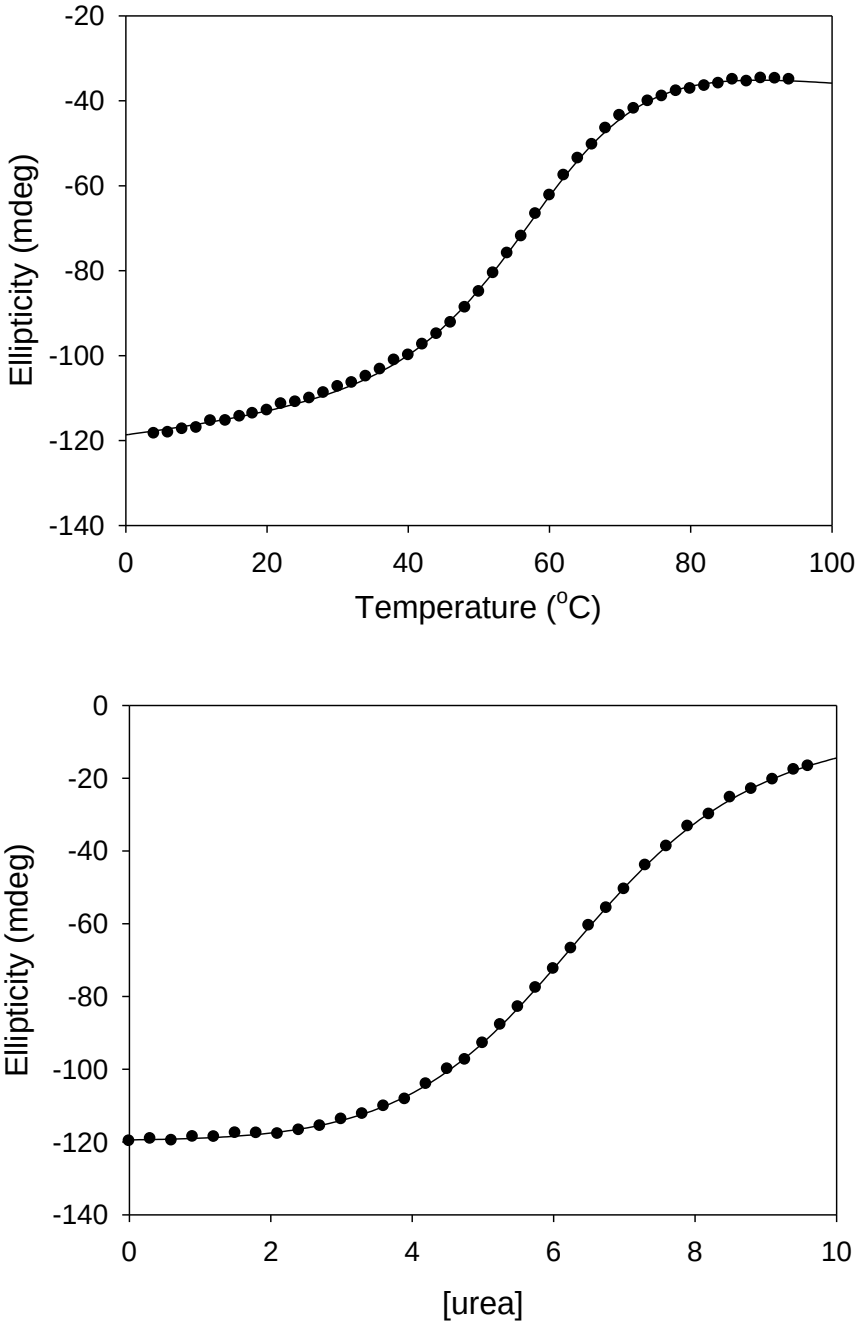


Figure 5-4. CD monitored thermal unfolding and urea unfolding transition for the D44NK48M mutant. The thermal unfolding experiment was performed at 222 nm, in 10 mM sodium acetate 150 mM sodium chloride, pH 3.0. The urea unfolding experiment was performed at 25 ^{o}C and 222 nm, in 10 mM sodium acetate 150 mM sodium chloride, pH 6.0. A 1.0 cm cuvette was used. The solid line represents the best fit to a two-state folding transition.

References

1. Anfinsen, C. B. (1972) The formation and stabilization of protein structure, *Biochem. J. 128*, 737-749.
2. Selkoe, D. J. (2003) Folding proteins in fatal ways, *Nature 426*, 900-904.
3. Hart, T., Hosszu, L. L., Trevitt, C. R., Jackson, G. S., Waltho, J. P., Collinge, J., and Clarke, A. R. (2009) Folding kinetics of the human prion protein probed by temperature jump, *Proc. Natl. Acad. Sci. USA. 106*, 5651-5656.
4. Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A., and Hofrichter, J. (2006) Sub-microsecond protein folding, *J. Mol. Biol. 359*, 546-553.
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. USA. 102*, 16662-16667.
6. Levinthal, C. (1968) Are there pathways for protein folding?, *J. Chim. Phys. 65*, 44-45.
7. Chan, H. S., Bromberg, S., and Dill, K. A. (1995) Models of cooperativity in protein folding, *Philos. trans. R. Soc. Lond. B Biol. Sci. 348*, 61-70.
8. Matthews, C. R., and Crisanti, M. M. (1981) Urea-induced unfolding of the alpha subunit of tryptophan synthase: evidence for a multistate process, *Biochemistry 20*, 784-792.
9. Bai, Y., and Englander, S. W. (1996) Future directions in folding: the multi-state nature of protein structure, *Proteins: Struct., Funct., Genet. 24*, 145-151.
10. Brandts, J. F. (1964) The thermodynamics of protein denaturation. I. The denaturation of chymotrypsinogen, *J. Am. Chem. Soc. 86*, 4291-4301.
11. Tanford, C. (1968) Protein denaturation, *Adv. Protein Chem. 23*, 121-282.
12. Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of

denaturation, *Adv. Protein Chem. 24*, 1-95.

13. Ikai, A., and Tanford, C. (1971) Kinetic evidence for incorrectly folded intermediate states in refolding of denatured proteins, *Nature 230*, 100-102.

14. Tsong, T. Y., Baldwin, R. L., and Elson, E. L. (1971) The sequential unfolding of Ribonuclease A: Detection of a fast initial phase in kinetics of unfolding, *Proc. Natl. Acad. Sci. USA. 68*, 2712-2715.

15. Vallee-Belisle, A., and Michnick, S. W. (2007) Multiple tryptophan probes reveal that ubiquitin folds via a late misfolded intermediate, *J. Mol. Biol. 374*, 791-805.

16. Rea, A. M., Simpson, E. R., Crespo, M. D., and Searle, M. S. (2008) Helix mutations stabilize a late productive intermediate on the folding pathway of ubiquitin, *Biochemistry 47*, 8225-8236.

17. Whittaker, S. B. M., Spence, G. R., Grossmann, J. G., Radford, S. E., and Moore, G. R. (2007) NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7, *J. Mol. Biol. 366*, 1001-1015.

