Stony Brook University

Studies of the Folding and Stability of the Villin Headpiece Subdomain

A Dissertation Presented

by

Yuan Bi

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Chemistry

Stony Brook University
May 2008

Stony Brook University

The Graduate School

Yuan Bi

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

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

Nicole S. Sampson, Ph. D., Chairperson
Department of Chemistry, Stony Brook University

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

Carlos de los Santos, Ph. D., Outside Member
Department of Pharmacology, Stony Brook University

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

Studies of the Folding and Stability of the Villin Headpiece Subdomain

by

Yuan Bi

Doctor of Philosophy
in
Chemistry

Stony Brook University

2008

The helical C-terminal subdomain of the villin headpiece is denoted HP36. HP36 is one of the smallest naturally occurring sequences which folds cooperatively. My work focuses on the folding and stability studies of HP36.

Using fusion technology, I demonstrate the high level expression of a small, 36 residue, three helix bundle, the villin headpiece subdomain. HP36 is linked via a Factor Xa cleavage sequence to the C-terminus of the N-terminal domain of the ribosomal protein L9 (NTL9). The yield of purified fusion protein is 70 mg per liter of culture and the yield of purified villin headpiece subdomain is 24 mg per liter of culture. I also demonstrate the use of the fusion system to express a smaller marginally folded peptide fragment of the villin headpiece domain.

The contribution of charged residues to the stability of HP36 was investigated by conducting a systematic mutational study of twelve point mutants. Every charged residue is individually mutated to a neutral residue or reverse charged residue. Thermodynamic measurements were carried out by CD. The results show that mutation of any of the acidic residues leads to a decrease in T_m and ΔG^o of unfolding; in contrast three of the seven mutations which targeted basic residues lead to significant increase in stability.

A hyperstable double mutant N68AK70M of HP36 was designed. CD and Amide H/D exchange experiments were used to measure the stability of the N68AK70M double mutant. I demonstrate that the double mutant was significant more stable than wild type. The double mutant did not display a complete unfolding transition either by thermal unfolding or by urea unfolding. In 3M urea the Tm of wild type is 54.8 ^oC while that of the double mutant is 73.9 ^oC. Amide H/D exchange studies show that the pattern of exchange is very similar for HP36 and the double mutant. The structure of the K70M single mutant and the double mutant were determined by x-ray crystallography and are identical to wild type. Analytical ultracentrifugation demonstrates that the proteins are monomeric.

The side chain interactions of HP36 were also investigated. The thermal and urea

unfolding experiments have shown the difference of the stability between WT HP36 and its mutants (D44N, R55M, F47L, D44NR55M, and F47LR55M). All the mutants are less stable than the WT HP36. Experimental double mutant cycles confirmed that the F47/R55 pair has a larger energetic coupling than the D44/R55 interaction, suggesting the formation of a cation-pi interaction between F47 and R55 which is consistent with molecular dynamics simulations.

Table of Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

AUC	Analytical ultracentrifugation
a_d	The y intercept for a post-transition line
a_n	The y intercept for a pre-transition line
b_d	The slope for a post-transition line
b_n	The slope for a pre-transition line
CD	Circular dichroism
C_M	The midpoint concentration of a denaturant induced unfolding transition
D	Deuterium
DSSP	Dictionary of Secondary Structure Prediction
RMSD	Root Mean Squared Deviation; GB, Generalized Born
PME	Particle Mesh Ewald
IPS	Isotropic Periodic Sum
MD	Molecular Dynamics
PMF	Potential of Mean Force
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ESI-MS	Electrospray ionisation mass spectrometry
F	The folded state
GB	Generalized Born
Gnd HCl	Guanidine hydrochloride
HSQC	Heteronuclear single-quantum coherence
IEC	Ion exchange chromatograph
IPS	Isotropic periodic sum
IPTG	Isopropylthio-β-D-galactoside
k_{ex}	Exchange rate, $k_{ex} = k_f + k_u$
k_f	Folding rate
k_u	Unfolding rate
MALDI-TOF	Matrix assisted laser desorption and ionization time of flight
NMR	Nuclear magnetic resonance
m	Slope of $\Delta G°$ vs. denaturant concentration
p_D	Population of unfolded (denatured) protein
p_N	Population of native protein
PME	Particle mesh Ewald
PMF	Potential of mean force
PMSF	Phenylmethanesulfonyl fluoride
ppm	Parts per million
R	Universal gas constant
RMSD	Root mean square deviation
RP-HPLC	Reverse phase high performance liquid chromatography
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TFA	Trifloroacetic acid
T	Temperature

T_m	Midpoint of thermal denaturation transition
TMAO	Trimethylamine *N*-oxide
TSP	3-(Trimethylsilyl)propionate
U	The unfolded state
UV	Ultraviolet
$\Delta C°_p$	Heat capacity change of unfolding
$\Delta G°_{D-N}$	Standard state free energy of unfolding
$\Delta H°$	Standard state enthalpy of unfolding
$\Delta S°$	Standard state entropy of unfolding
[den]	Concentration of denaturant
$[\theta]$	Mean residue ellipticity
1D ^{1}H-NMR	One-dimensional proton nuclear magnetic resonance

Acknowledgments

I would like to thank my advisor Daniel Raleigh for his excellent guidance. I would not accomplish all my work without his great ideas, detailed suggestions, encouragement and patience. Especially for his encouragement and support when I was experiencing the hard time in my research and my life in Stony Brook.

I want to thank professor Nicole Sampson, professor Carlos Simmerling and professor Carlos de los Santos for serving as my committee members. I thank them for their valuable suggestions on my research projects.

Thanks for the friendship from all the members of the Raleigh group past and present: Dr. Farhana Chowdhury, Dr. Jia-Cherng Horng, Dr. Sylvia Tracz, Dr. Jae-hyun Cho, Dr. Andisheh Abedini, Dr. Burcu Anil, Dr. Ying Li, Dr. Yuefeng Tang, Dr. Benben Song, Lauren Wickstrom, Ruchi Gupta, Bing Shan, Humeyra Taskent, Fanling Meng, Shifeng Xiao, Wenli Meng, Konstantine Aprilakis, Peter Marek.

I thank my family members for their encouragement and unconditional support, especially my parents and parents in law, for their great help to me for taking care of my lovely son Evan. Lastly, I thank my husband Jun for his love and being always on my side.

List of Publications

1. F. A. Chowdhury, R. Fairman, **Y. Bi**, D. J. Rigotti, and D. P. Raleigh (2004) Protein dissection experiments reveal key differences in the equilibrium folding of alpha-lactalbumin and the calcium binding lysozymes. *Biochemistry* 43: 9961-9967.
2. **Y. Bi**, Y. Tang, D. P. Raleigh and J. Cho (2006) Efficient high level expression of peptides and proteins as fusion proteins with the N-terminal domain of L9: application to the villin headpiece helical subdomain. *Protein Expr Purif* 47, 234-40.
3. L. Wickstrom, **Y. Bi**, V. Hornak, D. P. Raleigh and C. L. Simmerling (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* 46,3624-34.
4. **Y. Bi**, J. Cho, E. Kim, B. Shan, H. Schindelin and D. P. Raleigh (2007) Rational Design, Structural and Thermodynamic Characterization of a Hyperstable Variant of the Villin Headpiece Subdomain. *Biochemistry* 46, 7497-7505.
5. **Y. Bi**, B. Shan and D. P. Raleigh, Contribution of Charged Residues to the Stability of a Miniature Protein Domain: the Villin Headpiece Subdomain. (*To be Submitted*)

1. Introduction

1.1. Protein folding and staility

1.1.1. Overview of protein folding

In all living systems, proteins carry out the most important tasks that make life possible. In order to carry out their biological functions, proteins must fold to their proper, unique three-dimensional structure. Failure to fold the native conformation may lead to serious problem. A number of diseases have been linked to protein folding problems. These include prion diseases such as bovine spongiform encephalopathy (BSE) and its human equivalent Creutzfeld-Jakob disease (CJD), and also Alzheimer's disease, Parkinson's disease and type II (non-insulin dependent) diabetes (*1-9*). Understanding the mechanism of protein folding may help the development of the therapies for those diseases. Improved knowledge of the factors that affect protein stability is also critical for understanding and controlling misfolding.

1.1.2. The mechanism of protein folding

The protein folding mechanism has been studied for more than fifty years (*10*). Protein folding are often classified as either a two-state process (*11*) or a multi-state process (*12, 13*):

$U \rightleftharpoons F$ (two-state process, without intermediates)

Where "F" refers to the native state and "U" is the denatured state of a protein. In this case, the protein only has two thermodynamically stable states, folded and unfolded. The folding or unfolding process is cooperative without any intermediate(s), or with very unstable intermediate(s) that cannot be detected. Some proteins form intermediates beside the folded and unfolded states. The simplest case is the three-state transition model, which involves a stable folding intermediate:

$U \rightleftharpoons I \rightleftharpoons F$ (three-state process, "I" stands for the stable intermediate)

Much more complicated models are possible with parallel pathways and with off pathway intermediates, these more complicated mechanisms often apply to larger proteins. More recently, a "new view" of protein folding termed "folding funnel" theory has been developed (*14-16*). In this theory, the energy landscape associates every protein conformation with its potential energy and the folding process can be viewed as a kind of free-energy funnel. The unfolded states are characterized by a high degree of conformational entropy and relatively high energy corresponding to the top of the funnel which has more conformational states, while the bottom of the funnel corresponds to the native state which has only a single conformation for a specific protein of course there are fluctuations around that minimum . A free-energy funnel is shown in Figure 1.1, representing various folding pathways from the

unfolded to native states: instead of a simple folding pathway, a protein molecule may follow the (yellow) fast folding or (green) slow folding pathway that crosses the high energy barrier, or (red) slow folding pathway which returns to a less folded state before following the pathway for fast folding. The significance of this free-energy funnel theory is that it describes the folding pathways in a more general and comprehensive way, which gives information about the local energy minima during the folding steps, and the energy-barriers, between distinct folding intermediates.

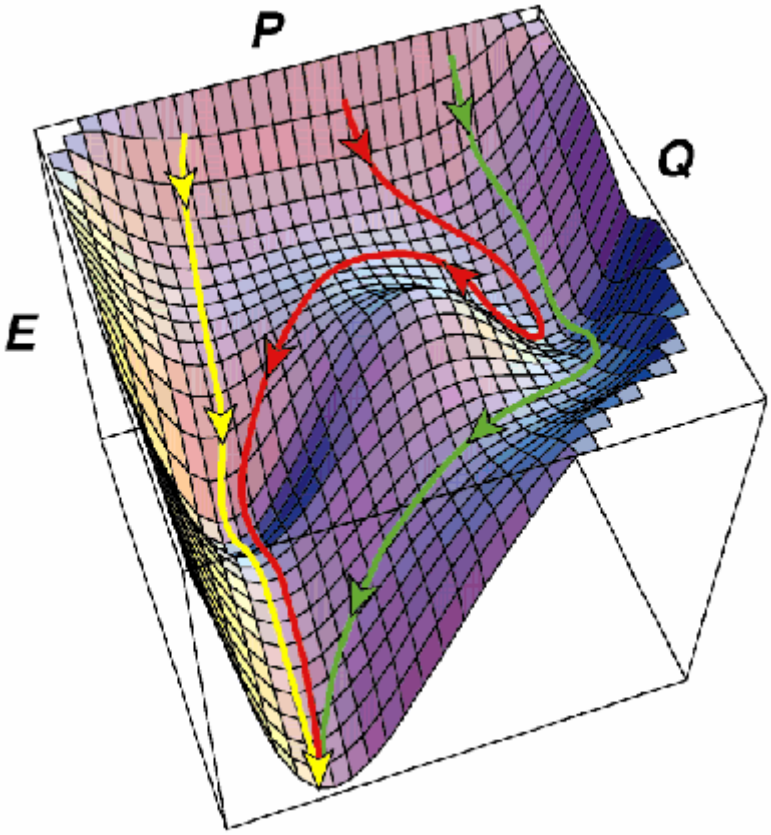


Figure 1.1: Three-dimensional representation of folding funnel. E represents the energy of the system, Q is the proportion of native contacts formed, and P is a measure of the available conformational space. Adapted from Dobson *et al.*, 1998 (*16*).

1.1.3. Major forces contributing to protein stability

When a protein folds, most of the nonpolar side chains are removed from water and form a hydrophobic core. This burial of the nonpolar groups is referred to as the hydrophobic effect. It is believed that the hydrophobic effect is the major driving force for the folding of globular proteins (*17-19*). The importance of the hydrophobic effect is most easily probed by site-directed mutagenesis experiments combined with thermal and spectroscopic stability experiments on proteins. Many proteins have been studied by mutating the hydrophobic core residues (*20, 21*). The burial of a -CH_2- contributes 1.2 kcal mol^{-1} per $Å^3$ to the stability of a protein based on the experimental studies of hydrophobic interactions in protein (*22*).

Hydrogen bond is the other major force stabilizing the native state of proteins (*23*). In 1951, Pauling's group described the α-helix and the β-sheet, two most important secondary structures found in proteins (*24*). In these structures, the polar groups are almost completely hydrogen bonded (*25*). To understand the contribution

of hydrogen bonds to protein stability, mutational studies of Tyr→Phe in 52 proteins and Thr→Val in 40 proteins clearly demonstrated that hydrogen bonds contribute favorable to protein stability (*22, 26, 27*). Based on the experiments, when the Tyr-OH (Thr-OH) group was hydrogen bonded, the Tyr→Phe (Thr→Val) mutants were 1.4 kcal mol^{-1} (0.9 kcal mol^{-1}) less stable than wild type; they were only slightly less stable than wild type when the Tyr-OH (Thr-OH) group was not hydrogen bonded. A recent review argued that the contributions of hydrogen bond and van der Waals interactions to protein stability is significant (*28*).

Electrostatic interactions are not the major contributor to protein stability but they do affect protein stability (*29-33*). Electrostatic interactions vary, sometimes making favorable contributions and sometimes making unfavorable contributions to protein stability due to the fact that they can be attractive and repulsive. They are also related to the ordering and desolvation of charges (*34*). Early, the role of surface charged residues in protein stability received much less attention comparing with hydrophobic effect or hydrogen bonds (*28, 35*), and it was thought that the surface charged residues are not important for protein stability, because their interactions with the solvent should be similar in the native and unfolded states. Furthermore the high dielectric of the solvent will decrease the strength of the charge-charge interactions and thus lead to a further decrease in the contribution of these residues to protein stability (*36*). However, several studies including work from our laboratory showed the importance of charge-charge interactions for protein stability (*29, 32, 37-41*). Several studies have shown that protein stability can be increased by targeting charged residues.

1.2. The thermodynamics and kinetics of protein folding

Understanding the physical determinates of protein stability is very important for understanding how proteins fold(*42*). The conformational stability of a protein is the difference in the Gibbs free energy between the folded state and the unfolded state.

$$\Delta G^o_{N-D} = G^o_D - G^o_N \tag{1.1}$$

In a two-state cooperative mechanism, at standard conditions, the Gibbs free energy is defined as:

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

Where $\Delta H^o_{N\text{-}D}$ is the difference in enthalpy and $\Delta S^o_{N\text{-}D}$ is the difference in entropy.

$$\Delta H^o_{N-D} = \Delta C^o_P\left(T - T_o\right) + \Delta H^o_{N-D}\left(T_o\right) \tag{1.3.1}$$

$$\Delta S^o_{N-D} = \Delta C^o_P \ln\left(\frac{T}{T_o}\right) + \Delta S^o_{N-D}\left(T_o\right) \tag{1.3.2}$$

Substituting the equations 1.3.1 and 1.3.2 into equation (1.2) gives the Gibbs-Helmholtz equation (1.4):

$$\Delta G^o_{N-D}\left(T\right) = \Delta H^o\left(T_o\right) - T\Delta S^o\left(T_o\right) + \Delta C^o_P\left[T - T_o - T\ln\left(\frac{T}{T_o}\right)\right] \tag{1.4}$$

This equation describes the temperature dependence of the Gibbs free energy of unfolding. When the reference temperature T_o is the midpoint of the thermal

denaturation experiment, T_m, the system is in an equilibrium state with equal amount of folded and unfolded states, $\Delta G^o_{N-D}(T_m) = 0$ and $\Delta S^o = \Delta H^o(T_m)/T_m$. Then we can write equation (1.4) as:

$$\Delta G^o_{N-D}(T) = \Delta H^o(T_m)\left(1 - \frac{T}{T_m}\right) + \Delta C^o_P\left[T - T_m - T\ln\left(\frac{T}{T_m}\right)\right] \quad (1.5)$$

$\Delta H^o(T_m)$ can be measured from fitting thermal unfolding curves or from differential scanning calorimetry (DSC). The change of heat capacity (ΔC^o_p) upon unfolding can also be obtained from DSC or from pH dependent thermal unfolding. The change of heat capacity is assumed as independent of temperature when deriving equation (1.4) and it is proportional to the change of non-polar surface area that is buried in the folded state compared to that exposed in the unfolded state. The dependence of ΔC^o_p on non-polar surface area is due to the hydophobic effect and classically is thought to reflect the solvent ordering around the hydophobic solute. The value of ΔC^o_p of unfolding is estimated to be ~12 cal mol^{-1} K^{-1} per residue of protein (*42*).

Protein denaturation can also be studied at constant temperature by varying the denaturant (Urea or Guanidine hydrochloride) concentration or pH conditions. The denaturants solubilize all the constituent parts of a protein, from its polypeptide backbone to its hydrophobic side chains. Because the denatured state is more exposed to solvent than the native state, the denatured state is preferentially stabilized by the denaturant. The free energy of unfolding (ΔG^o_{N-D}) is assumed to be linearly proportional to the concentration of denaturant:

$$\Delta G^o_{N-D} = \Delta G^o_{N-D}(H_2O) - m[denaturant] \quad (1.6)$$

where $\Delta G^o_{N-D}(H_2O)$ is the value in water, m is the slope, equation (1.6) shows the linear dependence of ΔG^o_{N-D} on the denaturant concentration and reveals quantitative information about the buried surface area change upon unfolding.

All of the denaturation experiments can be monitored by Circular dichroism (CD) spectroscopy. CD can monitor the spectroscopic properties change between the folded and unfolded state of the protein. In a two-state mechanism, there is an equilibrium between the native state (N) and the denatured state (D), the equilibrium constant can be expressed as:

$$K_{eq} = \frac{[D]}{[N]} = \frac{P_D}{1 - P_D} \quad (1.7)$$

where [N] and [D] are the concentration of the native state and the denatured state respectively. P_D represents the fraction of unfolded state, (1- P_D) or P_N represents the fraction of folded state. In any equilibrium condition, the observed signal (y) is related to both the native signals (y_N) and the denatured signals (y_D) and their corresponding populations:

$$y = y_N P_N + y_D P_D \quad (1.8)$$

By combining equation 1.7 and 1.8, $\Delta G^{o}_{N\text{-}D}$ can be calculated as equation 1.9:

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

All thermal denatureations can be analyzed by a non-linear least squares curve fitting equation:

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

Where a_n and b_n represents the intercept and slope of the pre-transition. a_d and b_d represent intercept and slope of the post-transition. $\Delta G^{\circ}_{N\text{-}D}(T)$ is given by the Gibbs-Helmholtz equation.

The fraction of unfolded state is:

$$P_D = \frac{y(T) - (a_n + b_n T)}{(a_d + b_d T) - (a_n + b_n T)} \tag{1.11}$$

Chemical denaturation curves can also be analyzed by a non-linear least squares curve fitting equation:

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

Amide hydrogen (NH) exchange measurements by nuclear magnetic resonance (NMR) is another method to measure protein stability at residue-level resolution (*43, 44*). When a protein is transferred to D_2O, the amide hydrogens starts to exchange with deuterium, and the rate constants for each individual residues can be measured by NMR. The exchange rate can be described:

$$(\text{N-H}\cdots\text{O=C})_{\text{cl}} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} (\text{N-H})_{\text{op}} \xrightarrow{k_{ch}} \text{N-D} \tag{1.13}$$

k_{op} and k_{cl} are the rate constants for the structure opening and closing reaction, and k_{ch} is the rate constant for exchange from the open state (*45*). Under conditions where $k_{cl} >> k_{ch}$, the exchange rate constant $k_{ex} = (k_{op}/k_{cl})k_{ch}$, this is known as EX2 exchange. $K_{op} = k_o / k_{cl}$ is the equilibrium constant for structure opening, so the free energy change for structure opening can be expressed by equation 1.14:

$$\Delta G^{o}_{ex} = -RT \ln K_{op} = -RT \ln\left(k_{ex}/k_{ch}\right) \tag{1.14}$$

The exchange rate constants were calculated by fitting the normalized peak volumes to a first-order exponential decay:

$$V(t) = V_o e^{-(k_{ex}t)} + b \tag{1.15}$$

where $V(t)$ is normalized peak volume, V_o is normalized peak volume at time zero, k_{ex} is exchange rate constant and b is the baseline correction term.

1.3. Introduction to analytical ultracentrifugation

Analytical ultracentrifugation is a very versatile and powerful technique for characterizing the solution-state behavior of macromolecules (*46*). An analytical ultracentrifuge combines a preparative ultracentrifuge with an optical detection system that can directly measure the sample concentration inside the centrifuge cell during sedimentation.