18. Connell, K. B., Horner, G. A., and Marqusee, S. (2009) A single mutation at residue 25 populates the folding intermediate of E. coli RNase H and reveals a highly dynamic partially folded ensemble, *J. Mol. Biol. 391*, 461-470.

19. Gsponer, J., Hopearuoho, H., Whittaker, S. B. M., Spence, G. R., Moore, G. R., Paci, E., Radford, S. E., and Vendruscolo, M. (2006) Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7, *Proc. Natl. Acad. Sci. USA. 103*, 99-104.

20. Dixon, R. D. S., Chen, Y. W., Feng, D., Khare, S. D., Prutzman, K. C., Schaller, M. D., Campbell, S. L., and Dokholyan, N. V. (2004) New insights into FAK signaling and

localization based on detection of a FAT domain folding intermediate, *Structure 12*, 2161-2171.

21. Karplus, M., and Weaver, D. L. (1976) Protein folding dynamics, *Nature 260*, 404-406.

22. Karplus, M., and Weaver, D. L. (1994) Protein folding dynamics: The diffusion-collision model and experimental data, *Protein Sci. 3*, 650-668.

23. Kim, P. S., and Baldwin, R. L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding, *Annu. Rev. Biochem. 51*, 459-489.

24. Baldwin, R. L. (1989) How does protein folding get started?, *Trends Biochem. Sci. 14*, 291-294.

25. Tanford, C. (1962) Contribution of hydrophobic interactions of stability of globular conformation of proteins, *J. Am. Chem. Soc. 84*, 4240-4247.

26. Fersht, A. R. (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications, *Proc. Natl. Acad. Sci. USA. 92*, 10869-10873.

27. Fersht, A. R. (1997) Nucleation mechanisms in protein folding, *Curr. Opin. Struct. Biol. 7*, 3-9.

28. Larson, S. M., Ruczinski, I., Davidson, A. R., Baker, D., and Plaxco, K. W. (2002) Residues participating in the protein folding nucleus do not exhibit preferential evolutionary conservation, *J. Mol. Biol. 316*, 225-233.

29. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscaple of protein folding: A synthesis, *Proteins: Struct., Funct., Genet. 21*, 167-195.

30. Dill, K. A., Bromberg, S., Yue, K. Z., 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.

31. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels, *Nat. Struct. Biol. 4*, 10-19.

32. 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.

33. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves, *Methods Enzymol. 131*, 266-280.

34. Gruenewald, B., Nicola, C. U., Lustig, A., Schwarz, G., and Klump, H. (1979) Kinetics of the helix-coil transition of a polypeptide with nonionic side groups, derived from ultrasonic relaxation measurements, *Biophys. Chem. 9*, 137-147.

35. Jones, C. M., Henry, E. R., Hu, Y., Chan, C. K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, J., and Eaton, W. A. (1993) Fast events in protein folding initiated by nanosecond laser photolysis, *Proc. Natl. Acad. Sci. USA. 90*, 11860-11864.

36. Lu, H. S. M., Volk, M., Kholodenko, Y., Gooding, E., Hochstrasser, R. M., and DeGrado, W. F. (1997) Aminothiotyrosine disulfide, an optical trigger for initiation of protein folding, *J. Am. Chem. Soc. 119*, 7173-7180.

37. Volk, M., Kholodenko, Y., Lu, H. S. M., Gooding, E. A., DeGrado, W. F., and Hochstrasser, R. M. (1997) Peptide conformational dynamics and vibrational stark effects following photoinitiated disulfide cleavage, *J. Phys. Chem. B 101*, 8607-8616.

38. Phillips, C. M., Mizutani, Y., and Hochstrasser, R. M. (1995) Ultrafast thermally induced unfolding of RNase A, *Proc. Natl. Acad. Sci. USA. 92*, 7292-7296.

39. Nolting, B. (1996) Temperature-jump induced fast refolding of cold-unfolded protein, *Biochem. Biophys. Res. Commun. 227*, 903-908.

40. Nolting, B., Golbik, R., and Fersht, A. R. (1995) Submillisecond events in protein folding, *Proc. Natl. Acad. Sci. USA. 92*, 10668-10672.

41. Jacob, M., Holtermann, G., Perl, D., Reinstein, J., Schindler, T., Geeves, M. A., and Schmid, F. X. (1999) Microsecond folding of the cold shock protein measured by a pressure-jump technique, *Biochemistry 38*, 2882-2891.

42. Chan, C. K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A., and Hofrichter, J. (1997) Submillisecond protein folding kinetics studied by ultrarapid mixing, *Proc. Natl. Acad. Sci. USA. 94*, 1779-1784.

43. Takahashi, S., Yeh, S. R., Das, T. K., Chan, C. K., Gottfried, D. S., and Rousseau, D. L. (1997) Folding of cytochrome c initiated by submillisecond mixing, *Nat. Struct. Biol. 4*, 44-50.

44. Park, S. H., Shastry, M. C. R., and Roder, H. (1999) Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing, *Nat. Struct. Biol. 6*, 943-947.

45. Knight, J. B., Vishwanath, A., Brody, J. P., and Austin, R. H. (1998) Hydrodynamic focusing on a silicon chip: Mixing nanoliters in microseconds, *Phys. Rev. Lett. 80*, 3863-3866.

46. Pollack, L., Tate, M. W., Darnton, N. C., Knight, J. B., Gruner, S. M., Eaton, W. A., and Austin, R. H. (1999) Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering, *Proc. Natl. Acad. Sci. USA. 96*, 10115-10117.

47. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997)

The energy landscape of a fast-folding protein mapped by Ala->Gly substitutions, *Nat. Struct. Biol. 4*, 305-310.

48. Huang, G. S., and Oas, T. G. (1995) Submillisecond folding of monomeric lambda-repressor, *Proc. Natl. Acad. Sci. USA. 92*, 6878-6882.

49. Vugmeyster, L., Raleigh, D. P., Palmer, A. G., and Vugmeister, B. E. (2003) Beyond the decoupling approximation in the model free approach for the interpretation of NMR relaxation of macromolecules in solution, *J. Am. Chem. Soc. 125*, 8400-8404.

50. Pauling, L., Corey, R. B., and Branson, H. R. (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain, *Proc. Natl. Acad. Sci. USA. 37*, 205-211.

51. Fersht, A. R., Shi, J. P., Knilljones, J., Lowe, D. M., Wilkinson, A. J., Blow, D. M., Brick, P., Carter, P., Waye, M. M. Y., and Winter, G. (1985) Hydrogen bonding and biological specificity analyzed by protein engineering, *Nature 314*, 235-238.

52. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins, *FASEB J. 10*, 75-83.

53. Dill, K. A. (1990) Dominant forces in protein folding, *Biochemistry 29*, 7133-7155.

54. Honig, B., and Yang, A. S. (1995) Free energy balance in protein folding, *Adv. Protein Chem. 46*, 27-58.

55. Privalov, P. L., and Makhatadze, G. I. (1993) Contribution of hydration to protein folding thermodynamics II. The entropy and Gibbs energy of hydration, *J. Mol. Biol. 232*, 660-679.