There are two types of experiments that are performed with the analytical ultracentrifuge: sedimentation velocity and sedimentation equilibrium. Sedimentation

velocity experiments measure the rate at which a macromolecule moves when given radial acceleration in an ultracentrifuge. The rate depends on the mass and shape of the macromolecular species. Sedimentation equilibrium only depends on the mass of the macromolecular species. Sedimentation equilibrium experiments can help to determine the quaternary structure of associating macromolecules and to study the stoichiometry of heterogeneous binding reactions. This technique depends on the measurement of the solute concentration distribution in the sample cell when equilibrium conditions have been attained, that is, when the distribution of solute due to sedimentation is balanced by the distribution due to diffusion (*47*). That means: upon reaching equilibrium, there will be no net transport of matter if the ultracentrifuge continues to operate at the same rotor speed and temperature. So the flow J=0 (*48*):

$$L[\omega^2 rM(1-\bar{v}\rho) - \frac{RT}{C}\frac{dC}{dr}] = 0 \tag{1.16}$$

$$\frac{1}{C}\frac{dC}{dr} = \frac{\omega^2 rM(1-\bar{v}\rho)}{RT} \tag{1.17}$$

Equation (1.17) describes the concentration gradient at equilibrium for a single solute component in an ideal two-component solution. By integrating between the meniscus r_0 and some point r, we can see that C depends on r in an exponential fashion:

$$C_r = C_{r_0} e^{[M(1-\bar{v}\rho)\omega^2 r/2RT](r^2 - r_0{}^2)} \tag{1.18}$$

Alternatively, we can write this as:

$$\frac{d\ln(C_r)}{dr^2} = \frac{\omega^2 M(1-\bar{v}\rho)}{2RT} \tag{1.19}$$

where C_r is the concentration of macrosolute at any radial distance, C_{r_0} is the concentration of macrosolute at the reference radial distance r_0, $\bar{v}$ is the partial specific volume of macrosolute, ω is the angular velocity (ω=2π rpm/60, where rpm is revolutions per minute), the ρ is the density of solution, R is the gas constant, T is the absolute temperature, and M is the molecular weight.

In practice, we often write this equation with absorbance, A, substituted for concentration, C, assuming by Beer's law that absorbance is proportional to concentration. If plotting ln (C_r) versus r^2, if it gives a straight line with a slope that equals $M\omega^2(1-\bar{v}\rho)/2RT$, we can get M. However a downward curving plot indicates the molecular weight decreases with increasing concentration, is an indication that the solution is nonideal, and that the molecular weight exhibits concentration dependence. An upward curving plot is an indication that the macrosolute is polydisperse. I.e. there is a distribution of molecular weights either because it is impure or because the material is aggregated. In this case $d\ln(C_r)/dr^2$ yields an average molecular weight that,

providing all macrosolute species have the same partial specific volume, gives the weight-average molecular weight, M_w. Then we can use global methods (*49*) to analyze the data.

1.4. Rational strategies for the stabilization of proteins

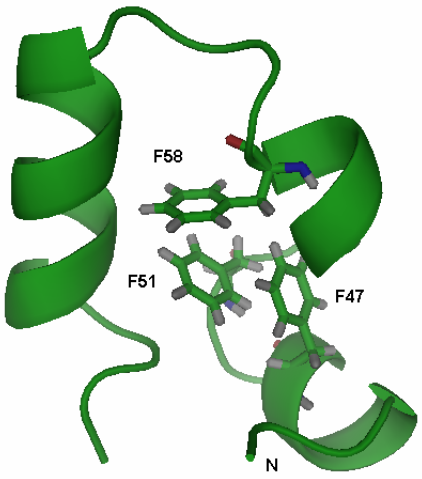
Protein design is becoming a powerful approach for understanding the relationship between amino acid sequence and three-dimensional structure, as well as an useful tools for optimizing protein drugs and creating novel biotherapeutics (*50*). Rational design is a protein engineering strategy in which people use detailed knowledge of the structure and function of the protein to make desired changes. Site-directed mutagenesis techniques are used to introduce these changes into the protein (*51-53*). The advantage of this method is that site-directed mutagenesis is generally inexpensive and easy. Nowadays, computer based rational design of protein is becoming more and more popular because computational methods can, in theory identify amino acid sequences that have low energies in target structure (*54-56*). However, a major drawback is that the detailed structural knowledge of a protein is needed depending on the design. Even if the structure is known, it can be extremely difficult to predict the effects of various mutations, especially long-range effects. More and more researchers use both methods instead of using them alone. There is a concept in protein design referred as "design cycle" in which the theory and experiments are used together (*57*). First, based on the protein three-dimensional structure and sequence as well as the experimental analysis of the target protein, a molecular model in combination with an algorithm is built. The second step is to model the selected mutants and check whether the desired properties have improved and whether the mutation causes other problems, such as less favorable torsion angles in the side chain or less optimal packing. The third step is to make the designed proteins and analyze their properties. If the experimental result is not close to what is expected, a next round of the design cycle is started.

A typical goal of protein engineering is to make a structure or product more stable, since the performance of protein pharmaceuticals and proteins used as biocatalysts can often be improved by stabilization (*58-60*). There are many rational strategies to increase protein stability (*39, 40, 61-63*). The role of hydrophobic effect has received much attention since the hydrophobic effect is a dominant factor that defines protein stability. Thus, redesigning the hydrophobic core is an obvious strategy to increase the protein stability because of improved packing of the hydrophobic core (*41, 64-68*). Another strategy is to introduce disulfides (*69-72*), Jeong and co-workers (*71*) designed a disulfide bond by using geometry based on a design program and an amino acid sequence analysis. The newly designed disulfide bridge increased the thermostability of the protein about 5 ^{o}C. People also attempt to increase the protein stability by optimizing of the surface electrostatic interactions (*38-41, 73, 74*). Spector and co-workers have shown that a charged reverse mutation in a small peripheral subunit-binding domain leads to a 1.1 kcal mol^{-1} increase in the stability (*39*). Strickler and co-workers computationally redesigned the sequence of five different proteins to optimize the surface charge-charge interactions. They also experimentally measured their stability by CD and DSC: the stability of all the

designed proteins was significantly higher than the wild type proteins (*40*) . Thus the optimization of charge–charge interactions on the surface of proteins can be a useful strategy in the design of thermostable proteins. Design of charge–charge interactions has many important advantages. First, mutations at well-exposed positions in the native structure are less likely to be disruptive and significantly affect the structure and/or the function of the protein. Second, and most importantly, it appears that very simple electrostatic models can be used to decide which surface residues are to be mutated.

1.5. The chicken villin headpiece helical subdomain (HP36)

Villin headpiece which is denoted HP76 is 76 residues and is the C terminal domain of villin (*75, 76*). It retains full F-actin-binding activity in isolation (*77-79*). The protein villin plays an important role in the formation of microvilli of the gut and kidney (*80*). The first 9 residues in the N-terminus can be removed without affecting the function, stability and the structure of the protein (*81*). This protein is designated HP67. HP67 is the smallest biologically active unit of villin (*81, 82*). The numbering system used is that of the full length 76-residue villin headpiece, e.g. the first residue in HP67 is denoted as residue 10. The helical C-terminal subdomain of the villin headpiece is denoted HP36 which consists of residues 42-76 of the villin headpiece with an additional N-terminal Methionine from expression in E.*coli*. This Met is designated as number 41. In this thesis the C terminal subdomain of the villin headpiece (HP36) is used as an experimental model system.



N-terminus

41 51 61 71

MLSDE DFKAV FGMTR SAFAN LPLWK QQNLK KEKGLF

Figure 1.2: Ribbon diagram and primary sequence of the villin headpiece subdomain HP36. The N-terminus is labeled. The side chains of three phenylalanine residues (F47, F51, and F58) are shown in stick representation, created by PyMol, version 0.99

(*83*). The pdb code is 1VII (*84*). In the sequence, color green represents the loop regions and color orange represents helical regions.

HP36 folds very fast and it is one of the smallest peptides that folds cooperatively without disulfide bonds or ligand binding (*85*). The N-terminal 31 residues of HP67 cannot fold in isolation, but HP36 can fold independently. HP36 is composed of three helices (residues Asp44-Lys48, Arg55-Phe58 and Leu63-Lys70) which are packed together to share a hydrophobic core mainly contributed by the phenylalanine and leucine residues (*21*) (Figure 1.2). The three helices are connected by a loop (residues Ala49-Thr54) and a turn (residues Ala59 to Pro62) (*84*). The equilibrium folding transitions of HP36 have been shown to be two-state (*85*). It has also been demonstrated that HP36 folds on the microsecond timescale (*86, 87*). Its small size and rapid folding have made it an exceptionally popular model protein for theoretical and computational studies of protein folding and dynamics (*85-91*).

1.6. Aim of the thesis

This thesis describes the folding and stability of the chicken villin helical subdomain (HP36) and seventeen mutants. The expression and purification of the wild type HP36 using fusion technology is described in Chapter 2. In order to study the contribution of charged residues to the stability of HP36, twelve point mutants were studied and compared with WT HP36. Each of the acidic residues was individually mutated to the corresponding amide while the Lys and Arg residues were individually mutated to Met. This work is described in Chapter 3. A hyperstable double mutant of HP36 was designed and characterized and those studies are descried in Chapter 4. Finally, the interaction between Phenylalanine 47 and Arginine 55 as well as the interaction between Phenylalanine 47 and Aspartic acid 44 were investigated using double mutant cycle analysis and are discussed in Chapter 5.

2. Efficient high level expression of peptides and proteins as fusion proteins with the N-terminal domain of L9: Application to the Villin Headpiece Helical Subdomain

Abstract

The efficient expression of small to midsize polypeptides and small marginally stable proteins can be difficult. A new protein fusion system is developed to allow the expression of peptides and small proteins. The polypeptide of interest is linked via a Factor Xa cleavage sequence to the C-terminus of the N-terminal domain of the ribosomal protein L9 (NTL9). NTL9 is a small, (56 residue) basic protein. The C-terminus of the protein is part of an α-helix which extends away from the globular structure thus additional domains can be fused without altering the fold of NTL9. NTL9 expresses at high levels, is extremely soluble, and remains fully folded over a wide temperature and pH range. The protein has a high net positive charge, facilitating purification of fusion proteins by ion exchange chromatography. NTL9 fusions can also be easily purified by reverse phase HPLC. As a test case we demonstrate the high level expression of a small, 36 residue, three helix bundle, the villin headpiece subdomain. This protein is widely used as a model system for folding studies and the development of a simple expression system should facilitate experimental studies of the subdomain. The yield of purified fusion protein is 70 mg per liter of culture and the yield of purified villin headpiece subdomain is 24 mg per liter of culture. We also demonstrate the use of the fusion system to express a smaller marginally folded peptide fragment of the villin headpiece domain.

Acknowledgements

The material presented in this chapter has been published (Yuan Bi, Yuefeng Tang, Daniel P. Raleigh and Jae-Hyun Cho 2006 *Protein Expr Purif 47*, 234-40). This chapter contains direct excerpts from the manuscript written by Yuan Bi with suggestions and revisions from Jae-Hyun Cho and Professor Daniel P. Raleigh. Dr. Yuefeng Tang contributed to the work of the fragment of the villin headpiece helical subdomain (HP21).

2.1. Introduction

Expression of small to midsize peptides or marginally stable proteins can often be problematic. The ability to produce large amounts of high purity material efficiently and cheaply is critical for structural and biophysical studies, especially if isotopic labeling is desired. The cost-effective production of small peptides is also of interest due to the increasing use of peptides as pharmaceutical agents. Some, but

certainly not all small peptides can be synthesized chemically, however large peptides still present challenges for peptide synthesis. In addition, the incorporation of isotopic labels by peptide synthesis is prohibitively expensive. Fusion technology is the preferred method to enhance protein expression and to simplify protein purification (*92*). Here, we introduce a new fusion protein system based upon a small highly soluble basic protein, the N-terminal domain of L9 (NTL9). This 56 residue protein has been shown to be expressed at high yield (50 mg per liter or higher) (*93*), is very soluble, highly stable and is easy to purify via ion exchange chromatograph (IEC) or reverse phase high performance liquid chromatography (HPLC). We demonstrate that small marginally stable proteins and partially structured peptides can be expressed at a high final yield as NTL9 fusions using *Escherichia coli* (*E.coli*.) based systems. As a test case we examine the expression of the villin headpiece helical subdomain (HP36) (Figure 2.1A). This 36 residue domain is the smallest naturally occurring sequence that has been shown to fold cooperatively (*85*). As such it has become a favorite model system for studies of protein folding. A large number of computational and theoretical studies have been reported but a much smaller number of experimental investigations have been conducted (*87, 91, 94-97*). This may reflect the fact that the domain expresses at relatively low levels using standard systems. In fact, published studies of its folding have made use of chemically synthesized material (*86, 87, 98*).

We have also been interested in studying fragments of this domain for use as models of the unfolded state (*88*). Small partially structured peptides are difficult to express, although they sometimes can be prepared as fusion proteins or by systems which target them to inclusion bodies. Targeting to inclusion bodies can protect against degradation but can lead to problems in solubilization and purification. One often wishes to prepare ^{15}N, ^{13}C labeled material for NMR studies. Expression of a small peptide fused to a large globular protein means that a significant fraction of the expressive isotopic label will be incorporated into the uninteresting fusion partner. One advantage of NTL9 is that its small size means that a significantly smaller fraction of the total label is incorporated into the fusion partner. Recently, there is growing interest in natively unstructured proteins (*99-104*). This is not only due to their importance in regulating biological function (*104*) but also due to increasing interest in model systems for the unfolded state of protein (*99, 101, 102, 104*). Expression and purification of unstructured proteins or peptides, however, is not trivial. Because of the lack of well defined structure, their solubility is generally poor and it is easy to form inclusion bodies during expression or aggregates during purification. We show that expression of peptides as NTL9 fusions can provide a solution to these problems. As a test case, we use a 21 residue fragment of HP36. This fragment designated HP21 (Figure 2.1A) adopts some structure in isolation but is not well folded (*88*).

NTL9 is a small α-β protein whose folding has been under active investigation in our group (*105*). The protein is very thermal stable with a melting midpoint of 77.5 ^{o}C at pH 5.45 (*106*). It remains fully folded over an exceptionally wide pH range (pH 1 to at least pH 11) and is highly soluble. NTL9 is an RNA binding domain and is rich in Lysine and Arginine residues. The domain contains 12 Lysine and Arginine

residues and has a calculated net charge of +6.0 at pH 7.0. The basic nature of the protein makes purification by IEC very simple. In addition NTL9 can be easily purified by HPLC (*105*).

2.2. Materials and methods

2.2.1. Materials.

The primers for PCR amplification were from Operon. *E. coli* BL21 (DE3) cells were purchased from Promega. Factor Xa was purchased from Novagen. Enzymes for recombinant DNA technology such as cloned pfu polymerase, dNTPs and Dpn I were purchased from Stratagene. Quick T4 DNA ligase was purchased from New England Biolabs, Inc. Kits for purification of PCR fragments, plasmid extraction, and extraction of restriction fragments separated on agarose gels were purchased from Qiagen. All other chemicals and biochemicals were purchased from Sigma–Aldrich, Fisher Scientific and Shelton Scientific.

2.2.2. Design and construction of the pET3a-NTL9-FXa-HP36 expression vector.

Forward and reverse oligomers were designed to introduce the sequences containing Factor Xa and HP36 at the 3' end of the NTL9 sequence in pET3a. This vector was initially designed for preparing recombinant NTL9 for folding studies (*93*). Six DNA oligomers were designed and 13 to 17 codons were introduced by each PCR amplification. The sequences of the oligonucleotides used are shown in Table 2.1. The underscored sequences indicate codons that were introduced by PCR. PCR was performed under the following conditions: 5 min at 95^{o}C for full denaturation, 30 s at 95 ^{o}C, 1 min at 50 ^{o}C, 10 min at 72 ^{o}C for 18 cycles for amplification, and 10 min at 72 ^{o}C for an additional extension. The final plasmid consists of NTL9-FXa-HP36 cloned into pET3a vector. The stop codon was inserted at the position of residue 22 by PCR to construct the expression vector for HP21 (pET3a-NTL9-FXa-HP21), the truncated version of HP36. The primer sequence for adding the stop codon (the underscored sequences) is 5' GCC TTT GCC AAC TTG <u>TAA</u> TTG TGG AAA CAG CAG 3' and the corresponding reverse sequence.

2.2.3. Overexpression and purification of fusion protein.

The pET3a-NTL9-FXa-HP36 plasmid was transformed and overexpressed in E.coli BL21 (DE3). The transformed cells were grown at 37 °C in a shaking incubator using Luria-Bertani media containing 50 mg/ml of ampicillin. Overexpression of fusion protein was induced by IPTG (1mM) for four hours at 37 C when the OD_{600} reached 1.5. Cells were harvested by centrifugation at 5000 **g** at 4 °C for 10 min. The pellet was resuspended in 20 mM Tris–HCl (pH 7.5) and lysed using sonication. The insoluable fraction was removed by centrifugation at 12,000 **g** at 4 °C for 30 min.

The clear supernatant was collected and loaded onto an ion exchange column (S-Sepharose column, Pharmacia) equilibrated with 20 mM Tris–HCl (pH 7.5). The column was then washed for 30 min with the same buffer. The fusion protein was eluted with a 0–2 M NaCl gradient and purified further by HPLC with a Vydac C8 preparative column. The fusion protein was eluted with linear A-B gradient where the

buffer A was 0.1% (v/v) TFA in H_2O, and the buffer B was 90% acetonitrile, 9.9% H_2O and 0.1% (v/v) TFA. The gradient used was 0% to 100% buffer B in 100 min after 10 min of 100% buffer A. For purification of the fusion protein of HP21, EDTA and PMSF were added to final concentration of 1mM to prevent degradation by proteases during IEC.

2.2.4. Cleavage of fusion protein and purification of HP36.

Cleavage of HP36 from the NTL9-FXa-HP36 fusion protein was carried out by incubating fusion protein with Factor Xa in buffer (50 mM Tris–HCl, 100 mM NaCl, and 5 mM $CaCl_2$, pH 8.0). One milligram of the fusion protein was treated with 2 unit of Factor Xa for 16 h at ~23 °C. The solution was purified by HPLC using a C8 preparative column. HP36 was eluted with linear A-B gradient where the buffer A was 0.1% (v/v) TFA in H_2O, and the buffer B was 90% acetonitrile, 9.9% water and 0.1% (v/v) TFA. The gradient used was 0% to 100% buffer B in 100 min after 10 min 100% buffer A. The HP36 fractions were collected, lyophilized and kept at -20 °C. Cleavage of HP21 from the NTL9-FXa-HP21 fusion protein was carried out by incubating fusion protein with Factor Xa in buffer (50 mM Tris–HCl, 100 mM NaCl, and 5 mM $CaCl_2$, pH 8.0). One milligram of the fusion protein was treated with 0.5 unit of Factor Xa for 16 h at 4 °C.

2.2.5. Characterization of protein products.

The identity of fusion proteins, HP36 and HP21 were confirmed by ESI mass spectrometry (ESI-MS) using a Platform LCZ instrument in the Stony Brook Proteomics center. The structure of the proteins was confirmed by 1D ^{1}H-NMR and Far-UV CD. NMR spectra were recorded on Varian Instruments Inova 500 MHz spectrometer. The samples for NMR experiments were prepared in 10% D_2O, 90% 10 mM sodium acetate and 150 mM sodium chloride, 1 mM HP36 at pH5.0 and 400 uM HP21 at pH 5.5. All CD experiments were performed using Aviv 62A DS and 202SF CD spectrophotometers. All samples for CD experiments were prepared in 10 mM Sodium acetate and 150 mM Sodium chloride buffer at pH 5.0. Both NMR and CD experiments were performed at 25 ^{o}C.

2.3. Results

2.3.1. Overexpression and purification of fusion protein.

We constructed the fusion protein system by cloning the gene for HP36 at the C-terminus of the pET3a-NTL9 vector. Initially, we had tried to express HP36 alone using the pVHP42-76b plasmid in E.coli BL21(DE3) (*85*). This system, however, produced relatively low levels of HP36. Our maximum observed yield was 7.5 mg per liter for wildtype. Furthermore, some mutations of the protein reduced the expression level so much that we could not detect any expressed protein in the total lysate of the induced cells (data not shown). Similar problems are well known when small proteins or peptides are expressed in bacterial systems. This is often due to the marginal stability/or unstructured confirmation of small proteins.

Use of the fusion expression system, pET3a-NTL9-FXa-HP36 (Figure 2.1B),

leads to the successful overexpression of the fusion protein (Figure 2.2, lane 1). The expression conditions for fusion protein followed those for expression of NTL9 (*93*). Most of the fusion protein was present in the cytosolic fraction (Figure 2.2, lane 2).

The fusion protein was purified by IEC (Figure 2.3A). Fractions 19-23 were pooled and the purity of the purified fusion protein was checked by SDS-PAGE (Figure 2.2, lane 3) and HPLC (Figure 2.3B). The fusion protein eluted at 47% buffer B in the HPLC and the peak was well resolved from other peaks. The calculated peak area for the fusion protein represents 93% of the total peak area. The simple high efficient purification is likely due to the unique characteristics of the NTL9 tag. NTL9 is a highly charged basic protein since the pI of NTL9 is around 10 (*107*). The fusion provides an excellent way to purify target proteins by IEC. The highly charged nature of NTL9 not only aids the purification but also helps keep the target protein soluble. The molecular weight of the purified fusion protein was confirmed by ESI-MS. The measured molecular weight is 10846.46±0.79 Da (Figure 2.4A) and the expected molecular weight is 10846.70 Da. The yield of the fusion protein was around 70 mg from one liter of culture. The purified fusion protein was lyophilized and stored at -20 °C .We also used the fusion to express the K70M mutant of HP36 at the yields that were similar to wildtype. All attempt to express this mutant using the pVHP42-76b plasmid failed.

2.3.2. Purification of HP36 and HP21.

Factor Xa cleavage of NTL9-FXa-HP36 fusion protein was performed as described in the methods section. The cleavage was clean, resulting in two products, NTL9-FXa and HP36. The efficiency of the cleavage reaction was confirmed by HPLC (Figure 2.3C). Both peaks were collected separately and identified by ESI-MS. The second peak (49% buffer B) was pure HP36 (Figure 2.2, lane 5). The molecular weight of the purified HP36 was determined to be 4190.13±0.06 Da by ESI-MS (Figure 2.4B), and the expected molecular weight is 4190.92 Da. The final purification yield for HP36 is around 24 mg from one liter of culture. This compares very well to the expected yield based on the relative masses of HP36 and NTL9-FXa-HP36 (i. e. 27 mg per liter).

The yield of purified NTL9-FXa-HP21 was 60mg per liter. The identity of the product was confirmed by ESI-MS, observed 9006.99±0.13 Da and expected 9006.38 Da. The cleavage reaction for HP21 was performed at lower temperature (4 °C) because of nonspecific cleavage by Factor Xa at 23 °C. At 23°C Factor Xa cleavaged site between R55 and S56 in addition to cleavage at the desired site. The nonspecific cleavage could be avoided by lowering the incubation temperature. The identity of HP21 was confirmed by ESI-MS, observed 2351.10±0.13 Da and expected 2350.60 Da.

2.3.3. Characterization of expressed proteins.

Structural characterization of HP36 was performed by far-UV CD and 1D ^{1}H-NMR measurements. HP36 has a compact hydrophobic core which is made up of three alpha helices (*84*). The far-UV CD spectrum (Figure 2.5A) shows a double

minimum at 208nm and 222nm which is a characteristic feature of helical proteins. The 1D ^{1}H-NMR spectra for HP36 (Figure 2.6A, 2.6B) showed several characteristic resonances, that indicate a well folded structure. In particular, the peaks at -0.10 ppm (Val-50) and 5.50 ppm (Phe-47) are due to packing in the hydrophobic core. The far-UV CD and NMR spectra are identical to spectra of a known sample of wildtype (*84, 86*).

The thermal unfolding of HP36, as monitored by CD, is fully reversible and was well fit by a two-state (un)folding model (Figure 2.5B). The transition midpoint for thermal unfolding, T_m, was 73 ^{o}C and the enthalpy change of unfolding (ΔH^{o}) at T_m was estimated to be 31.8 kcal mol^{-1}. These numbers are very closed to those reported previously (*85*) indicating that expressed HP36 behaves normally. Chemical denaturation using guanidine chloride (GdmCl) also showed two-state (un)folding behavior (Figure 2.5C). The estimated free energy for unfolding (ΔG^{o}_{U}) was 2.93 kcal mol^{-1} which is very close to the reported value (*85*).

Far-UV CD spectra of HP21 showed a double minimum at 208nm and 222nm, however, the signal intensity of HP21 is less than that of HP36 (Figure 2.5A). The CD spectrum of recombinant HP21 is similar to chemically synthesized material (*88*). The previous characterized chemical synthesized sample has an amidated C-terminus but this was only a small effect on secondary structure. The 1D ^{1}H-NMR spectrum of HP21 (Figure 2.6C, 2.6D) is also very similar to that of the chemically synthesized sample. This data show that the recombinant peptide behaves similarly to the synthesized material.

2.4. Discussion

We have showed that the new fusion-tag, NTL9 works extremely well for expressing and purifying small marginally stable proteins and partially structured peptides. The development of a sample expression system for HP36 should facilitate work on this interesting protein. There are several considerations for designing an ideal fusion-tag system. Firstly, the fusion protein alone should have a high expression level. Secondly, the fusion protein should not affect the structure of target protein. Thirdly, it would be ideal if the fusion tag can provide an easy purification method. The NTL9 fusion system satisfies all three requirements. NTL9 has a long C-terminal alpha helix which prevents non-desired interactions between the target protein and the NTL9 tag. NTL9 is stable and has simple two-state (un)folding behavior in a wide range of experimental conditions. The highly charged NTL9 provides additional advantages in expressing and purifying unstructured protein and/or peptides. The fusion system helps keep the target protein soluble during expression and purification. For example, even though the solubility of HP21 alone is low (the maximum solubility is around 400 uM), our fusion system could express NTL9-FXa-HP21 at more than 60mg from one liter of culture.

Table 2.1: The sequence of primers used to introduce the Factor Xa cleavage site and the sequence of HP36 into the 3' end of the NTL9 sequence in pET3a.

PCR-1	Forward 1	5'AGA GAG CAT TCG ACC TTC AAT GCG CTG CTC TTT TTG TTT TTG CGC CTC3'
	Reverse 1	5' AAG AAG GAG AAA GGA CTC TTC TAG GGA TCC GGC TGC TAA CAA AGC CCG 3'
PCR-2	Forward 2	5'AAC AGC CTT GAA GTC CTC GTC AGA GAG CAT TCG ACC TTC AAT GCG CTG3'
	Reverse 2	5' TGG AAA CAG CAG AAC CTC AAG AAG GAG AAA GGA CTC TTC TAG GGA TCC GGC3'
PCR-3	Forward 3	5' GGT CAT GCC AAA AAC AGC CTT GAA GTC CTC GTC AGA GAG CAT TCG ACC TTC AAT 3'
	Reverse 3	5' CGC TCT GCC TTT GCC AAC TTG CCC TTG TGG AAA CAG CAG AAC CTC AAG AAG GAG AAA GGA3'

A

```
          41              51              61              71
HP36:  MLSDEDFKAV  FGMTRSAFAN  LPLWKQQNLK  KEKGLF
HP21:  MLSDEDFKAV  FGMTRSAFAN  L
```

B

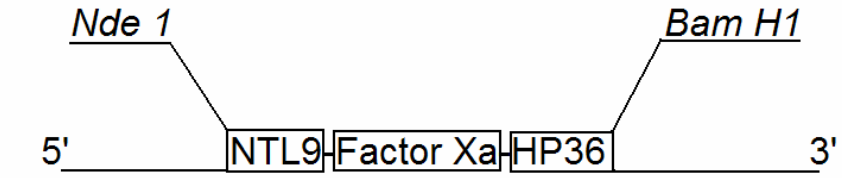


Figure 2.1: (A) Primary sequence of HP36 and HP21. (B) Schematic representation of the expression vector for NTL9-FXa-HP36 fusion system.

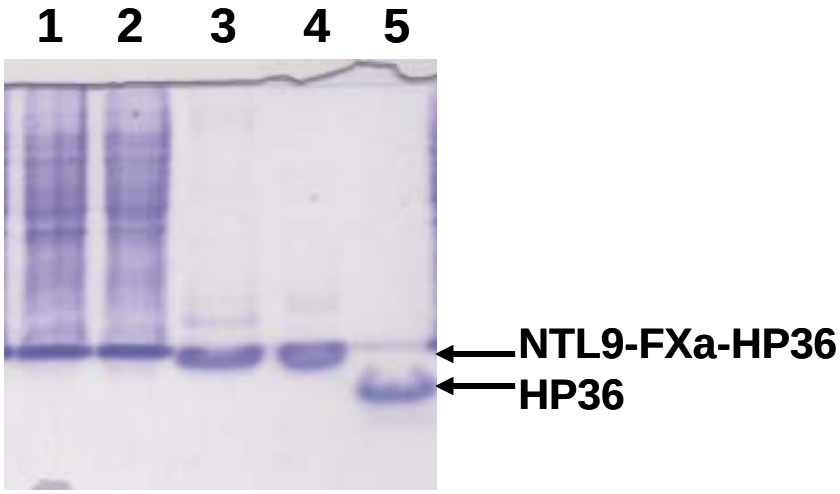


Figure 2.2: SDS-PAGE analysis (15%) of purification of the NTL9-Fxa-HP36 fusion protein and HP36. Lane 1, total lysate of the induced cells; lane 2 , soluble fraction (total lysate after centrifugation); lane 3, factions 19-23 from ion exchange chromatography; lane 4, purified NTL9-FXa-HP36 fusion protein; lane 5, purified HP36.

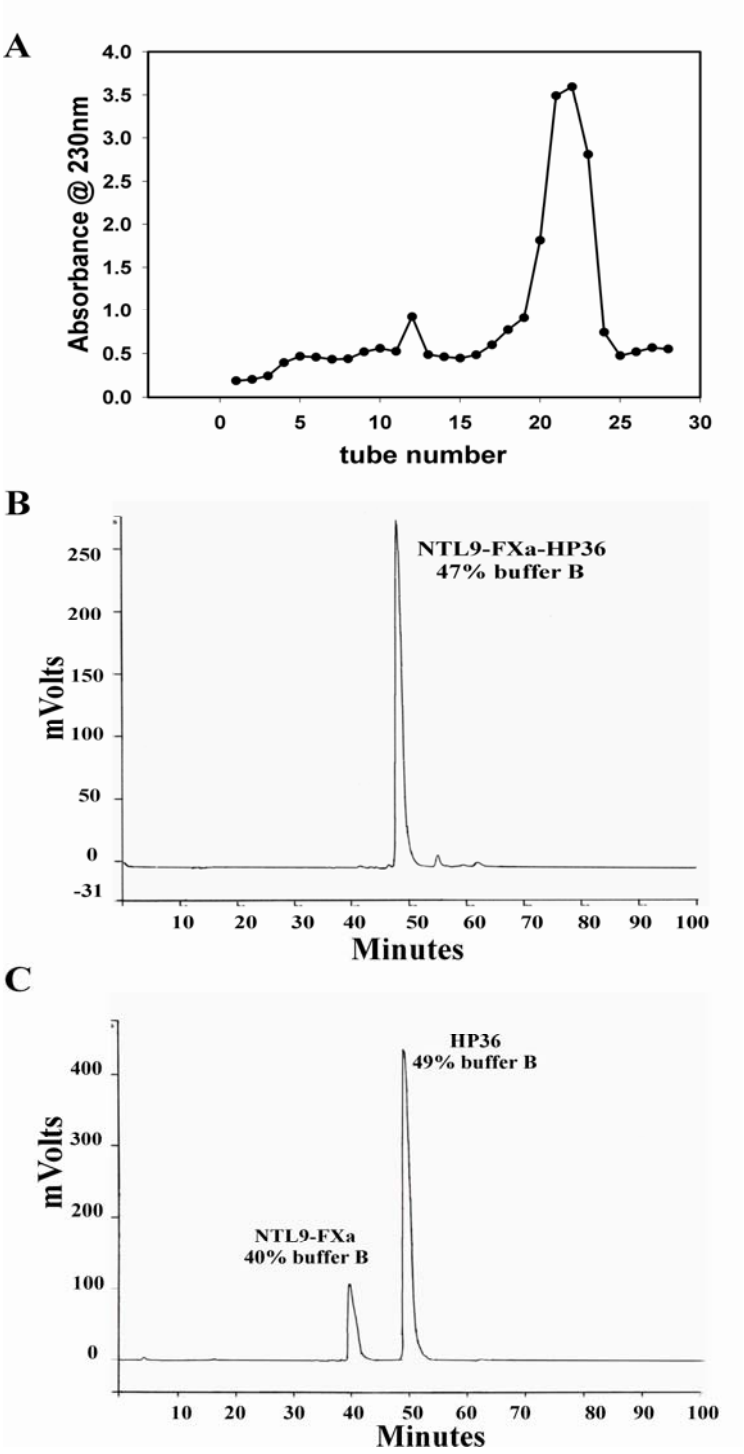


Figure 2.3: (A) Purification of the fusion protein by ion exchange chromatogram. The major peak (fractions 19-23) contains the NTL9-FXa-HP36 fusion protein. (B) HPLC trace of the fusion protein from the pooled fractions from ion exchange chromatogram. (C) HPLC trace of HP36 after cleavage from the fusion protein. The peak at 40% buffer B corresponds to NTL9-FXa. The peak at 49% buffer B corresponds to HP36. HPLC traces display measured absorbance at 280 nm versus time. The molar absorption coefficient is 1290 $M^{-1}cm^{-1}$ for NTL9 (one Tyr) and 5690 $M^{-1}cm^{-1}$ for HP36 (one Trp) *(122)*.

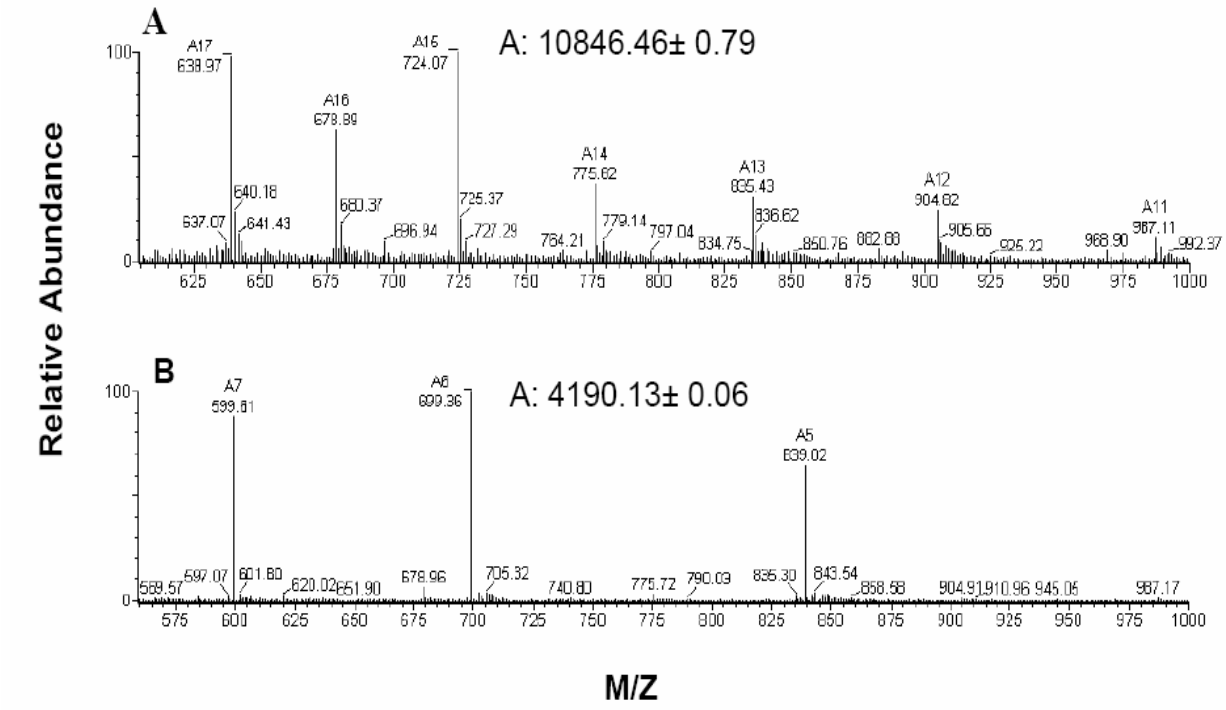


Figure 2.4: (A) ESI mass spectrum of NTL9-FXa-HP36 fusion protein. (B) ESI mass spectrum of HP36.

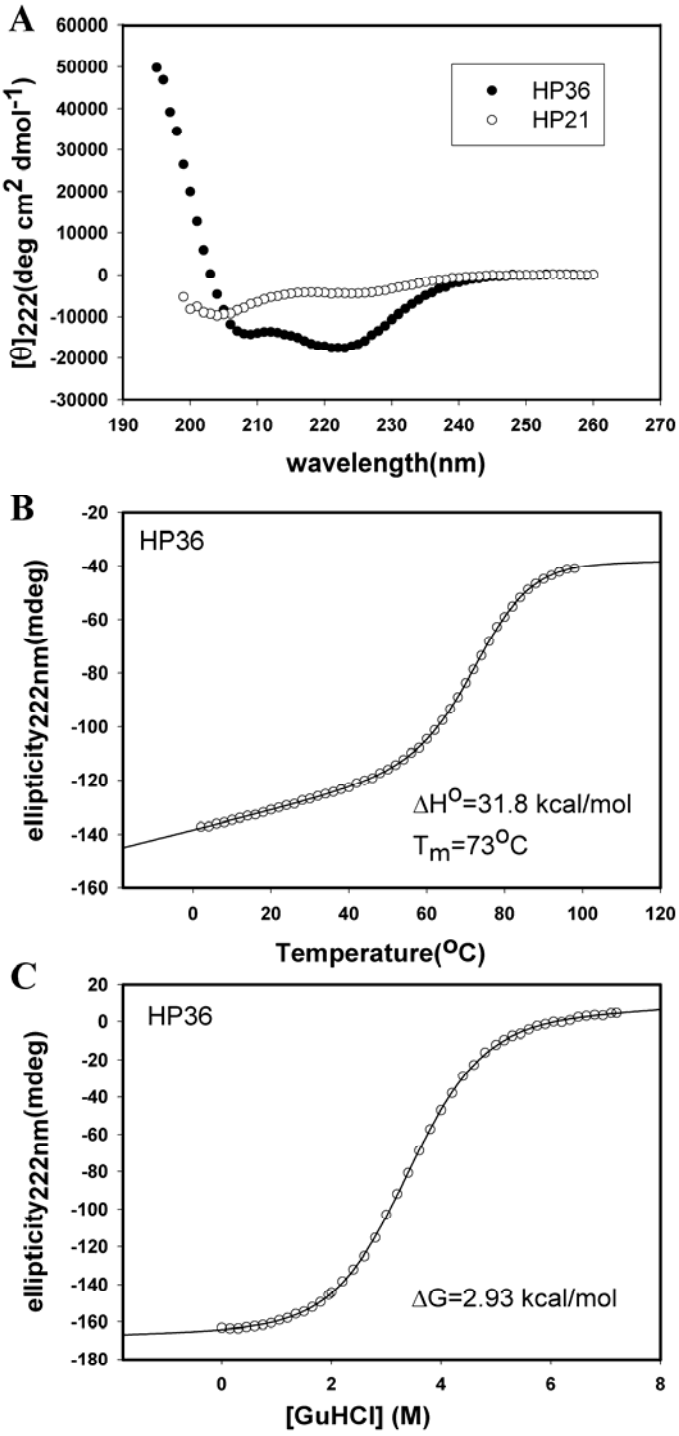


Figure 2.5: (A) Far-UV CD spectra of HP36 (closed circles) and HP21 (open circles). (B) Thermal unfolding curve of HP36. (C) Guanidine hydrochloride denaturation curve of HP36.

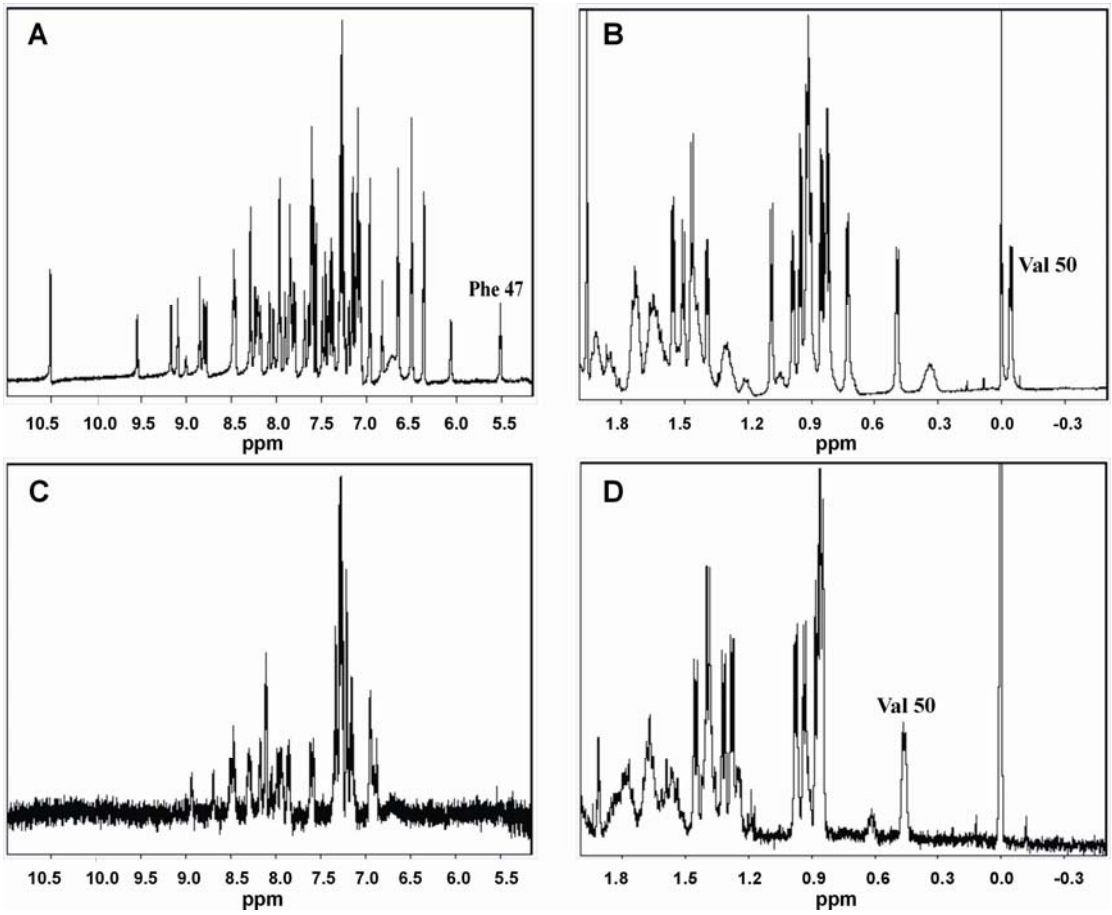


Figure 2.6: 1D ^{1}H-NMR spectrum of HP36 (A) Downfield region (5.0 ppm-11.0 ppm) (B) Upfield region (-0.5 ppm-2.0 ppm). 1D ^{1}H-NMR spectrum of HP21 (C) Downfield region (5.0 ppm-11.0 ppm) (D) Upfield region (-0.5 ppm-2.0 ppm).