56. Weber, G. (1996) Persistent confusion of total entropy and chemical system entropy in chemical thermodynamics, *Proc. Natl. Acad. Sci. USA. 93*, 7452-7453.

57. Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) pH-induced denaturatoin of proteins: A single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme, *Biochemistry 29*, 2403-2408.

58. Hendsch, Z. S., and Tidor, B. (1999) Electrostatic interactions in the GCN4 leucine zipper: Substantial contributions arise from intramolecular interactions enhanced on binding, *Protein Sci. 8*, 1381-1392.

59. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. F., and Thomas, S. T. (2003) Contribution of surface salt bridges to protein stability: Guidelines for protein engineering, *J. Mol. Biol. 327*, 1135-1148.

60. 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.

61. 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.

62. Kumar, S., and Nussinov, R. (1999) Salt bridge stability in monomeric proteins, *J. Mol. Biol. 293*, 1241-1255.

63. Hendsch, Z. S., and Tidor, B. (1994) Do salt bridges stabilize proteins? A continuum electrostatic analysis, *Protein Sci. 3*, 211-226.

64. Wong, K. B., Freund, S. M. V., and Fersht, A. R. (1996) Cold denaturation of barstar: H-1, N-15 and C-13 NMR assignment and characterisation of residual structure, *J. Mol. Biol. 259*, 805-818.

65. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol. 6*, 197-208.

66. Bai, Y. W., Milne, J. S., Mayne, L., and Englander, S. W. (1993) Primary structure effects on peptide group hydrogen exchange, *Proteins: Struct., Funct., Genet. 17*, 75-86.

67. 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.

68. Mishima, T., Ohkuri, T., Monji, A., Imoto, T., and Ueda, T. (2006) Amyloid formation in denatured single-mutant lysozymes where residual structures are modulated, *Protein Sci. 15*, 2448-2452.

69. Flory, P. J. (1965) Principles of polymer chemistry, *Cornell University Press, Ithaca, NY*.

70. Tanford, C., Kawahara, K., Lapanje, S., Hooker, T. M., Zarlengo, M. H., Salahudd.A, Aune, K. C., and Takagi, T. (1967) Proteins as random coils. 3. Optical rotatory dispersion in 6M guanidine hydrochloride, *J. Am. Chem. Soc. 89*, 5023-5029.

71. Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P., Sosnick, T. R., Hasan, M. Z., Pande, V. S., Ruczinski, I., Doniach, S., and Plaxco, K. W. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins, *Proc. Natl. Acad. Sci. USA. 101*, 12491-12496.

72. Goldenberg, D. P. (2003) Computational simulation of the statistical properties of unfolded proteins, *J. Mol. Biol. 326*, 1615-1633.

73. Yi, Q., Scalley-Kim, M. L., Alm, E. J., and Baker, D. (2000) NMR characterization of residual structure in the denatured state of protein L, *J. Mol. Biol. 299*, 1341-1351.

74. Bu, Z. M., Cook, J., and Callaway, D. J. E. (2001) Dynamic regimes and correlated

structural dynamics in native and denatured alpha-lactalbumin, *J. Mol. Biol. 312*, 865-873.

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

76. Plattt, G. W., McParland, V. J., Kalverda, A. P., Homans, S. W., and Radford, S. E. (2005) Dynamics in the unfolded state of beta(2)-microglobulin studied by NMR, *J. Mol. Biol. 346*, 279-294.

77. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M. V., and Fersht, A. R. (2005) Solution structure of a protein denatured state and folding intermediate, *Nature 437*, 1053-1056.

78. Marsh, J. A., Neale, C., Jack, F. E., Choy, W.-Y., Lee, A. Y., Crowhurst, K. A., and Forman-Kay, J. D. (2007) Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure, *J. Mol. Biol. 367*, 1494-1510.

79. Cliff, M. J., Craven, C. J., Marston, J. P., Hounslow, A. M., Clarke, A. R., and Waltho, J. P. (2009) The denatured state of N-PGK is compact and predominantly disordered, *J. Mol. Biol. 385*, 266-277.

80. Meng, W., Shan, B., Tang, Y., and Raleigh, D. P. (2009) Native like structure in the unfolded state of the villin headpiece helical subdomain, an ultrafast folding protein, *Protein Sci. 18*, 1692-1701.

81. Shan, B., Eliezer, D., and Raleigh, D. P. (2009) The unfolded state of the C-terminal domain of the ribosomal protein L9 contains both native and non-native structure, *Biochemistry 48*, 4707-4719.

82. Bruun, S. W., Iesmantavicius, V., Danielsson, J., and Poulsen, F. M. (2010) Cooperative formation of native-like tertiary contacts in the ensemble of unfolded states of a four-helix protein, *Proc. Natl. Acad. Sci. USA. 107*, 13306-13311.

83. Chow, C. Y., Wu, M.-C., Fang, H.-J., Hu, C.-K., Chen, H.-M., and Tsong, T.-Y. (2008) Compact dimension of denatured states of staphylococcal nuclease, *Proteins: Struct., Funct., Bioinf. 72*, 901-909.

84. Fitzkee, N. C., and Bertrand Garcia-Moreno, E. (2008) Electrostatic effects in unfolded staphylococcal nuclease, *Protein Sci. 17*, 216-227.

85. Patel, S., and Sasidhar, Y. U. (2008) A shorter peptide model from staphylococcal nuclease for the folding-unfolding equilibrium of a beta-hairpin shows that unfolded state has significant contribution from compact conformational states, *J. Struct. Biol. 164*, 60-74.

86. Feng, Y. G., Liu, D. S., and Wang, J. F. (2003) Native-like partially folded conformations and folding process revealed in the N-terminal large fragments of staphylococcal nuclease: A study by NMR spectroscopy, *J. Mol. Biol. 330*, 821-837.

87. Pappu, R. V., Srinivasan, R., and Rose, G. D. (2000) The Flory isolated-pair hypothesis is not valid for polypeptide chains: Implications for protein folding, *Proc. Natl. Acad. Sci. USA. 97*, 12565-12570.

88. Srinivasan, R., and Rose, G. D. (2002) Methinks it is like a folding curve, *Biophys. Chem. 101*, 167-171.

89. Fitzkee, N. C., and Rose, G. D. (2004) Steric restrictions in protein folding: An alpha-helix cannot be followed by a contiguous beta-strand, *Protein Sci. 13*, 633-639.

90. Ptitsyn, O. B. (1987) Protein folding: hypotheses and experiments, *J. Protein Chem. 6*,

273-293.

91. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins, *Annu. Rev. Biochem. 60*, 795-825.

92. Bromberg, S., and Dill, K. A. (1994) Side-chain entropy and packing in proteins, *Protein Sci. 3*, 997-1009.

93. Flanagan, J. M., Kataoka, M., Shortle, D., and Engelman, D. M. (1992) Truncated staphylococcal nuclease is compact but disordered, *Proc. Natl. Acad. Sci. USA. 89*, 748-752.

94. Sosnick, T. R., and Trewhella, J. (1992) Denatured states of ribonuclease A have compact dimensions and residual secondary structure, *Biochemistry 31*, 8329-8335.

95. Choy, W. Y., Mulder, F. A. A., Crowhurst, K. A., Muhandiram, D. R., Millett, I. S., Doniach, S., Forman-Kay, J. D., and Kay, L. E. (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques, *J. Mol. Biol. 316*, 101-112.

96. 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.

97. Baldwin, R. L. (2002) A new perspective on unfolded proteins, *Adv. Protein Chem. 62*, 361-367.

98. Lumb, K. J., and Kim, P. S. (1994) Formation of a hydrophobic cluster in denatured bovine pancreatic trypsin inhibitor, *J. Mol. Biol. 236*, 412-420.

99. Evans, P. A., Topping, K. D., Woolfson, D. N., and Dobson, C. M. (1991) Hydrophobic clustering in nonnative states of a protein: Interpretation of chemical shifts in NMR

spectra of denatured states of lysozyme, *Proteins: Struct., Funct., Genet. 9*, 248-266.

100. Whitten, S. T., and Garcia-Moreno, B. (2000) pH dependence of stability of staphylococcal nuclease: Evidence of substantial electrostatic interactions in the denatured state, *Biochemistry 39*, 14292-14304.

101. 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.

102. Bowler, B. E. (2007) Thermodynamics of protein denatured states, *Mol. BioSyst. 3*, 88-99.

103. Guzman-Casado, M., Parody-Morreale, A., Robic, S., Marqusee, S., and Sanchez-Ruiz, J. M. (2003) Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of Thermus thermophilus ribonuclease H, *J. Mol. Biol. 329*, 731-743.

104. Lindman, S., Bauer, M. C., Lund, M., Diehl, C., Mulder, F. A. A., Akke, M., and Linse, S. (2010) pKa values for the unfolded state under native conditions explain the pH-dependent stability of PGB1, *Biophys. J. 99*, 3365-3373.

105. Arbely, E., Rutherford, T. J., Neuweiler, H., Sharpe, T. D., Ferguson, N., and Fersht, A. R. (2010) Carboxyl pKa values and acid denaturation of BBL, *J. Mol. Biol. 403*, 313-327.

106. Kundrotas, P. J., and Karshikoff, A. (2002) Modeling of denatured state for calculation of the electrostatic contribution to protein stability, *Protein Sci. 11*, 1681-1686.

107. Tollinger, M., Forman-Kay, J. D., and Kay, L. E. (2002) Measurement of side-chain carboxyl pKa values of glutamate and aspartate residues in an unfolded protein by multinuclear NMR spectroscopy, *J. Am. Chem. Soc. 124*, 5714-5717.

108. Meng, W., and Raleigh, D. P. (2011) Analysis of electrostatic interactions in the denatured state ensemble of the N-terminal domain of L9 under native conditions, *Proteins: Struct., Funct., Bioinf. 79*, 3500-3510.

109. 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.

110. Cho, J.-H., Sato, S., Horng, J.-C., Anil, B., and Raleigh, D. P. (2008) Electrostatic interactions in the denatured state ensemble: Their effect upon protein folding and protein stability, *Arch. Biochem. Biophys. 469*, 20-28.

111. Finidori, J., Friederich, E., Kwiatkowski, D. J., and Louvard, D. (1992) In vivo analysis of functional domains from villin and gelsolin, *J. Cell Biol. 116*, 1145-1155.

112. Pope, B., Way, M., Matsudaira, P. T., and Weeds, A. (1994) Characterization of the F-actin binding domains of villin: Classification of F-actin binding proteins into 2 groups according to their binding sites on actin, *FEBS. Lett. 338*, 58-62.

113. 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.

114. 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.

115. Wang, M. H., Tang, Y. F., Sato, S. 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.

116. Brewer, S. H., Vu, D. M., Tang, Y. F., 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. USA. 102*, 16662-16667.

117. Islam, S. A., Karplus, M., and Weaver, D. L. (2002) Application of the diffusion-collision model to the folding of three-helix bundle proteins, *J. Mol. Biol. 318*, 199-215.

118. 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.

119. Shen, M. Y., and Freed, K. F. (2002) All-atom fast protein folding simulations: The villin headpiece, *Proteins: Struct., Funct., Genet. 49*, 439-445.

120. Srinivas, G., and Bagchi, B. (2002) Folding and unfolding of chicken villin headpiece: Energy landscape of a single-domain model protein, *Curr. Sci. 82*, 179-185.

121. 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.

122. van der Spoel, D., and Lindahl, E. (2003) Brute-force molecular dynamics simulations of Villin headpiece: Comparison with NMR parameters, *J. Phys. Chem. B 107*, 11178-11187.

123. Ensign, D. L., Kasson, P. M., and Pande, V. S. (2007) Heterogeneity even at the speed limit of folding: Large-scale molecular dynamics study of a fast-folding variant of the villin headpiece, *J. Mol. Biol. 374*, 806-816.

124. Jayachandran, G., Vishal, V., Garcia, A. E., and Pande, V. S. (2007) Local structure formation in simulations of two small proteins, *J. Struct. Biol. 157*, 491-499.

125. Xiao, S., Bi, Y., Shan, B., and Ralelgh, D. P. (2009) Analysis of core packing in a cooperatively folded miniature protein: The ultrafast folding villin headpiece helical subdomain, *Biochemistry 48*, 4607-4616.

126. Glasscock, J. M., Zhu, Y., Chowdhury, P., Tang, J., and Gai, F. (2008) Using an amino acid fluorescence resonance energy transfer pair to probe protein unfolding: Application to the villin headpiece subdomain and the LysM domain, *Biochemistry 47*, 11070-11076.

127. 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.

128. Urbanek, D. C., Vorobyev, D. Y., Serrano, A. L., Gai, F., and Hochstrasser, R. M. (2010) The two-dimensional vibrational echo of a nitrile probe of the villin HP35 protein, *J. Phys. Chem. Lett. 1*, 3311-3315.

129. Packer, L. E., Song, B., Raleigh, D. P., and McKnight, C. J. (2011) Competition between intradomain and interdomain interactions: A buried salt bridge is essential for villin headpiece folding and actin binding, *Biochemistry 50*, 3706-3712.

130. Chiu, T. K., Kubelka, J., Herbst-Irmer, R., Eaton, W. A., Hofrichter, J., and Davies, D. R. (2005) High resolution X-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein, *Proc. Natl. Acad. Sci. USA. 102*, 7517-7522.

131. Dobson, C. M., Sali, A., and Karplus, M. (1998) Protein folding: A perspective from theory and experiment, *Angew. Chem.-Int. Edit. 37*, 868-893.