3. The contribution of charged residues to the stability of the villin headpiece subdomain

Abstract

The helical subdomain of the villin headpiece (HP36) is a small three helix protein. HP36 is one of the smallest naturally occurring sequences which folds cooperatively. Its small size coupled with its very rapid folding have made it an extraordinarily popular model system for experimental and computational studies of protein folding. Unfortunately the modest stability of the subdomain causes technical difficulties with some experimental studies and variants with increased stability are desired. Here we analyze the contribution of charged residues to the stability of HP36 by conducting a systematic mutational study of twelve point mutants. Each of the acidic residues was individually mutated to the corresponding amide while the Lys and Arg residues were individually mutated to Met. Mutation of any of the acidic residues leads to a decrease in T_m and ΔG^o of unfolding. In contrast three of the seven mutations which targeted basic residues lead to significant increase in stability. The correlation between changes in T_m and the corresponding change in ΔG^o are well predicted by experimental relations derived from a database of large proteins. Potential unfolded state effects were probed by examining the linkage relationship between the change in $\Delta\Delta G^o$ in pH.

Acknowledgements

The material presented in this chapter contains direct excerpts from a manuscript written by Yuan Bi with suggestions and revisions from Professor Daniel P. Raleigh. Mr. Bing Shan contributed to the pKa measurements.

3.1. Introduction

It is very important to understand the factors that affect protein stability, folding and aggregation both because of the importance of aggregation in human diseases and because of practical problems associated with marginal stability or inefficient folding (*8, 9, 108, 109*). Enhanced protein stability can also be critical for biophysical studies of marginally stable proteins.

The villin headpiece helical subdomain (HP36) is one of the smallest known natural sequences that folds cooperatively to a compact globular structure. HP36 is one of the most widely studied systems for experimental (*21, 84-89, 110*), theoretical and computational (*90, 91, 94, 111-113*) investigations of protein folding. More than a dozen groups have presented computational or theoretical studies of its folding. There have been fewer experimental investigations of its folding, although our laboratory and the Eaton's group have independently reported that it folds on the microsecond time scale (*86, 87*). One practical difficulty with studies of small proteins is that while they may have a high Tm, they are typically marginally stable in terms of ΔG^o of unfolding (*114*). This can limit

experimental work, particularly mutational studies since even modestly destabilizing mutations may prevent the proteins from folding. HP36 is no exception. Although the domain has a high T_m, 73 ^oC, the free energy of unfolding at 25 ^oC is only 3.22 kcal mol^{-1}. An additional motivation for seeking stabilized variants of HP36 is that they could, in principle, allow one to test the prediction that certain proteins can be induced to fold in a downhill fashion. Downhill folding refers to folding in the absence of a free energy barrier. It is has been suggested that proteins which fold near the so called speed limit for folding, such as HP36, might be converted to downhill folders by stabilizing them (*115*).

A variety of approaches have been used in attempts to enhance protein stability. These include improving the packing of the hydrophobic core, the introduction of disulfides, new salt bridges, hydrogen bonding interactions, or the optimization of surface electrostatic interactions (*32, 39, 40, 58, 61-63, 73, 116-119*). Modification of surface charge-charge interactions is an attractive approach for several reasons. First, mutations at well-exposed positions in the native structure are less likely to be disruptive, or to affect the structure, or the function of the protein. Second, it appears that very simple electrostatic models can, in some cases, be used to decide which surface residues are to be mutated (*32, 39*). Here we describe a systematic analysis of HP36 in which every charged residue is individually mutated to a neutral residue and demonstrate that single points mutants can significantly stabilize the domain.

The helical subdomain of headpiece is located at the C terminus of the actin-binding protein villin. It contains an F-actin-binding site and endows villin with microfilament bundling activity (*76*). Figure 3.1A shows a ribbon diagram and the sequence of HP36. Our construct corresponds to the last 35 residues of the full length 76-residues villin headpiece with an additional N-terminal Methionine. This protein is denoted as HP36 and the sequence numbering is such that N-terminal Methionine is denoted residue 41. This notation is used to be consistent with previous studies. HP36 is made up of three short α-helices and the subdomain contains a tightly packed hydrophobic core mainly contributed by the phenylalanine and leucine residues. The three helices (residues Asp44-Lys48, Arg55-Phe58 and Leu63-Lys70) are packed together as shown in Figure 3.1A. They are connected by a loop (residues Ala49-Thr54) and a turn (residues Ala59 to Pro62) (*84*). HP36 has ten charged residues on its surface (Figure 3.1B): four acidic residues D44, E45, D46, E72 and six basic residues K48, R55, K65, K70, K71, K73, thus it is a good target for studying the role of charged surface residues on protein stability.

3.2. Materials and methods

3.2.1. Materials

The primers for generation of the mutants D44N, E45Q, D46N, K48M, R55M, K65M, K70M, K71M, E72Q and K73M were obtained from Operon. *E. coli* BL21 (DE3) cells were purchased from Promega. Factor Xa was purchased from Novagen. Enzymes for recombinant DNA technology such as cloned pfu polymerase, dNTPs and *Dpn*I were purchased from Stratagene. Kits for the purification of plasmid were purchased from Qiagen. All other chemicals and biochemicals were purchased from Sigma–Aldrich, Fisher Scientific and Shelton Scientific.

3.2.2. Cloning, expression, and purification

The plasmid (pET3a-NTL9-FXa-HP36) containing the gene for HP36 was prepared as described (*120*). The fusion proteins with coded amino acid substitutions were expressed and purified 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 labeled glucose. The desired proteins were cleaved from fusion protein and purified by RP-HPLC. All proteins were more than 95% pure as judged by HPLC. The identities of the proteins were analyzed by electro-spray ionization mass spectrometry (ESI MS) or matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI-TOF). The expected and observed molecular weights were as follows: D44N, expected 4188.9, observed 4189.0; E45Q, expected 4188.9, observed 4188.8; D46N, expected 4188.9, observed 4190.1; K48M, expected 4192.9, observed 4193.9; R55K, expected 4161.8, observed 4163.4; R55M, expected 4164.9, observed 4166.4; K65M, expected 4190.1, observed 4192.3; K70M, expected 4190.1, observed 4193.3; K70E, expected 4190.8, observed 4190.4 K71M, expected 4190.1, observed 4194.3; E72Q, expected 4188.9, observed 4187.8; K73M, expected 4190.1, observed 4194.4; ^{13}C, ^{15}N-labeled HP36: expected 4432.9, observed 4428.9.

3.2.3. Analytical Ultracentrifugation

Analytical ultracentrifugation was performed to examine whether the stabilizing mutants were monomeric. Molecular masses were determined with initial protein concentrations of 30, 60, 100uM in 10mM sodium acetate and 150mM sodium chloride buffer at pH 5.0. Equilibrium experiments were performed at 25ºC with a Beckman Optima XL-A analytical ultracentrifuge using rotor speeds of 38,000 rpm (24hours) and 48,000 rpm (24hours). Six-channel, 12 mm path length, charcoal-filled Epon cell with quartz windows were used. The wavelength used for this experiment was 280nm. Ten scans were averaged. The partial specific volume (0.7467 ml g^{-1}) and solution density (1.005 g l^{-1}) were calculated from the software program SEDNTERP (*121*). The HeteroAnalysis program from the Analytical Ultracentrifugation Facility at the University of Connecticut was used for data analysis.

3.2.4. Circular Dichroism Spectroscopy

All Circular Dichroism (CD) experiments were carried out on an Aviv 62A DS and a 202SF Circular Dichroism spectrophotometer. All samples for CD experiments were prepared in 10 mM sodium acetate, 150 mM sodium chloride buffer solution at pH 5.0. The protein concentrations were determined from absorbance measurements using the method of Pace & coworkers (*122*). The absorbance was measured at 280 nm in 6M guanidine hydrochloride, 20 mM sodium phosphate pH 6.5, an extinction coefficient of 5690 $M^{-1}cm^{-1}$ was used. Far-UV wavelength scans were performed with five repeats and a signal averaging time of 1 second per point, in a 1 mm quartz cuvette, over the range of 195 nm to 260 nm. Thermal unfolding and urea denaturation experiments were carried out in a 1 cm quartz cuvette by monitoring the signal at 222 nm. Concentrations of urea were measured by measuring the refractive index. The concentration of urea was increased from 0 M to 10 M by ~0.25 M each

step. Wavelength scans and urea denaturation experiments were performed at 25 ^{o}C. Thermal unfolding experiments were performed from 2 ^{o}C to 98 ^{o}C with a 2 ^{o}C interval. For pH dependent stability measurements, urea denaturations were used at pH3.0 to pH 6.0 for wildtype HP36 and HP36 K48M mutant. Thermal denaturations were used at pH 2.0 to pH 5.0 for wildtype HP36. All of the denaturation data were analyzed by nonlinear least squares curve fitting using SigmaPlot, as described (*123-125*).

3.2.5. NMR Spectroscopy

All NMR experiments were carried out on a Varian Instruments Inova 500Mhz nuclear magnetic resonance spectrometer at 25 ^{o}C. Samples of 1mM proteins for 1D NMR were prepared in 10% D_2O, 90% 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0. TSP was added as a chemical shift standard.

3.2.6. Determination of pKa values using NMR

The ^{13}C, ^{15}N-labeled HP36 was dissolved in 90% H_2O/10% D_2O containing 10 mM sodium acetate and 150 mM sodium chloride at a concentration of 1mM. The NMR experiments for the measurement of the chemical shifts of the side-chain carboxyl carbons were performed at 25 ^{0}C on a 600 MHz Bruker spectrometer with a cryoprobe at the New York Structural Biology Center.

Side-chain protons were assigned at pH 5.0 using the HCCCONH and the HCCH-TOCSY experiments. Side-chain carboxyl carbons were assigned using the correlations of the $H^{\beta}C^{\beta}C^{\gamma}$ (Asp) and $H^{\gamma}C^{\gamma}C^{\delta}$ (Glu) resonances in the HCACO experiment (*126, 127*). A 2D version of the 3D HCACO experiment was carried out at pH 5.0 to measure the chemical shifts of the side-chain carboxyl carbons by following the Asp (H^{β}-C^{γ}) and Glu (H^{γ}-C^{δ}) crosspeaks. The data were acquired using 16 or 32 transients with1024 (^{1}H) * 64 (^{13}CO) complex points. The spectrum widths were 8.01 ppm and 20 ppm for the ^{1}H and ^{13}CO dimensions, respectively. The titration of the Asp and Glu side-chain carboxyl carbons was performed by measurement of the chemical shifts over the pH range of 2.09 to 6.37. All chemical shifts were determined using the 2D version of the HCACO experiments The HCACO spectra were processed using Topspin software. The pH was adjusted using HCl and NaOH. The pH was checked at the start and the end of each experiment and agreed within 0.1.

Data were fit to the Henderson-Hasselbalch equation in order to determine pKa values:

$$\delta(pH) = \left(\delta_{base} + \delta_{acid} 10^{(pK-pH)}\right) / \left(1 + 10^{(pK-pH)}\right) \tag{3.1}$$

Where δ is the Carbonyl carbon 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. Data were analyzed by the nonlinear least squares curve fitting using SigmaPlot. Data were also fit to a Hill equation that assumes that there are n protons titrating simultaneously at each position:

$$\delta(pH) = \left(\delta_{base} + \delta_{acid} 10^{n(pK-pH)}\right) / \left(1 + 10^{n(pK-pH)}\right) \tag{3.2}$$

The pKa of the C-terminal residue, F76, was determined by following the ^{13}C Hα chemical shifts of as a function of pH. TOCSY experiments with 75 ms mixing time were conducted at various pHs in order to determine the ^{13}C Hα chemical shifts of F76. The TOCSY experiments were performed at 25 ^{0}C with 1 mM protein in 10% D_2O. The data were acquired with 2048 (direct ^{1}H dimension)* 128 (indirect ^{1}H dimension) complex points. The spectra width was 12 ppm in both dimensions. The chemical shifts were internally referenced using DSS. The data were fit using equation 3.1 to determine the pKa.

3.2.7. Calculation of solvent accessibility

Solvent accessibilities were calculated using the GETAREA program (*128*) and are reported as % exposed. NMR structure (PDB code: 1VII) and X-ray structure (PDB code: 1YFR) were used. The solvent-accessible surface area of Carboxyl group, ε–amino group and Guanido group is reported as $Å^2$. The "random coil" value used for Lys, Arg, Asp and Glu in the % exposed calculation was the average solvent-accessible surface area of Lys, Arg, Asp and Glu in the tripeptide Gly-X-Gly (X= Lys, Arg, Asp or Glu) in an ensemble of 30 random conformations (Lys 164.5 $Å^2$, Arg 195.5 $Å^2$, Asp 113.0 $Å^2$, and Glu 141.2 $Å^2$). The tripeptides were generated by Pymol v0.99 (*83*).

3.3. Results

3.3.1. Design of mutants

The charged residues in HP36 are not uniformly distributed across the surface; rather there is a cluster of basic residues on one face of the protein surface. D44, E45, D46 and K48 located in the first helix; R55 is located in the second helix while K65 and K70 are found in the third helix; and K71, E72 and K73 are near the C-terminus. All of the charged groups are at least partially exposed to solvent although the hydrophobic portion of many of the charged sidechains are partially buried and involved in packing interactions. In the NMR structure (Table 3.1), the solvent accessibility of the Lys sidechain ranges from 88% exposed to 31% exposed while it varies from 93% to 60% for the acidic residues. K65 located in the third helix is the most buried charged residue. The solvent accessibility of K48 is 88 % solvent accessible, K65 is 31 % solvent accessible, K70 is 52 % solvent accessible, K71 is 76 % solvent accessible and K73 is 43 % solvent accessible. The solvent accessibilities of the ε–amino group in K48, K65, K70, K71 and K73 are 43.1 $Å^2$, 27.5 $Å^2$, 42.9 $Å^2$, 38.9 $Å^2$ and 35.7 $Å^2$, respectively. For comparison, the solvent accessibility of a fully exposed ε–amino group in a Gly-Lys-Gly tripeptide is 43.2 $Å^2$. The solvent accessibility of the carboxyl groups of the acidic residues (Cγ, $Oδ_1$, $Oδ_2$) range from 65% to 54%. The solvent accessibility of R55 is 50% and the solvent accessibility of its guanido group is 51.8 $Å^2$, while that of a fully exposed guanido group is 118.15 $Å^2$. There is one classic salt bridge in the HP36 structure involving R55 and D44. R55 is also involved in an interesting net work of interactions involving an apparent π-cation interaction with F47 as well as the interaction with D44. While in the X-ray structure (Table 3.1), E45 is 100% buried and K70 is the most buried charged residue. The

solvent accessibility of K48 is 71.5 % solvent accessible, K65 is 53 % solvent accessible, K70 is 30 % solvent accessible, K71 is 70 % solvent accessible and K73 is 55 % solvent accessible. The solvent accessibility of R55 is 48% and the solvent accessibility of its guanido group is 39.1 $Å^2$.

Several of the positively charged sidechains are relatively close to each other, and thus mutation of these might be expected to lead to an increase of stability. For example, K70 and K71 (NMR: 6.9 Å; X-ray: 11.2 Å) are adjacent positive charged residues, while K70 and R55 are 8.5 Å apart from each other in the NMR structure and 5.3 Å apart in the X-ray structure.

3.3.2. Analysis of point mutaions

We individually mutated each of the Lys residues to Met and the single Arg to Met. Lys to Met mutations were chosen since they are a reasonably isosteric mutation. A much better approximation to an isosteric substitution can be made for acidic residues. We mutated the acidic residues to their respective amide analogs. At 25 ^oC, far UV CD indicates that all the mutants have significant α-helix content and the shape and intensity of the spectra are very similar to that of the wild type protein (Appendix 1). In all of the 1D NMR spectra, characteristic resonances appeared at -0.1 ppm (Val-50) and at 5.6 ppm (Phe-47). Due to the low solubility of R55M, sample of R55M was prepared in 1M urea. The C4 proton from Phe47 overlaps with urea peak. These strongly shifted resonances are indicative of the global fold of HP36, and their observation provides further evidence that all of the mutants fold to a native structure which is the same as wild type (Appendix 2).

To assess the effect of the point mutations on the stability, thermal unfolding experiments and urea denaturation experiments monitored by CD were performed. Wild type HP36 has a transition midpoint (T_m) of 73.0 ^oC and the estimated free energy for unfolding (ΔG^o_U) is 3.22±0.12 kcal mol^{-1} at 25 ^oC based upon urea denaturation (Table 3.2). Three of the Lys to Met mutants, K48M, K65M and K70M are more thermostable than wild type by 5.0, 4.2 and 9.8 ^oC, respectively (Figure 3.2A and Table 3.2). While the other two, K71M and K73M have T_m values 1.8 and 3.5 ^oC lower than wild type, respectively. Urea denaturation experiments confirmed that the K48M, K65M and K70M mutants increase the stability (Figure 3.3A and Table 3.2). The K65M mutation increases the free energy of unfolding at 25 ^oC by 0.6 kcal mol^{-1}. K48M and K70M have larger effects. K48M and K70M do not exhibit complete urea unfolding curves and the unfolded baseline is not reached even at the highest urea concentration. This is not surprising since small proteins are expected to have broad unfolding transitions owing to their small m-values. GuHCl is a more powerful denaturant than urea and complete GuHCl induced unfolding curves can be observed. However GuHCl is a salt and thus not an appropriate denaturant if electrostatic interactions are the subject of interest. Furthermore, we have observed different apparent stabilities from GuHCl vs urea denaturation experiments with wildtype HP36 (2.93 kcal mol^{-1} from GuHCl denaturation and 3.22 kcal mol^{-1} from urea denaturation) providing further evidence that GuHCl is not an appropriate denaturant for this system. Thus we are forced to use urea and deal with the associated baseline

issues. The stability of these mutants was estimated in two ways. First the experimental curve can be fit, bearing in mind, that the derived parameters are likely to have a high uncertainty. Alternatively, the stability can be estimated from the midpoint concentration, C_M. The advantage of this approach is that C_M can be accurately determined even from an incomplete unfolding curve although the m-value, often can not. However stability can be estimated using the measured C_M value and either the wild type m-value (0.52 kcal mol^{-1} M^{-1}) or the average m-value for all the other mutants (0.51 kcal mol^{-1} M^{-1}). Fitting the curve directly give estimated ΔG^o values of 3.63 kcal mol^{-1} and 3.78 kcal mol^{-1} for K48M and K70M respectively. However these are unlikely to be reliable. In particular K48M and K70M clearly have a significantly larger C_M value than K65M yet the direct curve fit predicts they are less stable. This is implausible and for the reasons outlined above we believe the stability estimated using the C_M value and the average m-value is more accurate. Using the wild type or average m-value and the measured C_M values gives ΔG^o values of 4.05-4.13 kcal mol^{-1} and 4.32-4.41 kcal mol^{-1} for K48M and K70M respectively. This corresponds to an increase in ΔG^o of unfolding of 0.8 kcal mol^{-1} to 1.1 kcal mol^{-1}, respectively. Given that replacing a positively charged residue with a neutral nonpolar residue increases stability, it is natural to inquire if replacement by a negatively charged sidechain leads to an even great increase in stability. The most conservative substitution would be to replace the positively charged ε-amino group with a negatively charged group. Obviously there is no coded amino acid which corresponds to this substitution and the most conservatively naturally occurring variant is a Lys to Glu substitution. This will reverse the charge but will also change the size of the hydrophobic portion of the sidechain and could hence alter packing. The free energy of unfolding of the K70E mutant at 25 ^oC is identical to that of wildtype while its T_m is 6 ^oC less than wildtype. Thus the favorable free energy increment expected from removal of the ε-amino group is opposed by unfavorable effects. Analysis of the X-ray structure of HP36 (PDB code 1YRF) indicates that the Cβ and Cγ carbons of the K70 are essentially completely buried while the Cδ and Cε carbons are more than 80% buried. Modeling of a K70E substitution with the same sidechain dihedral angles suggests that the Cβ, Cγ and Cδ carbons are buried. The ε-oxygens are predicted to be 73% and 54% buried. Thus the K70E replacement might lead to unfavorable desolvation effects, provided that the model based on the X-ray structure offers a valid representation of the conformation of the E70 sidechain. In addition, the model predicts that the carboxylate of the E70 sidechain is close to the c-terminal carboxyl group, leading to a possible unfavorable coulombic contribution.

K71M and K73M are 0.22 and 0.17 kcal mol^{-1} less stable than wild type. The R55M mutation is the most destabilizing of the basic mutations decreasing the Tm by 6.0 ^oC and ΔG^o by 1.03 kcal mol^{-1}. R55 packs against the sidechain of F47 and also interacts with the sidechain of D44. The destabilizing effect of the R55M mutation suggests that these interactions are energetically significant. Double mutant analysis (*129*) confirms the importance of the F47/R55 interaction. To further probe the interactions made by R55, we analyzed a R55K mutant. This preserves the positive charge but changes the shape of the sidechain. This mutant is 0.4 kcal mol^{-1} less stable

than wildtype but it is more stable than R55M, suggesting that both the positive charge and shape of the side chain are the critical features at this position.

In contrast to the basic residues, where several stabilizing mutations were found, done of the mutations at the acidic sites were stabilizing. The D44N, E45Q and D46N mutants were less thermostable than wild type by 15.2, 4.3 and 8.1 ^{o}C, respectively; while E72Q has a T_m similar to wild type (Figure 3.2B, table 3.2). The urea denaturation experiments show that the unfolding free enegy of all the Asp to Asn and Glu to Gln mutants were 0.25~0.84 kcal mol^{-1} lower than wild type (Figure 3.3B and Table 3.2).

3.3.3 pKa values and pH dependent stability of wildtype HP36

The pH dependene of ΔG^{o} for HP36 can be calculated using equation 3.3 by inputting the pKa values for the native and denatured state:

$$\Delta\Delta G^{o}(pH - pH5) = RT\sum_{i=1}^{j}\ln\left[\frac{\left(1+10^{(pH-pK_i^N)}\right)\left(1+10^{(5-pK_i^D)}\right)}{\left(1+10^{(5-pK_i^N)}\right)\left(1+10^{(pH-pK_i^D)}\right)}\right] \quad (3.3)$$

Where *i* identifies the residue, *j* represents the number of glutamates and aspartates in the protein, pKa_i^N is the pKa value for the *i*th group in the native state and pK_i^D is the pKa value for the *i*th group in the appropriated model peptide. Equation 3.3 is defined from the Tanford-Wyman linkage relationship:

$$\frac{\partial\Delta G^{o}}{\partial pH} = (2.303)RT\Delta Q \quad (3.4)$$

Where ΔQ is the difference in the number of protons bound to the folded and unfolded state. The Tanford-wyman equation is valid under a wide range of conditions while equation 3.3 assumes that the titration behavior of the folded and unfolded ensembles can be described by a non-interacting sites model, i.e. that the titration behavior is described as a set of independently titrating groups.

The pKa values of the native state were directly measured by NMR experiments (Figure 3.4, Table 3.3). The pKa of E72 is near model compound values and the pKa's of D44, E45, D46 are all depressed relative to model compound values indicating that they are involved in favorable nature state electrostatic interations. This is broadly consistent with the stability studies which show that each of these three point mutants is destabilizing. There is not a one to one correspondence between $\Delta\Delta G^{o}$ and ΔpKa since the values of $\Delta\Delta G^{o}$ for D44N and D46N are virtually identical but the pKa shift is 0.4 units larger for D44.

When equation 3.3 is valid $\Delta\Delta G^{o}$ vs pH can be calculated provided a set of native and unfolded state pKa's is known. The relevant unfolded state pKa's are those for the unfolded state which is in equilibrium with the folded state under native conditions. This state is not directly accessible and it is very important to realize that its properties do not have to be, and are almost certainly not, the same as the urea or thermal unfolded states. Thus the appropriate set of unfolded state pKa's is not accessible. Equation 3.3 can still be used. One can test if using model compound pKa values for the unfolded state reproduces the experimental curve. If there is a

statistically significant deviation, then there must be unfolded state interactions which perturb the actual unfolded pKa values away from the model compound pKa's. Note that agreement does not prove that there are no unfolded state effects since the effect of a positive pKa shift at one site on $\Delta\Delta G^o$ could be counteracted by a negative pKa shift at a second site.

The predicated pH dependent changes in protein stability for wildtype HP36 referenced to pH 5.0 were calculated using values "random coil" unfolded state pKa values (Figure 3.5). The pH dependent stability of HP36 was experimentally dertermined from pH 2.0 to pH 6.0 (table 3.4). Urea induced unfolding was used between pH 3.0 and pH 6.0. Thermal unfolding was used at lower pH because of baseline problems with urea denaturation experiments. The stability at 25 ^oC was calculated using the Gibbs-Helmholtz equation together with a range of estimates of ΔC^o_P. ΔC^o_P of HP36 was estimated by plotting ΔH^o vs Tm for thermal unfolding at different pH's, this yields a value of 0.38 kcal mol^{-1} K^{-1} (Figure 3.6A). A plot of ΔH^o vs Tm for different mutants gives a value of 0.23 kcal mol^{-1} K^{-1} (Figure 3.6B). Sanchez-Ruiz and coworkers performed differential scanning calorimetry experiments for HP36. Because of the small size of HP36, the DSC transition is very broad. Thus the native and unfolded baselines are not clearly defined in the experimental thermograms and it is very difficult to do accurate extrapolation of ΔC^o_P for HP36, giving a value of 0.73 to 0.80 kcal mol^{-1} K^{-1}. From the literature, the value ofΔC^o_P of unfolding is expected to be about 0.012 kcal mol^{-1} K^{-1} per residue of protein (*42*). To a first approximation, the ΔC^o_P for HP36 can be calculated to be 0.43 kcal mol^{-1} K^{-1}. Another small 41-residue helical protein, the peripheral subunit-binding domain, has a ΔC^o_P value of 0.43 kcal mol^{-1} K^{-1} (*124*), suggesting that the estimate for HP36 is reasonable.

3.3.4. Stablizing mutants are monomeric

K48M, K65M and K70M are the only point mutants which increase the stability of the protein and K70M has the most favorable $\Delta\Delta G^o$. Replacing a charged surface residue with a hydrophobic amino acid could cause association and an apparent increase in stability, thus it is important to check whether the stabilizing mutants are monomeric or not. Analytical ultracentrifugation experiments show that K48M, K65M and K70M are monomeric (Figure 3.7). The data were fit well by an ideal single-species model with a molecular weight within 5% of the monomer molecular weight. Therefore, the increase in the stability for K48M, K65M and K70M is not due to association or aggregation.

3.4. Discussion

In this work, we have individually mutated all of the charged residues in HP36 to hydrophobic residues or the respective amide analogs. The mutants do not alter the structure of the domain. All the Asp to Asn and Glu to Gln mutants were less stable than wild type, while three of the Lys to Met mutants are more stable than wild type.

K65 is the most buried charged residue in the NMR structure, but the NMR structure of HP36 was determined at low pH (pH 3.7) (*84*) and the pH dependent

stability measurements of HP36 showed that the stability of HP36 decreased at low pH. Moreover, the simulations and experiments have showed that the X-ray structure is a more accurate representation of the structure in solution at neutral pH (pH 6.7) than the NMR structure at low pH (pH 3.7) (*129*). So in the X-ray structure, K70 is the most buried charged residue. Its ε-amino group has a calculated solvent accessibility of 34.0 $Å^2$. For comparison, the calculated solvent exposure of the ε-amino group in a fully exposed Lysine is 43.2 $Å^2$. Mutation of a partially buried charged residue that is not involved in a salt bridge should stabilize a protein. K70 is reasonably close to K71 and also R55, therefore substitution of K70 by Met should also remove unfavorable electrostatic interactions and stabilize the folded state. Thus the increased stability of the K70M mutant can be easily rationalized. The ε-amino groups in K48 and K70 are the most exposed among all of the Lys residues. The ε-amino group of K48 is not close to any positively charged sites. In the NMR structure the closest is R55 at 14.2 Å (measured to the center of the guanidine group). In the X-ray structure this distance is 13.3 Å. One possibility is that K48 might take part in favorable non-native interactions in the unfolded state which are disrupted upon mutation. Along these lines simulations of peptide fragments of HP36 suggest a stabilizing salt bridge could be formed in the unfolded state between D44 and K48 (*90*). This interaction is not found in the native state where the measured distance from the K48 ε-amino group to the D44 carboxylate is 8.5 Å in the NMR structure and 5.3 Å in the X-ray structure. If the simulations of the fragments are correct, and if they are translatable to the unfolded state of the intact protein, then the K48M mutant should disrupt the favorable unfolded state interactions. This will raise the free energy of the unfolded state, which will in turn lead to a favorable contribution to protein stability.

Electrostatic interactions in the unfolded state can be probed by pH dependent stability measurements. If the difference between the measured pH dependent stability of the protein and that calculated from the pKa values is big, then there must be electrostatic interactions in the unfolded state. If the mutation abolishes the interactions the deviation between the calculated and experimental pH stability profile will be much smaller. From pH dependent stability measurement of wildtype HP36, it is difficult to tell whether the deviation between the calculated and experimental pH dependent stability of wildtype HP36 is significant or is due to the uncertainty of denatured state pKa values of acidic residues and c-terminal residue F76 (Figure 3.5). Further study will be needed to measure the pKa of peptide fragments of HP36.

The most destabilizing mutation of a basic residue is the R55M mutant. The decrease in stability is likely due to the disruption of a π-cation interaction involving R55 and F47 as the interactions with D44 (*129*).

The work described here provides thermodynamic data for 12 different mutants at 10 distinct sites. The thermal stability of the mutants spans a wide range with the T_m varying by up to 25 ^oC, this provides an interesting system to test a proposed relationship between $\Delta\Delta G^o$ and ΔT_m. Rees and Robertson (*130*) analyzed equilibrium unfolding data for a set of globular proteins and were able to show that for their

dataset, there was a simple relationship between $\partial\Delta G(T^*)/\partial\Delta T_m$ and the number of residues and the T_m of the protein.

$$\frac{1}{N}\frac{d\Delta G(T^*)}{\Delta T_m} = 0.029 - \frac{2316}{T_m^{\,2}} \tag{3.4}$$

In equation 3.4, $\Delta G(T^*)$ is the stability at the temperature of maximum stability, N is the number of residues and T_m is the midpoint of the thermal unfolding transition. The relationship predicts that a plot of $\Delta\Delta G(T^*)$ vs ΔT_m should be linear for small value of ΔT_m. For HP36, the predicted slope is 0.832 kcal mol^{-1} K^{-1}. Using the T_m of wildtype, a plot of $\Delta\Delta G(298K)$ vs ΔT_m is displayed in Figure 3.8. The plot is nicely linear with R^2 of 0.75 and a P-value of 0.003. The calculated slope is 0.0903±0.016 kcal mol^{-1} K^{-1} which is in good agreement with the predicted value.

This study provides an experimental view of the contribution of surface charged residues to the stability of HP36. The results should be useful for those who wish to test computational approaches to predict electrostatic interactions. The demonstration of monomeric variants of HP36 with increased stability will prove useful for protein folding studies involving this domain. Furthermore, this paper provides additional evidence that removing unfavorable surface interactions is a productive strategy for enhancing protein stability. A handful of such successful studies have been reported for a range of different folds supporting the notion that the approach is general and robust (*39, 73, 117, 118*). It is also interesting to note that three quite simple mutations lead to clear increases in stability, in contrast, repacking of the hydrophobic core of proteins often requires multiple substitutions to achieve similar increases in ΔG^o.

The data presented here also argues quite strongly that "mini-proteins" have not been subjected to strong evolutionary pressure to optimize stability. Instead, like their larger "normal" counterparts, mutations can be found that increase stability. In addition, the thermodynamic parameters of the HP36 mutants agree quite well with predictions based on a data base of large proteins, reinforcing our view that mini-proteins do not require unusually structural or thermodynamic features to be stable (*131*). The notion that miniature proteins require "special" interactions appears to result from an influential early analysis that estimated the minimum size of a stable globular protein domain. That analysis predicted that the minimum domain size is on the order of 50 residues. What has sometimes been forgotten is that the analysis was attempting to predict the minimum size required to insure that the stability will be on the order of 20RT (*132*). The stability of HP36 is only about 5.3RT and is similar to other mini-proteins. Thus there is no inherent contradiction between the early analysis and the existence of folded mini-proteins, hence there is no need to invoke special interactions when considering the thermodynamics of small proteins.

Table 3.1: Solvent accessibilities of charged residues in HP36. The solvent-accessible surface area of carboxyl group of Asp and Glu residues, ε–amino group of Lys residues and Guanido group of Arg residue, and comparing with their fully exposed carboxyl group for Asp is 73.9 $Å^2$and Glu is 86.5 $Å^2$; ε–amino group for Lys is 44 $Å^2$; Guanido group for Arg is 118.15 $Å^2$. The NMR structure (1VII), X-ray structure (1YRF), simulation structure and HP36 K70M X-ray structure were used.

	Solvent accessibility (%)				Carboxyl group($Å^2$) for Asp and Glu ε–amino group ($Å^2$) for Lys Guanido group($Å^2$) for Arg				Percentage of exposed surface area comparing with fully exposed Carboxyl, ε–amino, Guanido groups (%)			
	NMR	X-ray	Simulation	X-Ray K70M	NMR	X-ray	Simulation	X-Ray K70M	NMR	X-ray	Simulation	X-Ray K70M
D44	92.0	79.8	70.3	21.4	72.1	53.7	46.85	21.5	97.6	72.7	63.4	29.1
E45	93.7	91.6	88.6	32.8	81.0	67.0	56.1	11.4	93.6	77.5	65.0	13.2
D46	60.1	60.2	61.0	37.2	66.0	63.3	61.9	36.7	89.3	85.6	83.8	50.0
E72	65.5	65.7	65.8	38.6	75.5	78.1	74.0	37.1	87.3	90.3	85.5	42.9
K48	87.9	64.5	51.6	23.3	43.1	43.9	45.2	8.55	98.0	100	100	19.4
K65	31.2	42.8	51.9	22.4	27.5	42.8	39.8	16.6	62.5	97.3	90.0	37.7
K70	51.7	30.2	34.1	-----	42.9	34.0	26.2	---	97.5	77.3	60.0	-----
K71	75.5	64.9	83.9	28.4	38.9	36.9	42.95	22.67	88.4	83.9	97.6	51.5
K73	43.0	55.3	39.1	12.6	35.7	34.8	38.11	9.41	81.1	79.1	86.6	21.4
R55	50.3	47.8	47.9	23.2	51.8	39.1	29.0	23.0	43.8	33.1	24.5	19.5

Table 3.2: Summary of equilibrium stability measurements for wild type HP36 and its mutants, from thermal unfolding and from urea denaturation at 25 ^{o}C.

Protein	T_m (^{o}C)	$\Delta H(T_m)$ (kcal mol^{-1})	ΔG^o (kcal mol^{-1})	m value (kcal mol^{-1} M^{-1})
WT HP36	73.0±1.5[a]	31.82	3.22±0.12[a]	0.52±0.01[a]
D44N	57.8	32.08	2.48	0.55
E45Q	68.7	30.64	2.85	0.54
D46N	64.9	28.58	2.38	0.53
K48M	78.0	34.07	4.05-4.13 [b]	0.51-0.52 [b]
			3.63[c]	0.44[c]
R55M	67.3	26.30	2.19	0.43
R55K	68.1	31.04	2.86	0.49
K65M	77.2	33.20	3.87	0.56
K70M	82.2	33.44	4.32-4.41 [b]	0.51-0.52 [b]
			3.78[c]	0.45[c]
K70E	66.7	31.96	3.20	0.49
K71M	71.2	31.86	3.00	0.49
E72Q	73.6	31.23	2.97	0.51
K73M	69.5	30.42	3.05	0.49
N68A	76.13	37.13		
HP36-CONH$_2$	65.5[d]	23.78	2.86	0.50

[a] The uncertainty in Tm as estimated by repeated measurements of wild type is 1.5 ^{o}C. The estimated uncertainties in ΔG^o and the m values are on the order of 0.12 kcal mol^{-1} and 0.01 kcal mol^{-1} M^{-1} respectively based on repeated measurements of wild type.

[b] Estimated using the measured C_M value and either the wild type m-value (0.52 kcal mol^{-1} M^{-1}) or the average m-value for all the other mutants (0.51 kcal mol^{-1} M^{-1})

[c] Estimated by direct fitting of the experimentally observed partial denaturation curve.

[d] In 10 mM deuterium sodium acetate, 150 mM sodium chloride buffer solution at pD 5.0.

Table 3.3: Native state pKa values for the acidic residues and C-terminal residue in wildtype HP36.

Residue	pKa (Henderson-Hasselbalch)	pKa (Hill Equation)	N (Hill coefficient)	R^2
E45	3.95±0.02	3.95±0.01	0.90±0.03	1.00
E72	4.37±0.03	4.37±0.03	1.00±0.05	1.00
D44	3.04±0.12	3.10±0.11	1.30±0.32	0.96
D46	3.44±0.11	3.45±0.12	1.05±0.29	0.97
F76 (αH)	3.17±0.14	3.24±0.12	1.41±0.53	0.97
F76 (NH)	3.07±0.01	3.09±0.01	1.09±0.02	1.00

Solutions contained 10mM sodium acetate and 150 mM sodium chloride in 90% H_2O, 10% ^2H_2O at 25^oC

Table 3.4: pH dependent stability measurements for wildtype HP36 from thermal denaturations and urea denaturations at 25 ^{o}C. pH dependent stability measurements for K48M from urea denaturations at 25 ^{o}C.

Protein	**Wildtype HP36**				**K48M**	
pH	**T_m(^{o}C)**	**$\Delta H(T_m)$ (kcal mol^{-1})**	**C_M (M)**	**$\Delta G^{o}_{(m=0.51)}$[a] (kcal mol^{-1})**	**C_M (M)**	**$\Delta G^{o}_{(m=0.51)}$[a] (kcal mol^{-1})**
2.00[b]	46.26	24.06				
2.50[b]	45.99	19.91				
	46.51	20.03				
2.60[b]	48.26	20.19				
3.00[b]	52.10	26.15	2.53	1.29	3.96	2.02
3.50[b]	57.74	28.18	3.27	1.67	4.75	2.42
4.00[c]	63.76	30.45	4.33	2.21	5.58	2.85
4.50[c]	68.15	30.93	5.53	2.82	6.83	3.48
5.00[c]	73.02	31.82	6.22	3.17	7.96	4.06
5.50[c]			6.48	3.30	8.01	4.09
6.00[d]			6.64	3.39	8.43	4.30

[a] Estimated using the measured C_M value the average m-value for all the other mutants (0.51 (kcal mol^{-1} M^{-1})

[b] 10 mM citric buffer

[c] 10 mM acetate buffer

[d] 10 mM phosphate buffer

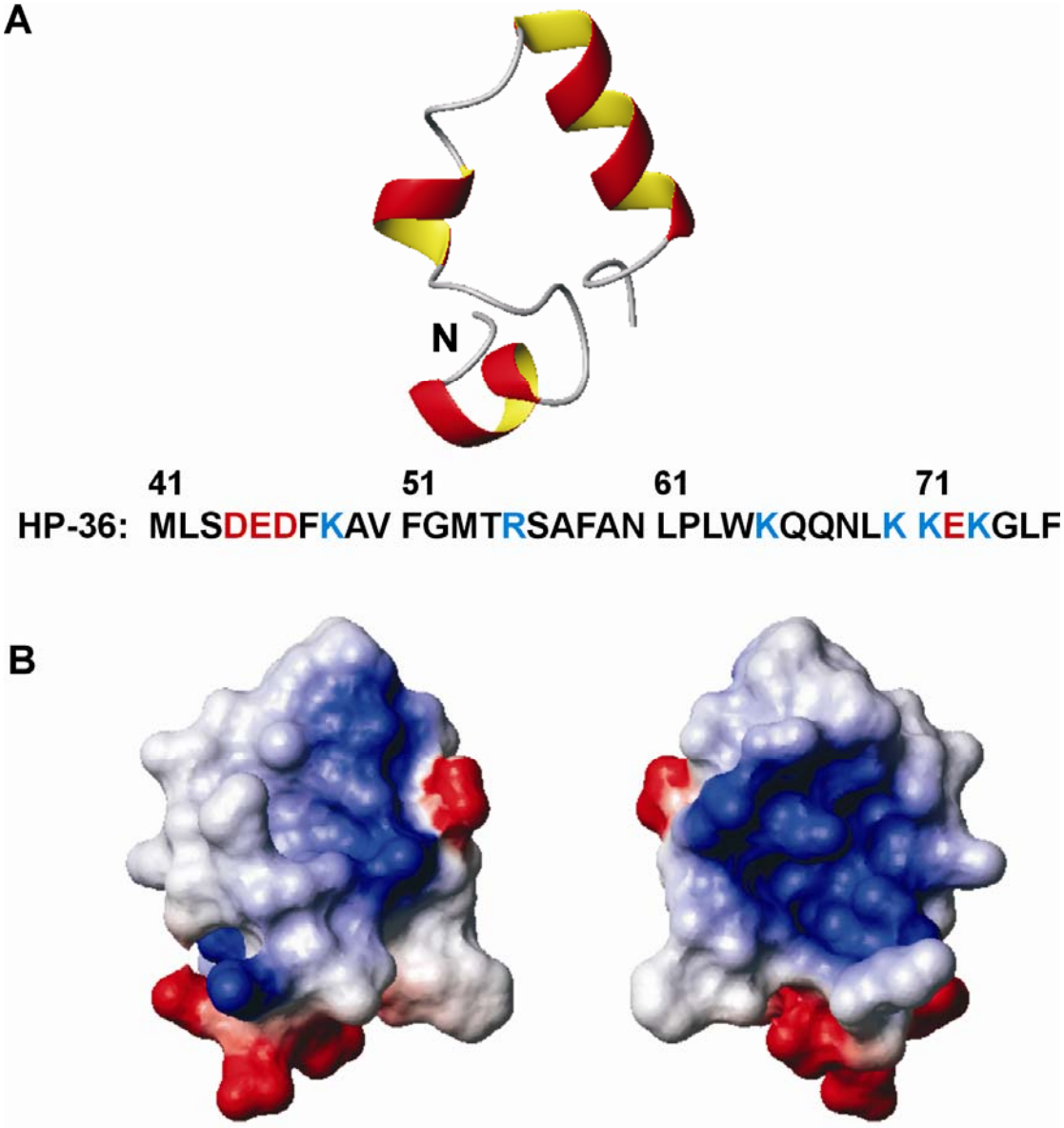


Figure 3.1: (A) Ribbon diagram and primary sequence of the villin headpiece subdomain (HP36); (B) Surface potential representation displayed in the same orientation as the ribbon diagram and after rotation by ~180^{o} about the y-axis. Positive potential is shown in blue and negative in red. All figures were created with the program MOLMOL (*133*) (PDB code 1VII).