132. Pace, C. N., Grimsley, G. R., and Scholtz, J. M. (2009) Protein ionizable groups: pK values and their contribution to protein stability and solubility, *J. Biol. Chem. 284*, 13285-13289.

133. 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.

134. Kubelka, J., Eaton, W. A., and Hofrichter, J. (2003) Experimental tests of villin

subdomain folding simulations, *J. Mol. Biol. 329*, 625-630.

135. 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. USA 95*, 9897-9902.

136. Sullivan, D. C., and Kuntz, I. D. (2002) Protein folding as biased conformational diffusion, *J. Phys. Chem. B 106*, 3255-3262.

137. 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.

138. 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.

139. Lin, C. Y., Hu, C. K., and Hansmann, U. H. E. (2003) Parallel tempering simulations of HP-36, *Proteins: Struct., Funct., Genet. 52*, 436-445.

140. He, J. B., Zhang, Z. Y., Shi, Y. Y., and Liu, H. Y. (2003) Efficiently explore the energy landscape of proteins in molecular dynamics simulations by amplifying collective motions, *J. Chem. Phys. 119*, 4005-4017.

141. Jang, S. M., 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.

142. Ripoll, D. R., Vila, J. A., and Scheraga, H. A. (2004) Folding of the villin headpiece subdomain from random structures. analysis of the charge distribution as a function of pH, *J. Mol. Biol. 339*, 915-925.

143. Kim, S. Y., and Lee, J. (2005) Folding simulations of small proteins, *Biophys. Chem. 115*,

195-200.

144. De Mori, G. M. S., 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: Struct., Funct., Bioinf. 58*, 459-471.

145. De Mori, G. M. S., Micheletti, C., and Colombo, G. (2004) All-atom folding simulations of the villin headpiece from stochastically selected coarse-grained structures, *J. Phys. Chem. B 108*, 12267-12270.

146. Godoy-Ruiz, R., Henry, E. R., Kubelka, J., Hofrichter, J., Munoz, V., Sanchez-Ruiz, J. M., and Eaton, W. A. (2008) Estimating free-energy barrier heights for an ultrafast folding protein from calorimetric and kinetic data, *J. Phys. Chem. B 112*, 5938-5949.

147. Bi, Y., Cho, J. H., Kim, E. Y., Shan, B., Schindelin, H., and Raleigh, D. P. (2007) Rational design, structural and thermodynamic characterization of a hyperstable variant of the villin headpiece helical subdomain, *Biochemistry 46*, 7497-7505.

148. Hansmann, U. H. E. (2002) Protein-folding simulations in generalized ensembles, *Int. J. Quantum Chem. 90*, 1515-1521.

149. Lei, H. X., and Duan, Y. (2007) Two-stage folding of HP-35 from ab initio simulations, *J. Mol. Biol. 370*, 196-206.

150. Micheletti, C., and Colombo, G. (2008) Study of the villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics, *Proteins: Struct., Funct., Bioinf. 70*, 309-309.

151. Mukherjee, A., and Bagchi, B. (2003) Correlation between rate of folding, energy landscape, and topology in the folding of a model protein HP-36, *J. Chem. Phys. 118*, 4733-4747.

152. Piana, S., Laio, A., Marinelli, F., Van Troys, M., Bourry, D., Ampe, C., and Martins, J. C. (2008) Predicting the effect of a point mutation on a protein fold: The villin and advillin headpieces and their Pro62Ala mutants, *J. Mol. Biol. 375*, 460-470.

153. Wei, Y., Nadler, W., and Hansmann, U. H. E. (2008) Backbone and side-chain ordering in a small protein, *J. Chem. Phys. 128*.

154. Yang, J. S., Wallin, S., and Shakhnovich, E. I. (2008) Universality and diversity of folding mechanics for three-helix bundle proteins, *Proc. Natl. Acad. Sci. USA. 105*, 895-900.

155. Tang, Y. F., Rigotti, D. J., Fairman, R., and Raleigh, D. P. (2004) Peptide models provide evidence for significant structure in the denatured state of a rapidly folding protein: The villin headpiece subdomain, *Biochemistry 43*, 3264-3272.

156. Tang, Y. F., Goger, M. J., and Raleigh, D. P. (2006) NMR characterization of a peptide model provides evidence for significant structure in the unfolded state of the villin headpiece helical subdomain, *Biochemistry 45*, 6940-6946.

157. Vermeulen, W., Van Troys, M., Bourry, D., Dewitte, D., Rossenu, S., Goethals, M., Borremans, F. A. M., Vandekerckhove, J., Martins, J. C., and Ampe, C. (2006) Identification of the PXW sequence as a structural gatekeeper of the headpiece C-terminal subdomain fold, *J. Mol. Biol. 359*, 1277-1292.

158. Cellmer, T., Henry, E. R., Hofrichter, J., and Eaton, W. A. (2008) Measuring internal friction of an ultrafast-folding protein, *Proc. Natl. Acad. Sci. USA. 105*, 18320-18325.

159. Beauchamp, K. A., Ensign, D. L., Das, R., and Pande, V. S. (2011) Quantitative comparison of villin headpiece subdomain simulations and triplet-triplet energy transfer experiments, *Proc. Natl. Acad. Sci. USA. 108*, 12734-12739.

160. Frank, B. S., Vardar, D., Buckley, D. A., and McKnight, C. J. (2002) The role of aromatic residues in the hydrophobic core of the villin headpiece subdomain, *Protein Sci. 11*, 680-687.

161. Gellman, S. H., and Woolfson, D. N. (2002) Mini-proteins Trp the light fantastic, *Nat. Struct. Biol. 9*, 408-410.

162. Neidigh, J. W., Fesinmeyer, R. M., and Andersen, N. H. (2002) Designing a 20-residue protein, *Nat. Struct. Biol. 9*, 425-430.

163. Cochran, A. G., Skelton, N. J., and Starovasnik, M. A. (2001) Tryptophan zippers: Stable, monomeric beta-hairpins, *Proc. Natl. Acad. Sci. USA 98*, 5578-5583.

164. Bhattacharyya, R., and Chakrabarti, P. (2003) Stereospecific interactions of proline residues in protein structures and complexes, *J. Mol. Biol. 331*, 925-940.

165. Tracz, S. M., Abedini, A., Driscoll, M., and Raleigh, D. P. (2004) Role of aromatic interactions in amyloid formation by peptides derived from human amylin, *Biochemistry 43*, 15901-15908.

166. Gazit, E. (2002) A possible role for pi-stacking in the self-assembly of amyloid fibrils, *FASEB J. 16*, 77-83.

167. Marshall, K. E., Morris, K. L., Charlton, D., O'Reilly, N., Lewis, L., Walden, H., and Serpell, L. C. (2011) Hydrophobic, aromatic, and electrostatic interactions play a central role in amyloid fibril formation and stability, *Biochemistry 50*, 2061-2071.

168. Bemporad, F., Taddei, N., Stefani, M., and Chiti, F. (2006) Assessing the role of aromatic residues in the amyloid aggregation of human muscle acylphosphatase, *Protein Sci. 15*, 862-870.

169. Fraczkiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the

accessible surface areas and their gradients for macromolecules, *J. Comp. Chem. 19*, 319-333.

170. 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.

171. Wickstrom, L., Bi, Y., Hornak, V., Raleigh, D. P., and Simmerling, C. (2007) Reconciling the solution and X-ray structures of the villin headpiece helical subdomain: Molecular dynamics simulations and double mutant cycles reveal a stabilizing cation-pi interaction, *Biochemistry 46*, 3624-3634.

172. Lattman, E. E., and Rose, G. D. (1993) Protein folding: What's the question?, *Proc. Natl. Acad. Sci. USA 90*, 439-441.

173. Xiao, S., and Raleigh, D. P. (2010) A critical assessment of putative gatekeeper interactions in the villin headpiece helical subdomain, *J. Mol. Biol. 401*, 274-285.

174. Richards, F. M., and Lim, W. A. (1993) An analysis of packing in the protein folding problem, *Q. Rev. Biophys. 26*, 423-498.

175. 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.

176. Ferguson, N., and Fersht, A. R. (2003) Early events in protein folding, *Curr. Opin. Struct. Biol. 13*, 75-81.

177. Scholtz, J. M., Grimsley, G. R., and Pace, C. N. (2009) Solvent denaturation of proteins and interpretations of the m value, *Methods Enzymol. 466*, 549-565.

178. Hua, L., Zhou, R., Thirumalai, D., and Berne, B. J. (2008) Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding, *Proc. Natl.*

Acad. Sci. USA. 105, 16928-16933.

179. Horovitz, A., and Fersht, A. R. (1990) Strategy for analyzing the co-operativity of intramolecular interactions in peptides and proteins, *J. Mol. Biol. 214*, 613-617.

180. Horng, J. C., Moroz, V., Rigotti, D. J., Fairman, R., and Raleigh, D. P. (2002) Characterization of large peptide fragments derived from the N-terminal domain of the ribosomal protein L9: Definition of the minimum folding motif and characterization of local electrostatic interactions, *Biochemistry 41*, 13360-13369.

181. 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.

182. Privalov, P. L. (1989) Thermodynamic problems of protein structure, *Ann. Rev. Biophys. Biophys. Chem. 18*, 47-69.

183. Horng, J. C., Moroz, V., and Raleigh, D. P. (2003) Rapid cooperative two-state folding of a miniature alpha-beta protein and design of a thermostable variant, *J. Mol. Biol. 326*, 1261-1270.

184. Rees, D. C., and Robertson, A. D. (2001) Some thermodynamic implications for the thermostability of proteins, *Protein Sci. 10*, 1187-1194.

185. Dyer, R. B. (2007) Ultrafast and downhill protein folding, *Curr. Opin. Struct. Biol. 17*, 38-47.

186. Snow, C. D., Sorin, E. J., Rhee, Y. M., and Pande, V. S. (2005) How well can simulation predict protein folding kinetics and thermodynamics?, *Annu. Rev. Biophys. Biomolec. Struct. 34*, 43-69.

187. Kubelka, J., Hofrichter, J., and Eaton, W. A. (2004) The protein folding 'speed limit', *Curr.*

Opin. Struct. Biol. 14, 76-88.

188. Chiti, F., and Dobson, C. M. (2006) Protein misfolding, functional amyloid, and human disease, *Annu. Rev. Biochem. 75*, 333-366.

189. Braisted, A. C., and Wells, J. A. (1996) Minimizing a binding domain from protein A, *Proc. Natl. Acad. Sci. USA 93*, 5688-5692.

190. Chin, J. W., and Schepartz, A. (2001) Concerted evolution of structure and function in a miniature protein, *J. Am. Chem. Soc. 123*, 2929-2930.

191. Mer, G., Kellenberger, E., and Lefevre, J. F. (1998) alpha-helix mimicry of a beta-turn, *J. Mol. Biol. 281*, 235-240.

192. Ottesen, J. J., and Imperiali, B. (2001) Design of a discretely folded mini-protein motif with predominantly beta-structure, *Nat. Struct. Biol. 8*, 535-539.

193. 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^o, ΔS^o, and ΔC_p^o, and evaluation of solvent isotope effects, *Protein Sci. 7*, 2405-2412.

194. Barua, B., Lin, J. C., Williams, V. D., Kummler, P., Neidigh, J. W., and Andersen, N. H. (2008) The Trp-cage: optimizing the stability of a globular miniprotein, *Protein Eng. Des. Sel. 21*, 171-185.

195. Bunagan, M. R., Gao, J. M., Kelly, J. W., and Gai, F. (2009) Probing the folding transition state structure of the villin headpiece subdomain via side chain and backbone mutagenesis, *J. Am. Chem. Soc. 131*, 7470-7476.

196. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains B1 and B2:

Why small proteins tend to have high denaturation temperatures, *Biochemistry 31*, 3597-3603.

197. Lei, H. X., Wu, C., Liu, H. G., and Duan, Y. (2007) Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations, *Proc. Natl. Acad. Sci. USA 104*, 4925-4930.

198. Bi, Y. (2008) Studies of the folding and stability of the villin headpiece subdomain, *PhD thesis*, Stony Brook University.

199. Munoz, V., and Serrano, L. (1994) Elucidating the folding problem of helical peptides using empirical parameters, *Nat. Struct. Biol. 1*, 399-409.

200. Sawle, L., and Ghosh, K. (2011) How do thermophilic proteins and proteomes withstand high temperature?, *Biophys. J. 101*, 217-227.

201. Baldwin, A. J., Knowles, T. P. J., Tartaglia, G. G., Fitzpatrick, A. W., Devlin, G. L., Shammas, S. L., Waudby, C. A., Mossuto, M. F., Meehan, S., Gras, S. L., Christodoulou, J., Anthony-Cahill, S. J., Barker, P. D., Vendruscolo, M., and Dobson, C. M. (2011) Metastability of native proteins and the phenomenon of amyloid formation, *J. Am. Chem. Soc. 133*, 14160-14163.

202. Jaeger, M., Dendle, M., Fuller, A. A., and Kelly, J. W. (2007) A cross-strand Trp-Trp pair stabilizes the hPin1 WW domain at the expense of function, *Protein Sci. 16*, 2306-2313.

203. Song, B. B. (2007) Folding of alfa-helical proteins and protein design using non-coded amino acids, *PhD thesis*, Stony Brook University.

204. Pace, C. N., Trevino, S., Prabhakaran, E., and Scholtz, J. M. (2004) Protein structure, stability and solubility in water and other solvents, *Philos. Trans. R. Soc. Lond. B Biol. Sci. 359*, 1225-1234.

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

206. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M. P., Thurlkill, R. L., Scholtz, J. M., and Pace, C. N. (1999) Increasing protein stability by altering long-range coulombic interactions, *Protein Sci. 8*, 1843-1849.

207. Spector, S., Wang, M. H., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B., and Raleigh, D. P. (2000) Rational modification of protein stability by the mutation of charged surface residues, *Biochemistry 39*, 872-879.

208. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M., and Makhatadze, G. I. (1999) Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface, *Biochemistry 38*, 16419-16423.

209. Strickler, S. S., Gribenko, A. V., Keiffer, T. R., Tomlinson, J., Reihle, T., Loladze, V. V., and Makhatadze, G. I. (2006) Protein stability and surface electrostatics: A charged relationship, *Biochemistry 45*, 2761-2766.

210. Schweiker, K. L., Zarrine-Afsar, A., Davidson, A. R., and Makhatadze, G. I. (2007) Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge-charge interactions, *Protein Sci. 16*, 2694-2702.

211. Schwehm, J. M., Fitch, C. A., Dang, B. N., Garcia-Moreno, B., and Stites, W. E. (2003) Changes in stability upon charge reversal and neutralization substitution in staphylococcal nuclease are dominated by favorable electrostatic effect, *Biochemistry 42*, 1118-1128.

212. Pace, C. N. (2000) Single surface stabilizer, *Nat. Struct. Biol. 7*, 345-346.

213. Hubbard, S. J., and Thornton, J. M. (1993) NACCESS Computer Program, Department of Biochemistry and Molecular Biology, University College London.

214. Miller, S., Janin, J., Lesk, A. M., and Chothia, C. (1987) Interior and surface of monomeric proteins, *J. Mol. Biol. 196*, 641-656.

215. 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.

216. Swintkruse, L., and Robertson, A. D. (1995) Hydrogen bonds and the pH-denpendence of ovomucoid third domain stability, *Biochemistry 34*, 4724-4732.

217. Tan, Y. J., Oliveberg, M., Davis, B., and Fersht, A. R. (1995) Perturbed pKa values in the denatured states of proteins, *J. Mol. Biol. 254*, 980-992.

218. Marti, D. N., and Bosshard, H. R. (2004) Inverse electrostatic effect: Electrostatic repulsion in the unfolded state stabilizes a leucine zipper, *Biochemistry 43*, 12436-12447.

219. Tollinger, M., Crowhurst, K. A., Kay, L. E., and Forman-Kay, J. D. (2003) Site-specific contributions to the pH dependence of protein stability, *Proc. Natl. Acad. Sci. USA. 100*, 4545-4550.

220. Trefethen, J. M., Pace, C. N., Scholtz, J. M., and Brems, D. N. (2005) Charge-charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa, *Protein Sci. 14*, 1934-1938.

221. Perez-Jimenez, R., Godoy-Ruiz, R., Ibarra-Molero, B., and Sanchez-Ruiz, J. M. (2005) The effect of charge-introduction mutations on *E coli* thioredoxin stability, *Biophys. Chem. 115*, 105-107.

222. Freddolino, P. L., and Schulten, K. (2009) Common structural transitions in

explicit-solvent simulations of villin headpiece folding, *Biophys. J. 97*, 2338-2347.

223. Bartlett, A. I., and Radford, S. E. (2009) An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms, *Nat. Struct. Mol. Biol. 16*, 582-588.

224. Havlin, R. H., and Tycko, R. (2005) Probing site-specific conformational distributions in protein folding with solid-state NMR, *Proc. Natl. Acad. Sci. USA. 102*, 3284-3289.

225. Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011) How fast-folding proteins fold, *Science 334*, 517-520.

226. Zagrovic, B., and Pande, V. S. (2006) Simulated unfolded-state ensemble and the experimental NMR structures of villin headpiece yield similar wide-angle solution X-ray scattering profiles, *J. Am. Chem. Soc. 128*, 11742-11743.

227. Wickstrom, L., Okur, A., Song, K., Hornak, V., Raleigh, D. P., and Simmerling, C. L. (2006) The unfolded state of the villin headpiece helical subdomain: Computational studies of the role of locally stabilized structure, *J. Mol. Biol. 360*, 1094-1107.

228. Khan, F., Chuang, J. I., Gianni, S., and Fersht, A. R. (2003) The kinetic pathway of folding of barnase, *J. Mol. Biol. 333*, 169-186.

229. Shen, J. K. (2010) A method to determine residue-specific unfolded-state pKa values from analysis of stability changes in single mutant cycles, *J. Am. Chem. Soc. 132*, 7258-7259.

230. Khandogin, J., Chen, J., and Brooks, C. L., III. (2006) Exploring atomistic details of pH-dependent peptide folding, *Proc. Natl. Acad. Sci. USA. 103*, 18546-18550.

231. Zdanowski, K., and Dadlez, M. (1999) Stability of the residual structure in unfolded BPTI in different conditions of temperature and solvent composition measured by

disulphide kinetics and double mutant cycle analysis, *J. Mol. Biol. 287*, 433-445.

232. Zarrine-Afsar, A., Wallin, S., Neculai, A. M., Neudecker, P., Howell, P. L., Davidson, A. R., and Chan, H. S. (2008) Theoretical and experimental demonstration of the importance of specific nonnative interactions in protein folding, *Proc. Natl. Acad. Sci. USA. 105*, 9999-10004.

Appendix 1. Summary of pH-dependent thermodynamic parameters for the unfolding of HP36 wildtype and the three Lys-to-Met mutants. All experiments were conducted in 10 mM sodium acetate, 150 mM sodium chloride solution.

pH	ΔG^o (kcal mol^{-1}) [1]			
	HP36 WT	K48M	K65M	K70M
2.0	1.11 ±0.05	1.62	1.41	1.90 ±0.08
2.5	1.32	1.78 ±0.07	1.52	2.01
3.0	1.67	2.11	1.86	2.61
3.5	1.95 ±0.13	2.33 ±0.12	2.25	2.81 ±0.10
4.0	2.21	2.88	2.80	3.12
4.5	2.82	3.51	3.52	3.75
5.0	3.17 ±0.06	4.11	3.87	4.26 ±0.19
5.5	3.3	4.11	3.90	4.42
6.0	3.39	4.26 ±0.07	4.03	4.42

Errors are the standard deviation of three independently repeated experiments.
1) ΔG^o values at pH from 2.0 to 3.0 are the calculated stabilities from thermal unfolding experiments using the Gibbs-Helmholtz equation with a ΔC_p^o of 0.38 kcal mol^{-1} K^{-1}. ΔG^o values at pH from 3.5 to 6.0 are the measured stabilities from urea induced unfolding experiments.