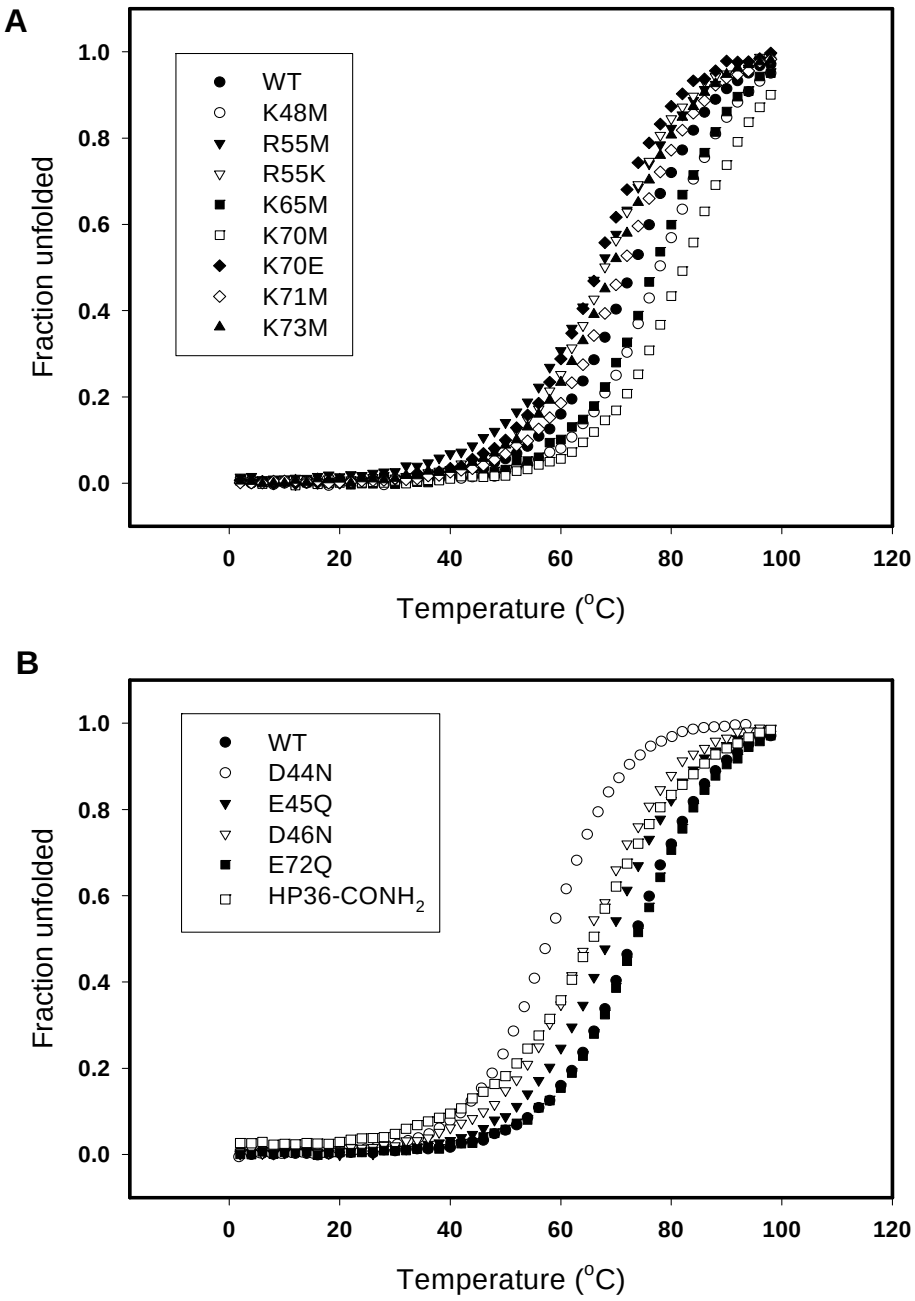


Figure 3.2: Thermal unfolding of HP36 and its mutants monitored by CD at 222 nm, data is plotted as fraction unfolded. All experiments were conducted at pH 5.0, 10 mM sodium acetate and 150 mm sodium chloride. (A)●: wild type HP36; ○: K48M; ▼: R55M; △: R55K; ■: K65M; □: K70M; ◆: K70E; ◇: K71M; ▲: K73M. (B) ● : wild type HP36; ○ : D44N; ▼ : E45Q; △ : D46N; ■ : E72Q; □ : HP36-$CONH_2$.

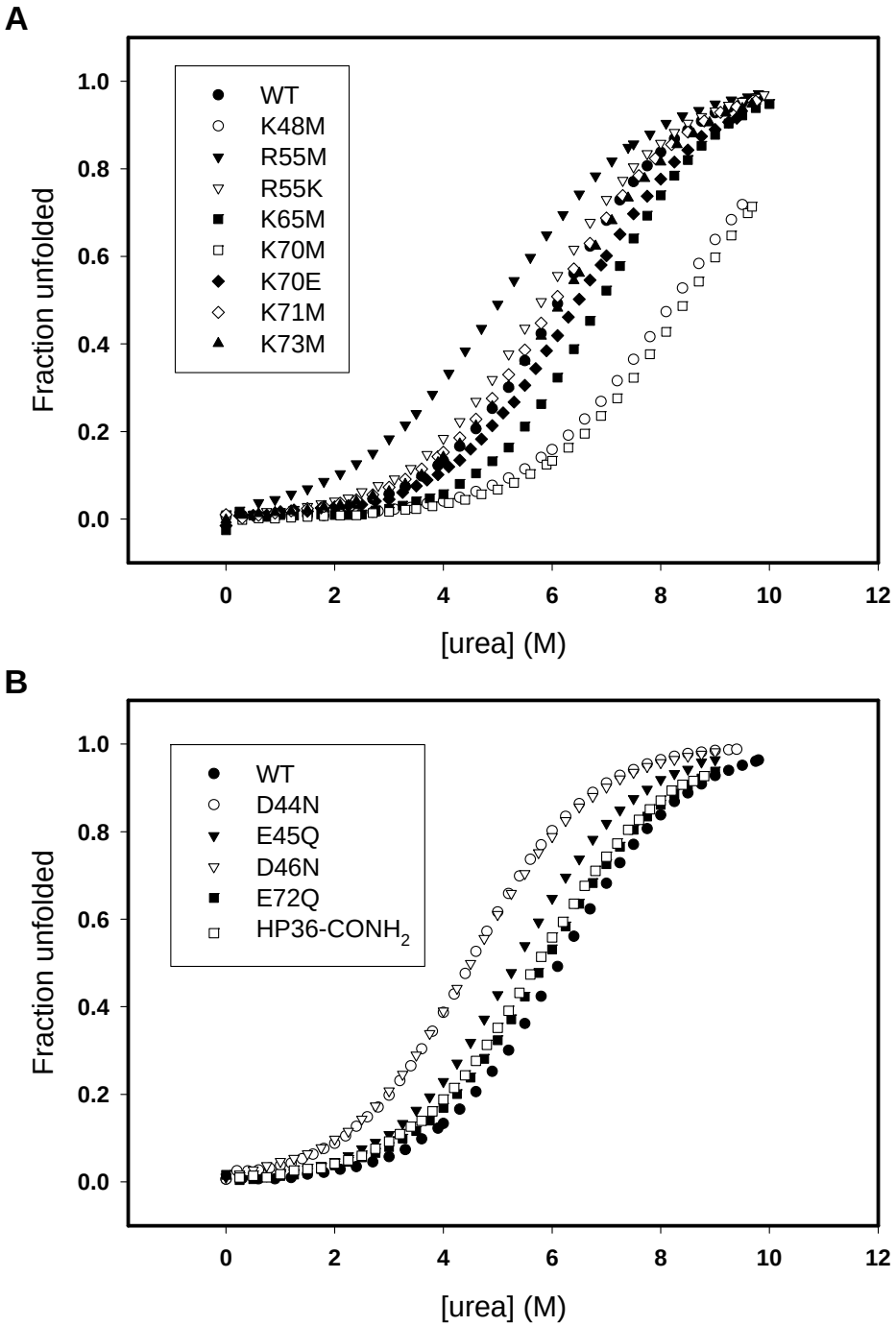


Figure 3.3: Urea denaturation of HP36 and its mutants. All experiments were conducted at 25 ^{o}C, pH 5.0, 10 mM sodium acetate and 150 mm sodium chloride. (A) ●: wild type HP36; ○: K48M; ▼: R55M; △: R55K; ■: K65M; □: K70M; ◆: K70E; ◇: K71M; ▲: K73M. (B) ●: wild type HP36; ○: D44N; ▼: E45Q; △: D46N; ■: E72Q; □: HP36-$CONH_2$.

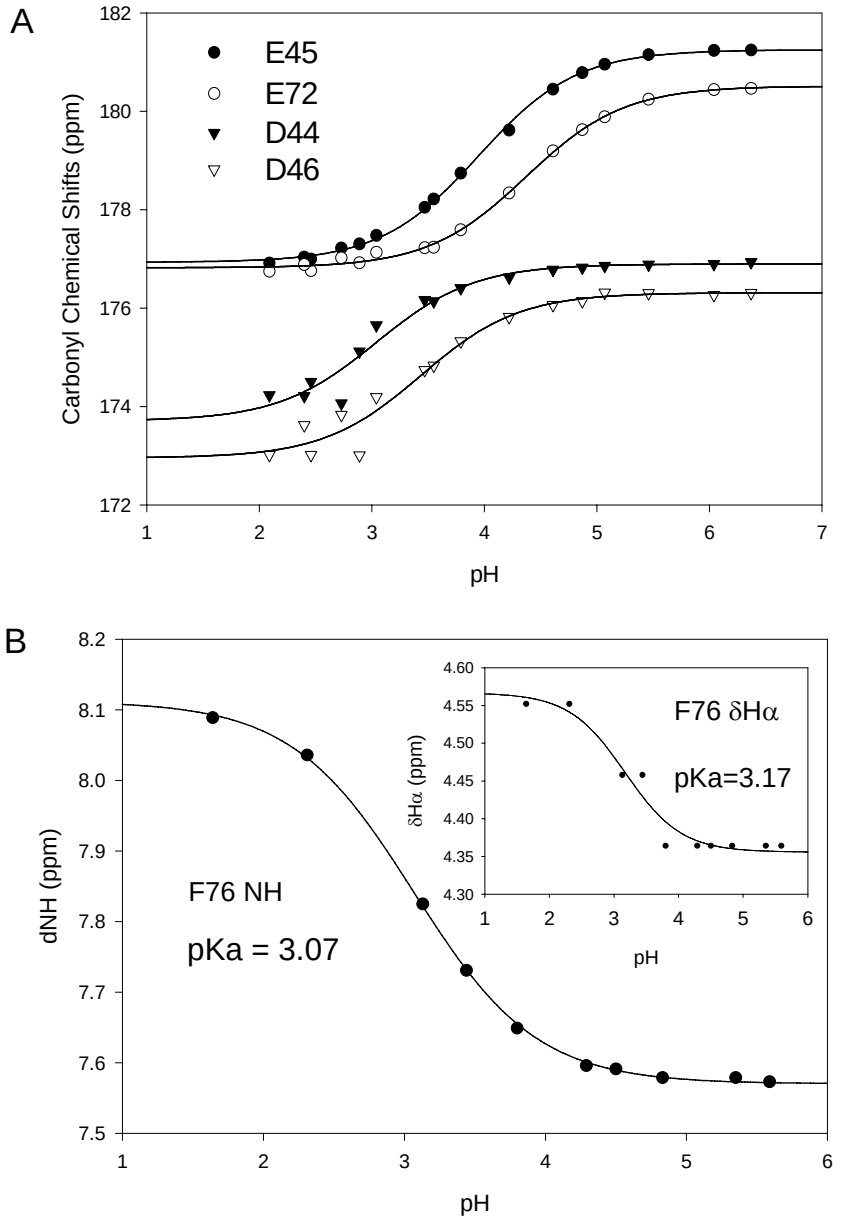


Figure 3.4: (A) Carbonyl carbon chemical shifts as a function of pH for wildtype HP36. The filled circles represent Glu 45, the open circles represent Glu 72, the filled triangles represent Asp 44 and the open triangles represent Asp 46. (B) Amide proton and α proton (insert) chemical shifts as a function of pH for wildtype of c-terminal residue F76. The data were fit with the Henderson-Hasselbalch equation (eq 4) to determine pKa values. The results are listed in Table 2. Measurement were made in 10 mM NaAC, 150 mM NaCl at 25 ^{o}C.

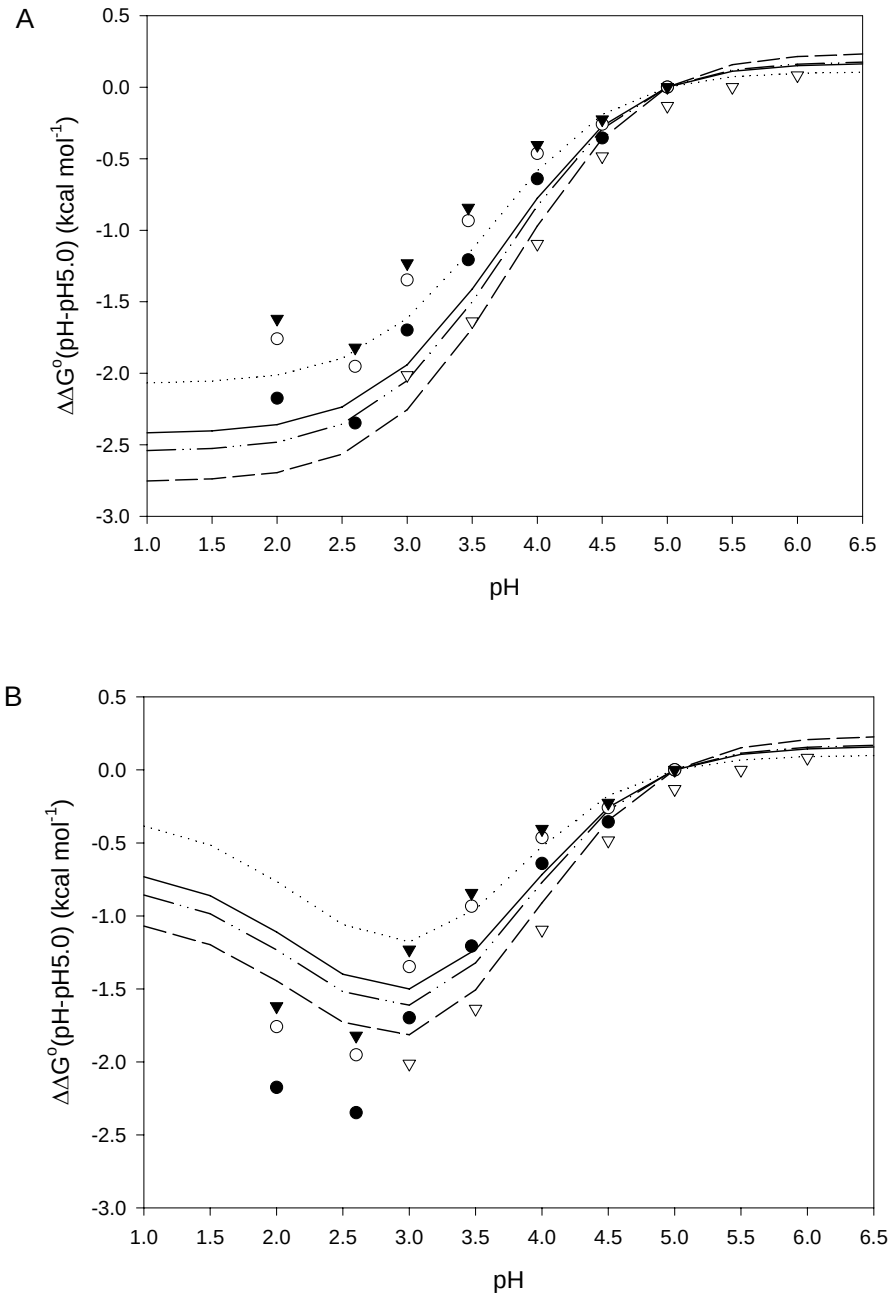


Figure 3.5: pH dependent changes in protein stability for wildtype HP36 referenced to pH 5.0. The continuous line (pKa^D : D3.95,E4.4), dashed line (pKa^D : D4.0, E4.5), dashed-dotted line (pKa^D : D4.0, E4.4) and dotted line (pKa^D : D3.9, E4.3) represent the predicted stability change using native state pKa values and peptide model pKa values as indicated (pKa^D for Asp ranging from 3.9-4.0 and pKa^D for Glu ranging from 4.3-4.5). The open triangles (▽) represent experimental values from urea denaturation. The filled triangles (▼), filled circles (●) and open circles (○) represent calculated values from the Gibbs-Helmholtz equation using different changes in heat capacity: "▼" 0.43 kcal mol^{-1} deg^{-1}, "●" 0.23 kcal mol^{-1} deg^{-1} and "○" 0.38 kcal mol^{-1} deg^{-1}. (A) without C-terminus (B) include C-terminus.

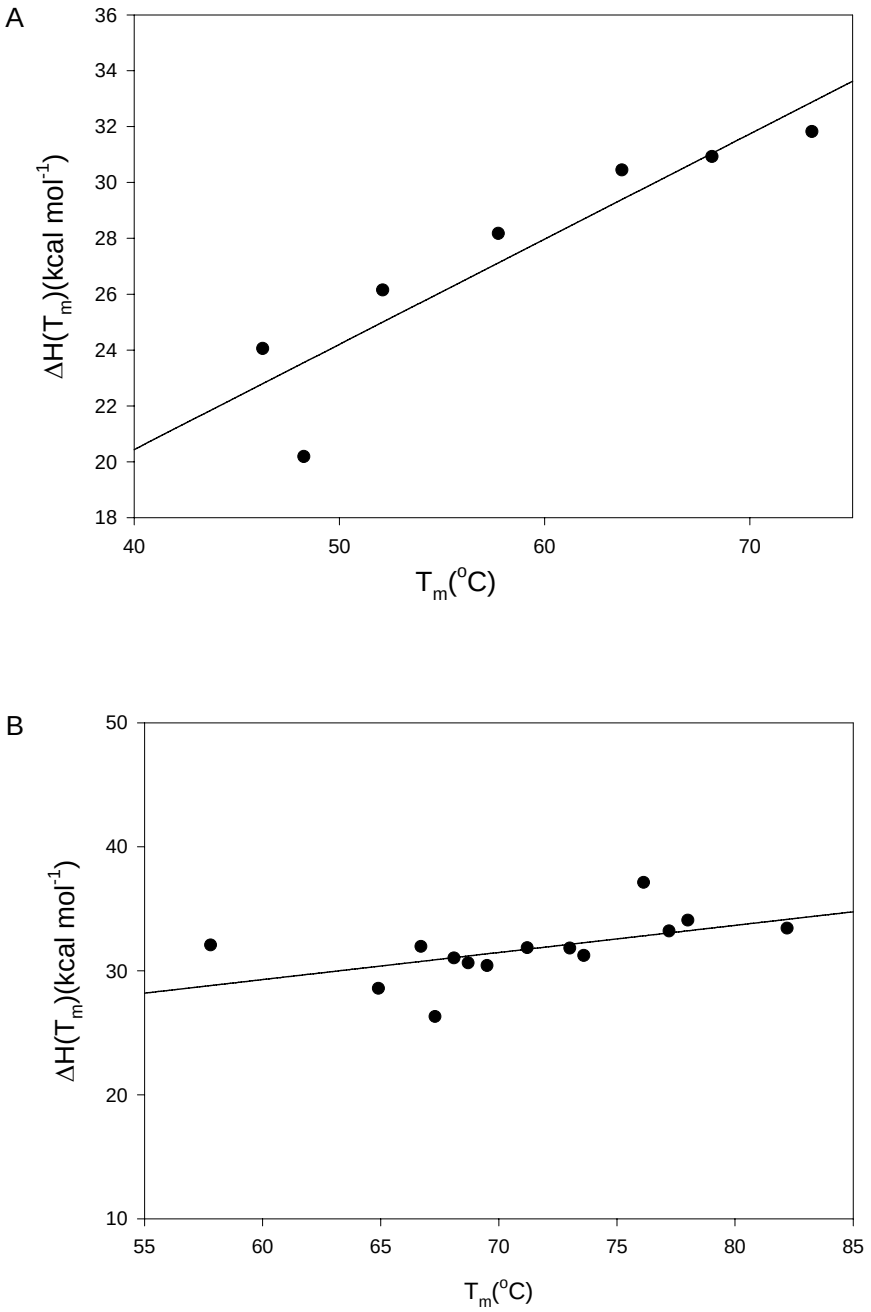


Figure 3.6: $\Delta H^o(T_m)$ vs T_m for (A) thermal unfolding of HP36 at different pH's and (B) for HP36 and different mutants (see Table 3.2). A linear fit to these points has a slope of 0.38 kcal mol^{-1} K^{-1} for (A) and 0.23 kcal mol^{-1} K^{-1} for (B).

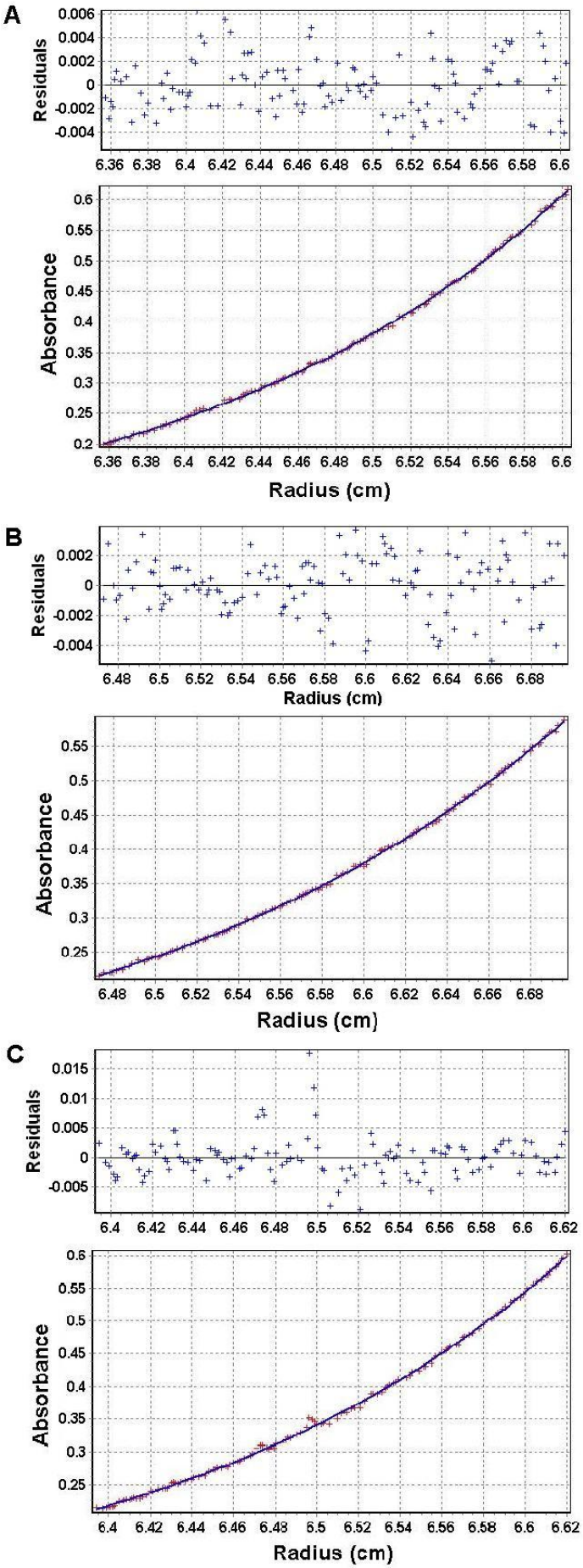

A
Residuals
0.006
0.004
0.002
0
-0.002
-0.004
6.36 6.38 6.4 6.42 6.44 6.46 6.48 6.5 6.52 6.54 6.56 6.58 6.6
Absorbance
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
6.36 6.38 6.4 6.42 6.44 6.46 6.48 6.5 6.52 6.54 6.56 6.58 6.6
Radius (cm)
B
Residuals
0.002
0
-0.002
-0.004
6.48 6.5 6.52 6.54 6.56 6.58 6.6 6.62 6.64 6.66 6.68
Radius (cm)
Absorbance
0.55
0.5
0.45
0.4
0.35
0.3
0.25
6.48 6.5 6.52 6.54 6.56 6.58 6.6 6.62 6.64 6.66 6.68
Radius (cm)
C
Residuals
0.015
0.01
0.005
0
-0.005
6.4 6.42 6.44 6.46 6.48 6.5 6.52 6.54 6.56 6.58 6.6 6.62
Absorbance
0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
6.4 6.42 6.44 6.46 6.48 6.5 6.52 6.54 6.56 6.58 6.6 6.62
Radius (cm)

Figure 3.7: Analytical ultracentrifugation analysis of (A) K48M (B) K65M and (C) K70M. Experiments were conducted at 25 ^{o}C, pH 5.0, in 10 mM sodium acetate buffer solution. The best fit to a single species is shown and the residuals are plotted below each curve. In all cases the calculated molecular weight is with 5% of the expected value. Expected molecular weight is 4190.1 Da and observed molecular weight: K48M, 4258 Da; K65M, 4320 Da; K70M, 4401Da.

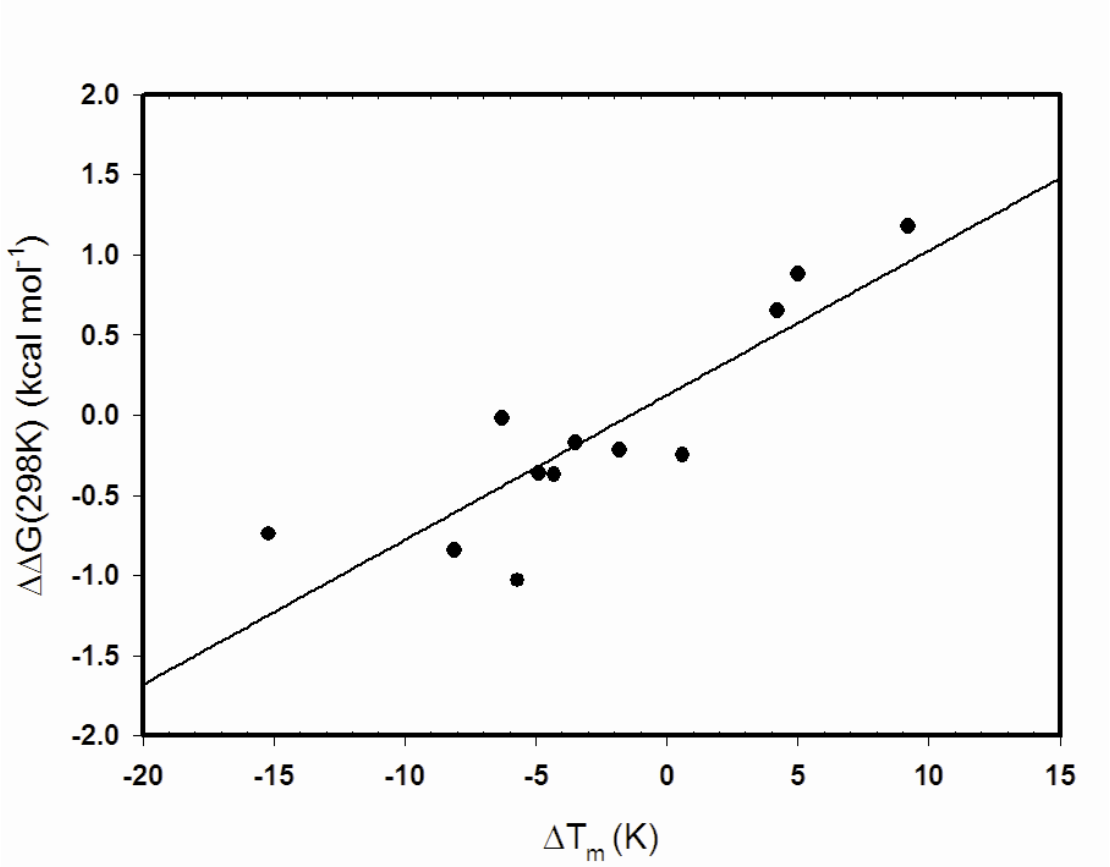


Figure 3.8: Relationship between ΔT_m (difference between mutants and wildtype HP36) and $\Delta\Delta G^o$(298K) (difference between mutants and wildtype HP36) at 25°C. A linear fit to these points has a slope of 0.0903±0.016 kcal mol^{-1} K^{-1} with R 0.866 and a P-value of 0.003.

4. Rational Design, Structural and Thermodynamic Characterization of a Hyperstable Variant of the Villin Headpiece Helical Subdomain

Abstract

A hyperstable variant of the small independently folded helical subdomain (HP36) derived from the F-actin binding villin headpiece was designed by targeting surface electrostatic interactions and helical propensity. A double mutant N68A, K70M was significantly more stable than wild type. The T_m of wild type in aqueous buffer is 73.0 ^oC while the double mutant did not display a complete unfolding transition. The double mutant could not be completely unfolded even by 10 M urea. In 3M urea the T_m of wild type is 54.8 ^oC while that of the N68AK70M double mutant is 73.9 ^oC. Amide H/^{2}H exchange studies show that the pattern of exchange is very similar for wild type and the double mutant. The structures of a K70M single mutant and the double mutant were determined by X-ray crystallography and are identical to wild type. Analytical ultracentrifugation demonstrates that the proteins are monomeric. The hyperstable mutant described here is expected to be useful for folding studies of HP36 since studies of the wild type domain have sometimes been limited by its marginal stability. The results provide direct evidence that naturally occurring miniature protein domains have not been evolutionarily optimized for global stability. The stabilizing effect of this double mutant could not be predicted by sequence analysis since K70 is conserved in the larger intact headpiece for functional reasons.

Acknowledgments

The material presented in this chapter has been published (Yuan Bi, Jae-Hyun Cho, Bing Shan, Eun-Young Kim, Hermann Schindelin and Daniel P. Raleigh 2007 *Biochemisty 46*, 7497-7505). This chapter contains direct excerpts from the manuscript written by Yuan Bi with suggestions and revisions from Professor Daniel P. Raleigh. Dr. Jae-Hyun Cho and Dr. Eun-Young Kim contributed to the work of crystallization. Mr. Bing Shan contributed to part of the assignment of HP36N68AK70M. I thank Prof. Mcknight for his continued interest in this work and for numerous helpful discussions. We also thank Dr. Yuefeng Tang for her help with the NMR experiments and for helpful discussions, Ms Benben Song for helpful discussions and Ms. Lauren Wickstrom for help with the RMSD calculations.

4.1. Introduction

A common goal of protein engineering studies is to increase protein stability. Increased protein stability is important for applications in biotechnology since the performance of protein pharmaceuticals and proteins used as biocatalysts can often be improved by stabilization (*58-60*). Enhanced stability can also be extremely useful in biophysical studies, particularly for the study of marginally stable proteins and protein domains. In this case, protein engineering experiments are often of limited utility

because even modestly destabilizing mutations can drastically depopulate the folded state. There are also theoretical reasons why developing more stable proteins is of interest particularly for proteins which fold rapidly. For example, recent work in protein folding has lead to the hypothesis that stabilizing a rapidly folding protein can lead to downhill folding, i.e. folding which does not involve crossing a free energy barrier (*115*). The analysis predicts that stabilizing a protein which folds at a rate near the speed limit for folding could lead to downhill folding. A wide variety of approaches have been applied to increase protein stability, sometimes with mixed success (*39, 40, 61-63*). Here we describe the rational design of a hyperstable variant of the small, fast folding villin headpiece helical subdomain (HP36).

HP36 is the helical C-terminal subdomain from the villin headpiece and can fold in isolation. The subdomain is one of the smallest known naturally occurring proteins that folds to a compact native state. The complete villin headpiece domain is a 76 residue F-actin binding domain that is found at the extreme C-terminus of villin and related actin binding proteins involved in cytoskeleton remodeling (*75, 81, 82, 134, 135*). Headpiece domains are not related to any other F-actin binding proteins. Intact villin is a multidomain protein comprised of a set of gelsolin domains and headpiece. It has the interesting property that it can acts as both an actin bundling protein and actin severing protein. Actin bundling requires two independent actin binding sites and the headpiece domain provides the second F-actin binding site (*75, 134*). The first 9 residues of headpiece can be removed without comprising actin binding activity and the resulting construct is commonly referred to as HP67 (*81*). The N-terminal potion of the protein is made up of a set of loops turns and a short helix, while the C-terminal portion, here designate HP36, comprises a small three helix bundle(*81, 85, 135*). HP67 is the smallest characterized fragment of headpiece which retains full actin binding activity. Extensive structural and mutational studies from the McKnight group have lead to a model of the binding site (*81, 82, 135*). Key residues include K38, E39, K65, the sequence comprised of residue 70 to 74 and a possibly L75 and F76. E39 and K70 are involved in a structurally important buried salt bridge and are important for maintaining the structural integrity of the protein, while the other residues group together on the protein surface. HP67 contains a single contiguous hydrophobic core but none the less the C-terminal helical subdomain can fold independently.

HP36 is composed of three short α-helices (Figure 4.1) which pack to form the hydrophobic core. The protein contains four acidic and six basic residues, all of which are located on the protein surface. They are not distributed uniformly on the molecular surface but rather form several clusters. The helical subdomain of headpiece starts at residue 42 which is a Leu and continues to the C-terminus. This subdomain is commonly denoted HP35. Many experimental studies of the subdomain use a construct which contains an additional N-terminal Met resulting from the expression system. This protein is usually referred to as HP36 and we adopt this notation here. The first residue of HP36 is denoted as Met-41 to conform to the notation used in earlier studies.

Small proteins have become useful model systems for experimental, theoretical and computational investigations of protein folding (*21, 84-91, 110, 111, 113, 136*).

HP36 in particular has been widely studied by a diverse set of investigators. Its small size and rapid folding have made it an exceptionally popular model protein for theoretical and computational studies of protein folding and dynamics. At least 17 separate research groups have published theoretical or computational studies of its folding (*90, 136-153*). There are fewer experimental studies of its folding, but we and independently the Eaton group have demonstrated that HP36 folds on the microsecond timescale (*86, 87*). One issue which has limited experimental studies is the modest stability of the domain. Wild type HP36 has a T_m of 73.0 ^oC, a relatively high value, but, its free energy of unfolding at 25 ^oC is only 3.22 kcal mol^{-1} at pH 5.0 (Table 4.1). This is typical for small globular proteins. Their small size normally means they have a small ΔCp^o and a small ΔH^o of unfolding which leads to relatively high T_m's even though ΔG^o at 25 ^oC can be quite modest. The small value of ΔG^o has limited mutational studies of the folding of HP36, particularly because these investigations typically involve temperature jump studies which are usually carried out above room temperature where the protein is even less stable. For example the hydrophobic core of HP36 contains an unusual triad of Phenylalanines which pack together in an approximately triangular arrangement. Studies of this interesting interaction have made use of Phe to Leu substitutions, but have been hindered because some of the mutants are so destabilizing that the domain is not completely folded (*21, 154*). In this work, we design and characterize a hyperstable mutant of HP36. We used X-ray crystallography to determine the structures of the single mutant K70M and the thermal hyperstable double mutant N68AK70M. Thermodynamic and H-D exchange experiments were also used to characterize the hyperstable mutant.

4.2. Materials and methods

4.2.1. Cloning, expression, and purification.

The plasmid (pET3a-NTL9-FXa-HP36) containing the gene for HP36 was prepared as described (*120*). The proteins with coded amino acid substitutions were expressed and purified as described (*120*). ^{15}N-labeled HP36 and N68AK70M were produced by using M9T minimal media with $^{15}NH_4Cl$. All proteins were more than 95% pure. The identities of the proteins were analyzed by electro-spray ionization mass spectrometry (ESI MS) or matrix-assisted laser desorption and ionization time-of-flight mass spectrometry (MALDI-TOF). The expected and observed molecular weights were as follows: N68A, expected 4146.8, observed 4146.8; N68AK70M, expected 4149.9, observed 4150.2. ^{15}N-labeled HP36, expected 4238.9, observed 4238.5; ^{15}N-labeled N68AK70M, expected 4195.9, observed 4195.6.

4.2.2. Circular Dichroism (CD) Spectroscopy.

All CD experiments were performed on an Aviv 62A DS and a 202SF Circular Dichroism spectrophotometer. A 10 mM sodium acetate buffer was used for all CD experiments at pH 5.0. The protein concentrations were determined from absorbance measurements using the method of Pace & coworkers (*122*). Far-UV wavelength scans were performed with five repeats and a signal averaging time of 1 second per point, in a 1 mm quartz cuvette, over the range of 195 nm to 260 nm. Thermal

unfolding and urea denaturation experiments were carried out in a 1 cm quartz cuvette by monitoring the signal at 222 nm. The concentration of urea solutions were measured by measuring the refractive index. The concentration of urea was increased from 0 M to 10 M in ~0.25 M steps. Wavelength scans and urea denaturation experiments were performed at 25 ^{o}C. Thermal unfolding experiments were performed from 2 ^{o}C to 98 ^{o}C with a 2 ^{o}C interval. For the thermal unfolding experiment in 3 M urea, the protein samples were dissolved in 10 mM sodium acetate, 150 mM sodium chloride and 3 M urea at pH 5.0. All the denaturation data were analyzed by nonlinear least squares curve fitting using SigmaPlot, as described (*123-125*).

4.2.3. Structural determination.

Both the HP36 K70M and HP36 N68AK70M mutants were lyophilized and dissolved in 10 mM sodium acetate (pH 5.0) and 150 mM NaCl. The final protein concentrations of K70M and N68AK70M were around 1mM and 500 uM, respectively. Crystals were obtained by incubating the sample at 4 °C. Diffraction data were collected on beamline X26C at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. Data were indexed, integrated and scaled using HKL. The structure of the both mutants were solved by molecular replacement with MOLREP using 1YRF (HP35-N68H) (*87*) as a search model. Parameters are given in Table 4.2. All structural Figures were generated using PyMOL v0.97 (*83*).

4.2.4. Analytical Ultracentrifugation.

Analytical ultracentrifugation of HP36 N68AK70M mutant was performed with a Beckman Optima XL-A analytical ultracentrifuge at 25ºC using rotor speeds of 38,000 rpm (24hours) and 48,000 rpm (24hours). Molecular masses were determined at initial concentrations of 30, 60, 100uM in 10mM sodium acetate and 150mM sodium chloride buffer (pH 5.0). Six-channel, 12 mm path length, charcoal-filled Epon cell with quartz windows were used. The wavelength used for this experiment was 280nm. Ten scans were averaged. The partial specific volume (0.7501 l g^{-1}) and solution density (1.005 g l^{-1}) were calculated from the software program SEDNTERP (*121*). The HeteroAnalysis program from the Analytical Ultracentrifugation Facility at the University of Connecticut was used for data analysis.

4.2.5. NMR Spectroscopy.

All NMR experiments were carried out on a Varian Instruments Inova 500MHz nuclear magnetic resonance spectrometer at 25 ^{o}C. Samples of 1mM proteins for 1D NMR were prepared in 10% D_2O, 90% 10 mM sodium acetate and 150 mM sodium chloride with pH 5.0. ^{15}N-labeled HP36 and ^{15}N-labeled N68AK70M for H1/H2 exchange experiments were dissolved in 10 mM sodium acetate and 150 mM sodium chloride buffer, adjusted to pH 5.0, and then they were frozen and lyophilized. The dry samples were dissolved in $^{2}H_2O$ and then transferred into the NMR tube. The H/H^2 exchange experimental procedures and data analysis were performed as described (*125*).