Appendix 2. Summary of pH-dependent sidechain carbonyl carbon chemical shift measurements of the acidic residues in the native state of HP36. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

pH	δ carbonyl carbon (ppm)			
	D44	E45	D46	E72
6.37	176.9	181.2	176.3	180.5
6.04	176.9	181.2	176.3	180.4
5.46	176.9	181.1	176.3	180.2
5.07	176.8	180.9	176.3	179.9
4.87	176.8	180.8	176.1	179.6
4.61	176.8	180.4	176.1	179.2
4.22	176.6	179.6	175.8	178.3
3.79	176.4	178.7	175.3	177.6
3.55	176.1	178.2	174.8	177.2
3.47	176.2	178.0	174.7	177.2
3.04	175.6	177.5	174.2	177.1
2.89	175.1	177.3	173.0	176.9
2.73	174.1	177.2	173.8	177.0
2.46	174.5	177.0	173.0	176.8
2.40	174.2	177.0	173.6	176.9
2.09	174.2	176.9	173.0	176.7

Appendix 3. Summary of pH-dependent α proton chemical shift measurements of the C-terminus in the native state of HP36. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

pH	δHα (ppm)
5.59	4.36
5.35	4.36
4.83	4.36
4.50	4.36
4.29	4.36
3.80	4.36
3.44	4.49
3.13	4.49
2.31	4.55
1.64	4.55

Appendix 4. Summary of pH-dependent proton chemical shift measurements of the acidic residues as well as the C-terminus in the native state of K48M mutant. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

pH	δHβ D44 (ppm)	δHγ E45 (ppm)	δHβ D46 (ppm)	δHγ E72 (ppm)	δHα C-terminus (ppm)
7.09	2.71	2.33	2.69	2.50	4.45
6.66	2.71	2.33	2.69	2.50	4.45
5.85	2.73	2.33	2.69	2.50	4.45
5.37	2.73	2.36	2.69	2.52	4.45
4.76	2.73	2.41	2.71	2.57	4.45
4.42	2.73	2.43	2.71	2.62	4.45
3.97	2.76	2.50	2.76	2.66	4.47
3.39	2.85	2.54	2.85	2.71	4.51
2.83	2.92	2.52	2.92	2.73	4.56
2.26	2.99	2.52	2.99	2.73	4.63

Appendix 5. Summary of pH-dependent proton chemical shift measurements of the acidic residues as well as the C-terminus in the native state of K65M mutant. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

pH	δHβ D44 (ppm)	δHγ E45 (ppm)	δHβ D46 (ppm)	δHγ E72 (ppm)	δHα C-terminus (ppm)
6.14	2.78	2.45	2.71	2.57	4.47
5.54	2.78	2.45	2.71	2.57	4.47
4.90	2.78	2.45	2.71	2.57	4.47
4.56	2.78	2.48	2.73	2.62	4.47
4.14	2.78	2.50	2.76	2.69	4.47
3.67	2.83	2.53	2.80	2.71	4.49
3.10	2.90	2.55	2.90	2.73	4.54
2.67	2.97	2.55	2.97	2.73	4.61
2.21	3.01	2.55	2.99	2.73	4.63

Appendix 6. Summary of pH-dependent proton chemical shift measurements of the acidic residues as well as the C-terminus in the native state of K70M mutant. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride.

pH	δHβ D44 (ppm)	δHγ E45 (ppm)	δHβ D46 (ppm)	δHγ E72 (ppm)	δHα C-terminus (ppm)
7.07	2.76	2.38	2.69	2.31	4.45
6.18	2.76	2.38	2.69	2.31	4.45
5.73	2.76	2.38	2.69	2.31	4.47
4.97	2.78	2.39	2.69	2.38	4.47
4.65	2.78	2.41	2.69	2.43	4.47
4.30	2.78	2.43	2.71	2.50	4.47
3.89	2.80	2.50	2.76	2.57	4.49
3.29	2.85	2.54	2.83	2.62	4.56
2.84	2.92	2.54	2.90	2.64	4.61
2.38	2.97	2.53	2.94	2.64	4.66

Appendix 7. Summary of pH-dependent proton chemical shift measurements of the acidic residues of the MLSDEDFKAVFGM peptide. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has a free N-terminus and an amidated C-terminus.

pH	δHβ D44 (ppm)	δHγ E45 (ppm)	δHβ D46 (ppm)
7.25	2.58	2.17	2.55
6.04	2.62	2.17	2.55
5.38	2.61	2.18	2.55
5.02	2.63	2.20	2.56
4.82	2.63	2.21	2.56
4.54	2.66	2.24	2.59
4.23	2.70	2.28	2.62
4.02	2.70	2.28	2.63
3.60	2.75	2.31	2.68
3.13	2.80	2.32	2.72
2.83	2.83	2.32	2.74
2.56	2.81	2.32	2.75
2.26	2.83	2.32	2.76
1.93	2.82	2.32	2.75

Appendix 8. Summary of pH-dependent proton chemical shift measurements of the acidic residues of the MLSDEDFMA peptide. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has a free N-terminus and an amidated C-terminus.

pH	δHβ D44 (ppm)	δHγ E45 (ppm)	δHβ D46 (ppm)
7.02	2.73	2.28	2.67
6.32	2.73	2.28	2.67
5.64	2.73	2.28	2.67
5.10	2.74	2.31	2.67
4.62	2.77	2.35	2.73
3.98	2.84	2.42	2.77
3.36	2.91	2.45	2.84
3.06	2.91	2.45	2.84
2.42	2.95	2.45	2.87
2.17	2.95	2.46	2.87

Appendix 9. Summary of pH-dependent proton chemical shift measurements of the acidic residues of the MLSDQDFKAVFGM peptide. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has a free N-terminus and an amidated C-terminus.

pH	δHβ D44 (ppm)	δHβ D46 (ppm)
6.32	2.73	2.66
5.90	2.73	2.67
5.20	2.74	2.67
4.56	2.77	2.70
4.13	2.80	2.70
3.80	2.84	2.73
3.28	2.91	2.79
2.85	2.95	2.81
2.38	2.95	2.82

Appendix 10. Summary of pH-dependent proton chemical shift measurements of the acidic residue as well as the C-terminus of the KEKGLF peptide. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has an acetylated N-terminus and a free C-terminus.

pH	δHγ E72 (ppm)	pH	δHα C-terminus (ppm)
7.18	2.26	7.18	4.43
5.58	2.28	5.58	4.43
5.07	2.30	5.07	4.43
4.74	2.32	4.74	4.44
4.55	2.34	4.55	4.44
4.26	2.37	4.26	4.46
3.94	2.40	3.94	4.48
3.49	2.44	3.49	4.53
3.16	2.45	3.40	4.53
2.81	2.46	2.90	4.58
2.51	2.46	2.68	4.62
2.12	2.47	2.30	4.65
		2.03	4.66

Appendix 11. Summary of pH-dependent proton chemical shift measurements of the acidic residue as well as the C-terminus of the KKEKGLF peptide. Measurements were performed at 25 ^{o}C in 90% H_2O, 10% D_2O 10 mM sodium acetate, 150 mM sodium chloride. The peptide has an acetylated N-terminus and a free C-terminus.

pH	δHγ E72 (ppm)	δHα C-terminus (ppm)
6.44	2.26	4.43
5.90	2.27	4.44
5.44	2.27	4.44
4.95	2.29	4.44
4.62	2.31	4.45
4.10	2.37	4.46
3.70	2.41	4.50
3.35	2.44	4.54
2.84	2.46	4.60
2.27	2.46	4.64
2.00	2.46	4.67