4.3. Results

4.3.1. Characterization of HP36 N68A and HP36 K70M.

Early Cys scanning studies of the entire villin headpiece domain showed that a K70C mutation lead to an increase in thermostability (*155*). In that study thermostability was measured using CD detected thermal melts. The wavelength used, 222nm, is dominated by the α–helices. The vast majority of the helical structure in villin headpiece is located in the HP36 region, thus we reasoned that this study was probing the thermal stability of the C-terminal helical subdomain. Inspired by this observation we sought to test if mutation of K70 would indeed stabilize the isolated C-terminal subdomain. We have recently analyzed the contribution of the six positively charged residues to the stability to the helical subdomain of headpiece (i.e. HP36). We replaced each charged residue by Met, reasoning that this is a more conservative mutation than a Lys to Cys or Arg to Cys mutation since it preserves the hydrophobic portion of the sidechain. We observed that mutation of three of the Lysines, K48, K65 and K70 lead to an increase in thermal stability of HP36. The HP36 K70M mutant was the most stabilizing among those mutations, and lead to an increase in T_m of 9.8 ^oC. The effect of the other two mutants, K48M and K65M, were smaller, leading to increase in T_m of 5 ^oC and 4.2 ^oC respectively. The K70M mutant also had the largest effect on ΔG^o of unfolding, increasing it by 1.19 kcal mol^{-1} (Table 4.1, Figure 4.2). In the absence of high resolution structural data on the HP36 K70M mutant, it is not clear what is responsible for the increase in stability. Consequently, we solved the structure of the mutant by X-ray crystallography at 1.41 Å resolution (Table 4.2, Figure 4.3). The structure is identical to that of the HP35 N68H (PDB code 1YRF) (*87*) mutant previously studied by the Eaton group. The RMSD for the two structures calculated for residues Leu42 to Leu75 of HP36 K70M with Leu42 to Leu75 of HP35 N68H is 0.32 Å for the backbone. The all atom RMSD, excluding the mutated site is 0.91 Å. The K70 and M70 sidechains in the two proteins adopt the same sidechain conformation with a χ_1 value of 124^o. The fact that the structures are identical strongly supports our conjecture that the K70M mutant exerts its effect by eliminating unfavorable electrostatic interactions rather than by altering core packing.

In this study, we also examined an additional substitution that did not include a charged residue: Asn-68 to Ala. The choice was again based on the early Cys scanning study of intact headpiece (*155*). Those studies reported that replacement of Asn-68 by Cys increased the T_m of domain but no detailed analysis was reported. Asn-68 is a surface residue located in the C-terminal helix of HP36 (Figure 4.1). Asn has a relatively low helix propensity thus we reasoned that its replacement by a residue with a high helical propensity could stabilize the domain (*42, 156*).

At 25 ^oC, the far UV CD spectrum indicates that the N68A mutant has significant α-helix content and the shape and intensity of the spectrum are very similar to that of HP36 (Appendix 1). In the 1D NMR spectrum, characteristic resonances appeared at -0.1 ppm (Val-50) and at 5.6 ppm (Phe-47) due to the ring current shift effect (Appendix 2). These resonances show that the N68A mutant folds to the same native state as wild type. The stability of the mutant was probed using both thermal unfolding experiments and urea denaturation experiments monitored by CD. The

HP36 N68A mutant is more thermostable than wild type HP36 with a T_m of 76.1 ^oC (Table 4.1). The HP36 N68A mutant does not exhibit complete urea unfolding curves and the unfolded baseline is not reached, even at the highest urea concentration (Figure 4.2). The estimated stability of the mutant from directly fitting the curve was 3.95 kcal mol^{-1}. However this value is not the most reliable since the complete transition can not be observed. Alternatively, the stability can be estimated from the midpoint concentration, C_M, which can be accurately determined even from an incomplete unfolding curve while the m-value often can not. However the stability can still be estimated using the measured C_M value and the wild type m-value (0.52±0.01 kcal mol^{-1} M^{-1}). Using this method gives 4.16±0.08 kcal mol^{-1} for the stability of the HP36 N68A mutant (Table 4.1).

4.3.2. N68AK70M a thermal hyperstable double mutant.

The analysis of the single mutants demonstrates that N68A and K70M significantly stabilize the domain. Examination of the structure of the domain suggests that those two mutations could be simultaneously accommodated by the protein. Consequently we prepared the N68AK70M double mutant. At 25 ^oC, the far UV CD spectrum and 1D NMR spectrum indicate that the double mutant adopts the same fold as wild type (Appendix 1 and 2). We solved the crystal structure of the double mutant to 1.79 Å resolution (Table 4.2, Figure 4.3). The structure is identical to the HP36 K70M and HP35 N68H mutants, the RMSD for residue 42 to 75 is 0.13 Å relative to HP36 K70M. The all atom RMSD, excluding the mutated residues is 0.25 Å.

The results of thermal unfolding with the double mutant are impressive. The protein does not completely thermally unfold and the melting profile does not show a post-transition (Figure 4.4A). The T_m for wild type is 73.0 ^oC while the T_m estimated for the double mutant is at least 90.6 ^oC (Table 4.1). We also compared the thermal stability of wild type and N68AK70M in 3 M urea (Figure 4.4B). Under these conditions, the T_m of WT HP36 is 54.8 ^oC while it is 73.9 ^oC for the N68AK70M mutant an increase of 19.1 ^oC (Table 4.1). Urea denaturation studies were attempted but the C_M of the mutant is near 10 M urea making it difficult to accurately measure stability (Figure 4.2). We estimate the stability by using the wild type m-value and C_M for the double mutant calculated from directly fitting the curve or from analysis of the derivative. ΔG^o of unfolding for the double mutant is 1.7 to 1.8 kcal mol^{-1} larger than the value for the wildtype. Guanidine hydrochloride induced unfolding is not a viable option for measuring the stability since electrostatic interactions make a significant contribution to the domain stability and preliminary studies have shown that guanidine hydrochloride and urea denaturation give different estimates of ΔG^o.

We next undertook amide $H/^2H$ exchange studies to probe the slow timescale dynamics of the double mutant and to compare it to wild type. Amide $H/^2H$ exchange experiments were performed for WT HP36 and HP36 N68AK70M at pD 5.0 and 25 ^oC under native conditions. For WT HP36, 21 amide ^1H-^{15}N cross-peaks were detected in the first HSQC spectrum, which was taken 27 min after the dry protein was mixed with ^2H_2O (Appendix 5). Only 2 amide ^1H-^{15}N cross-peaks (Val 50 and

Leu 61) were detected after 6 hours. While for HP36 N68AK70M, 23 amide ^{1}H-^{15}N cross-peaks were detected in the first HSQC spectrum and 7 amide ^{1}H-^{15}N cross-peaks (Phe 58, Ala 59, Leu 61, Lys 65, Asn 66, M70 and K71) were detected even after 17 hours (Appendix 6). A histogram of the apparent free energies of the opening reaction calculated from the observed exchange rates (Table 4.3) for wild type and the N68AK70M mutant is shown in Figure 4.5. The pattern of exchange is very similar for the two proteins arguing that the mutations have not significantly altered the slow dynamics of the domain. In principle, the free energy of unfolding of a protein can be accurately determined by amide exchange provided exchange occurs in the EX2 limit. The analysis requires a known subset of residues that exchange by global unfolding and do not experience local unfolding or conversely, so called super protection due to structure in the unfolded state. A priori it is not always clear which residues exchange strictly by global unfolding and furthermore stabilization of a domain can increase the fraction of residues that exchange by subglobal unfolding events. For these reasons we believe the amide $H/^{2}H$ exchange studies are the best viewed as a probe of slow dynamics. Nevertheless, it is interesting to compare the estimated ΔG^{o} of unfolding determined by $H/^{2}H$ exchange for the two proteins. Using the three most protected residues in wild type HP36 ΔG^{o} is 3.73 kcal mol^{-1} while the value calculated for the double mutant is 4.44 kcal mol^{-1}, an increase of 0.7 Kcal mol^{-1}. For reasons outlined above we believe the ΔG^{o} value determined by the urea denaturation experiment more accurately reflects the increase in stability caused by the double mutations.

4.3.3. The hyperstable mutant is monomeric.

HP36 N68AK70M is much more stable than WT HP36, however it is necessary to check whether this hyperstable mutant is monomeric or not since self association can lead to an apparent increase in stability. Analytical ultracentrifugation experiments show that HP36 N68AK70M is monomeric. Figure 4.6 shows representative sedimentation equilibrium data for HP36 N68AK70M. The data were fit well by an ideal single-species model with a molecular weight within 5% of the monomer molecular weight. The average experimental molecular mass determined from multiple experiments of HP36 N68AK70M is 4285, the expected mass is 4149.9.

4.4. Discussion

Using two simple surface mutations, we have developed a monomeric hyperstable variant of HP36 whose structure is identical to wild type. The mutations were designed to eliminate unfavorable electrostatic interactions (HP36 K70M) and to increase helical propensity (HP36 N68A). The increase in stability for the N68A mutant is very close to that expected based on helical propensity. Our high resolution structure of HP36 K70M is consistent with stabilization by modification of surface electrostatics rather than core repacking. The high resolution structure of the double mutant shows that there is no change in the hydrophobic core and again, is consistent with the design strategy which targeted surface interactions. It is interesting to

compare these results to experiments designed to improve protein stability by altering interactions in the core. In these cases multiple mutations are often required to achieve increases in stability comparable to what is observed here, suggesting that targeting surface residues may be a particular efficient strategy for stabilizing proteins (*63*). It is also noteworthy that the significant increase in T_m is associated with a relatively more modest increase in ΔG^o. This is, in fact, expected for small proteins. Rees and Robertson (*130*) analyzed a data base of larger proteins and derived a simple relationship between $\Delta\Delta G^o$ and ΔT_m, showing that $d\Delta G^o/dT_m$ scales with the number of residues, N. Thus the same increase in ΔG^o will lead to a large increase in T_m for small proteins and a small increase in T_m for large proteins. $d\Delta G^o/dT_m$ is predicted to be $N*0.0023$ Kcal.mol^{-1}K^{-1} at the T_m of the wild type (346 ^{o}K). Thus a 19 ^{o} increase in T_m is predicted to be linked to a 1.6 Kcal.mol^{-1} increase in ΔG^o which is in good agreement with the value estimated from analysis of the urea denaturation experiment. The observation that such simple mutations can drastically increase protein stability provides direct evidence that the stability of miniature protein domains has not been evolutionarily optimized; instead like large "normal" domains, their stability can be increased by optimizing interactions and removing unfavorable interactions. In addition, the fact that the relationship between the increase in T_m and $\Delta\Delta G^o$ is well predicted by the properties of large proteins provides future evidence that "miniature" proteins do not require specialized interactions to fold.

The modest stability of wild type HP36 has hampered mutational studies (*21*). The use of HP36 N68AK70M as a "pseudo-wild type" structure will allow more drastic mutations in the hydrophobic core than are tolerated in the normal wild type protein and will facilitate experimental studies of this interesting protein.

Finally it is interesting to contemplate the consequences of these mutations in the larger HP67 construct. The increase in stability caused by the K70M is not expected to propagate to the N-ternimal subdomain since K70 together with E39 form a buried salt bridge that links the two subdomains (*135*). Removal of the ε-amino group of K70 will leave the buried carboxylate E39 uncompenstated by a salt bridge partner and is thus expected to destabilize the N-terminal subdomain. Sequence analysis is often used to choose potential sites for mutation in protein engineering studies and can be used to design stabilizing substitutions in favorable cases (*157*). However analysis of aligned sequences of HP36 would fail to predicted that K70 is viable site for mutation since it is rigorously conserved in all HP36 sequences (*81*). The conservation arises because of the need to form the E39 K70 salt bridge and not because K70 is needed to directly stabilize the HP36 fold. It is interesting to note that virtually all characterized "miniature" proteins are domains or subdomains of large sequences, thus if the goal of a protein engineering exercise is to stabilize the protein, sequence alignments may not provide a useful guide since residues may be conserved for functional reasons in the intact protein, even if they make a unfavorable contribution to the stability of the smaller domain. K70 in HP36 provides a striking example of this.

Table 4.1: Summary of equilibrium stability measurement for wild type HP36 and its mutants. The quoted uncertainties represent the standard error to the fit. The uncertainty in T_m as estimated by repeated measurements of wild type is 1.5 ^oC. The estimated uncertainties in ΔG^o and the m values are on the order of 0.12 kcal mol^{-1} and 0.01 kcal mol^{-1} M^{-1} respectively as based on repeated measurements of wild type. Thermal unfolding studies were conducted at protein concentrations of 20.4 uM (Wild type HP36), 26 uM (HP36 K70M), 27 uM (HP36 N68A) and 28.7 uM (HP36 N68AK70M).

Protein	**T_m (^oC)**	**T_m in 3M urea (^oC)**	**ΔG^o (kcal mol^{-1})**	**m value (kcal mol^{-1} M^{-1})**
WT HP36	73.0±1.5	54.8	3.22±0.12	0.52±0.01
N68A	76.1	N/A	4.16±0.08[c]	0.52±0.01
K70M	82.2	N/A	4.41±0.09[c]	0.52±0.01
N68AK70M	90.6[a]，90[b]	73.9	4.94-5.06[d]	0.52±0.01

[a]Estimated value by directly fitting the thermal unfolding curve;
[b]Estimated value by taking the derivative of the curve;
[c]Estimated using the wild type m-value (0.52±0.01 kcal mol^{-1} M^{-1}) and the measured C_M value;
[d]Estimated value by using the wild type m-value and C_M determined either by directly fitting the curve or by taking the derivative of the curve.

Table 4.2: X-ray data

	HP36 K70M	**HP36 N68AK70M**
Wavelength (Å)	1.1	1.0
Resolution limits (Å)	50-1.41	50-1.79
Number of reflections	5,976	3,020
Completeness	0.942 (0.578)	0.942 (0.701)
Mean redundancy	3.9 (1.5)	5.9 (3.8)
R_{sym}[a]	0.062 (0.176)	0.038 (0.176)
$\langle I/\sigma I \rangle$[b]	21.1 (6.4)	34.5 (6.6)
Space group	P41	P41
Unit cell dimensions a, b, c (Å)	32.2, 32.2, 31.7	32.4, 32.4, 31.7

[a] $R_{sym} = \sum_{hkl}\sum_i|I_i - \langle I \rangle|/\sum_{hkl}\sum_i I_i$, where I_i is the i^{th} measurement and $\langle I \rangle$ is the weighted mean of all measurements of I.

[b] $\langle I/\sigma I \rangle$ indicates the average of the intensity divided by its standard deviation. Numbers in parentheses refer to the respective highest resolution data shell in each data set.

Table 4.3: Measured amide exchange rates for wild type HP36 and the N68AK70M double mutant.

Residue	$K_{observed}$ (min^{-1})	
	Wild type HP36	**HP36 N68AK70M**
47	0.130	0.051
48	0.030	0.029
49	0.031	0.019
50	0.008	0.006
51	0.009	0.004
52	0.034	0.035
53	0.021	0.007
54	0.038	0.021
57	0.105	0.018
58	0.013	0.004
59	0.013	0.004
60	0.023	0.011
61	0.003	0.018
65	0.010	0.005
66	0.014	0.002
67	0.029	0.005
68	0.065	0.006
69	0.019	0.077
70	0.029	0.002
71	0.068	0.011
72	----	0.023
75	----	0.071
76	0.081	0.072

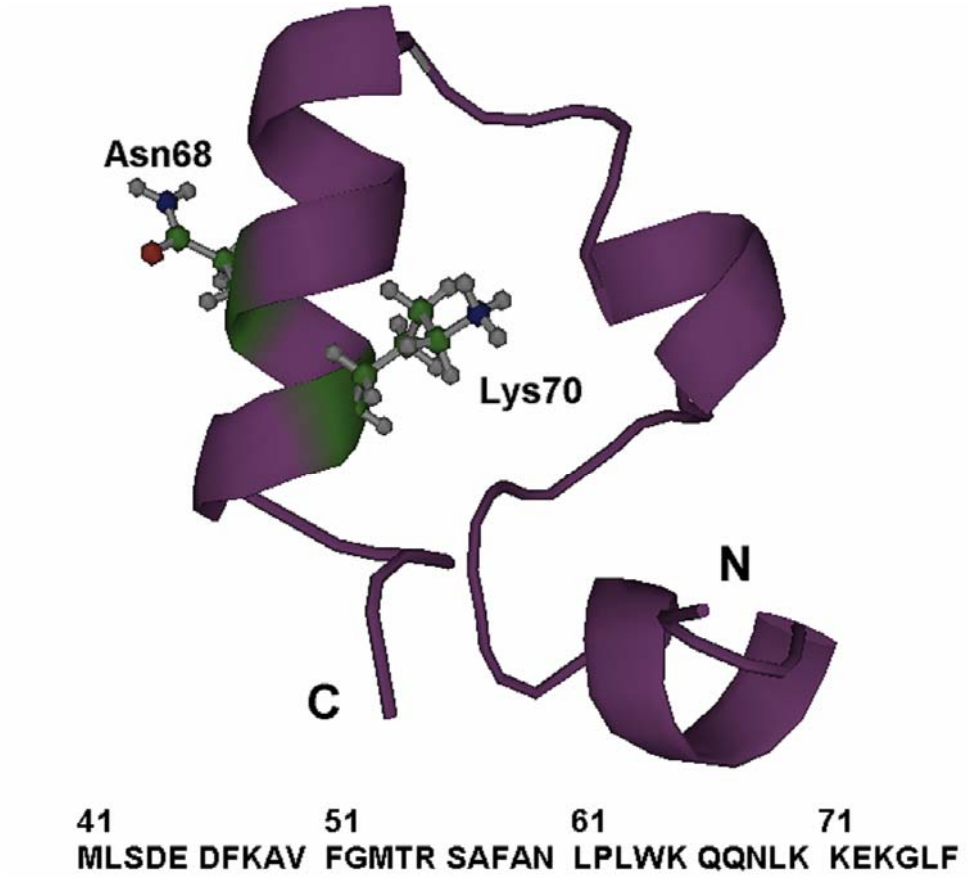


Figure 4.1: Ribbon diagram and primary sequence of the villin headpiece subdomain (HP36). The side chains of Asn68 and Lys70 are shown in ball-and-stick representation created by Pymol v0.99.(*83*)

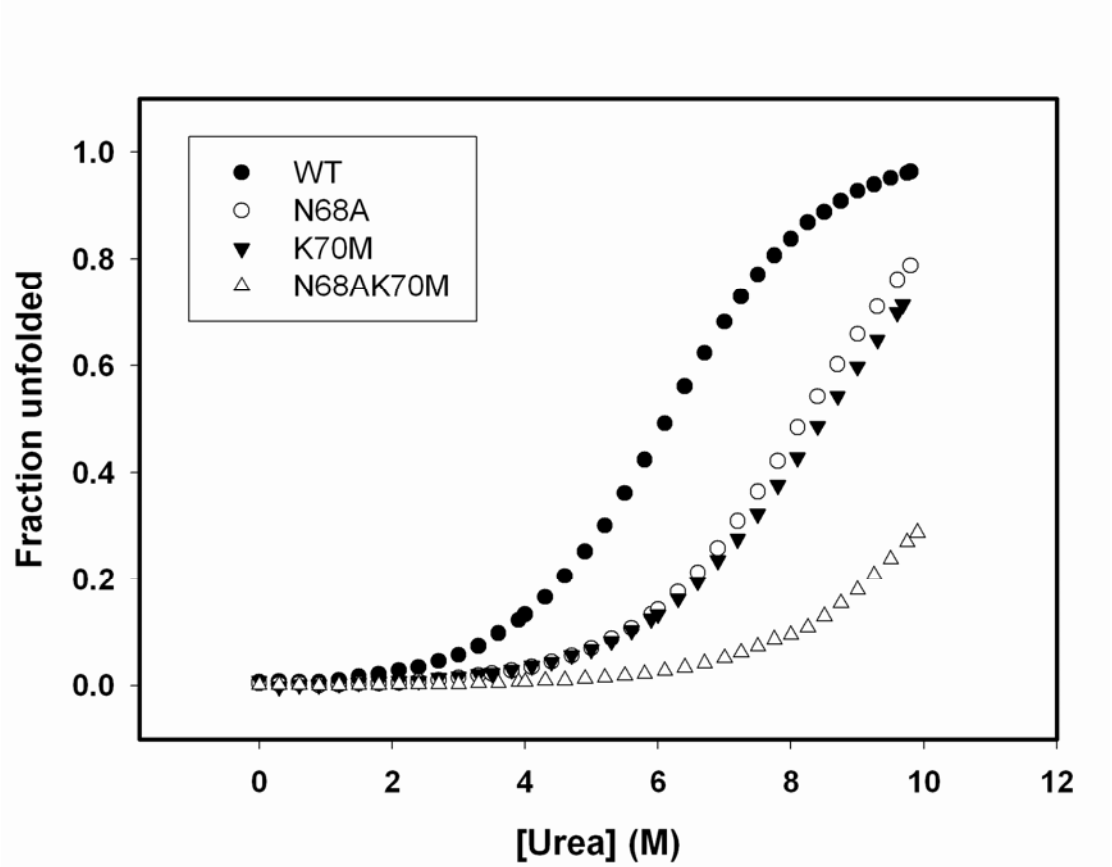


Figure 4.2: Urea denaturation of HP36 (●), N68A (○), K70M (▼) and N68AK70M (△). Measurements were made at 25 ^{o}C, pH 5.0, in 10 mM sodium acetate buffer. Data is plotted as fraction unfolded.

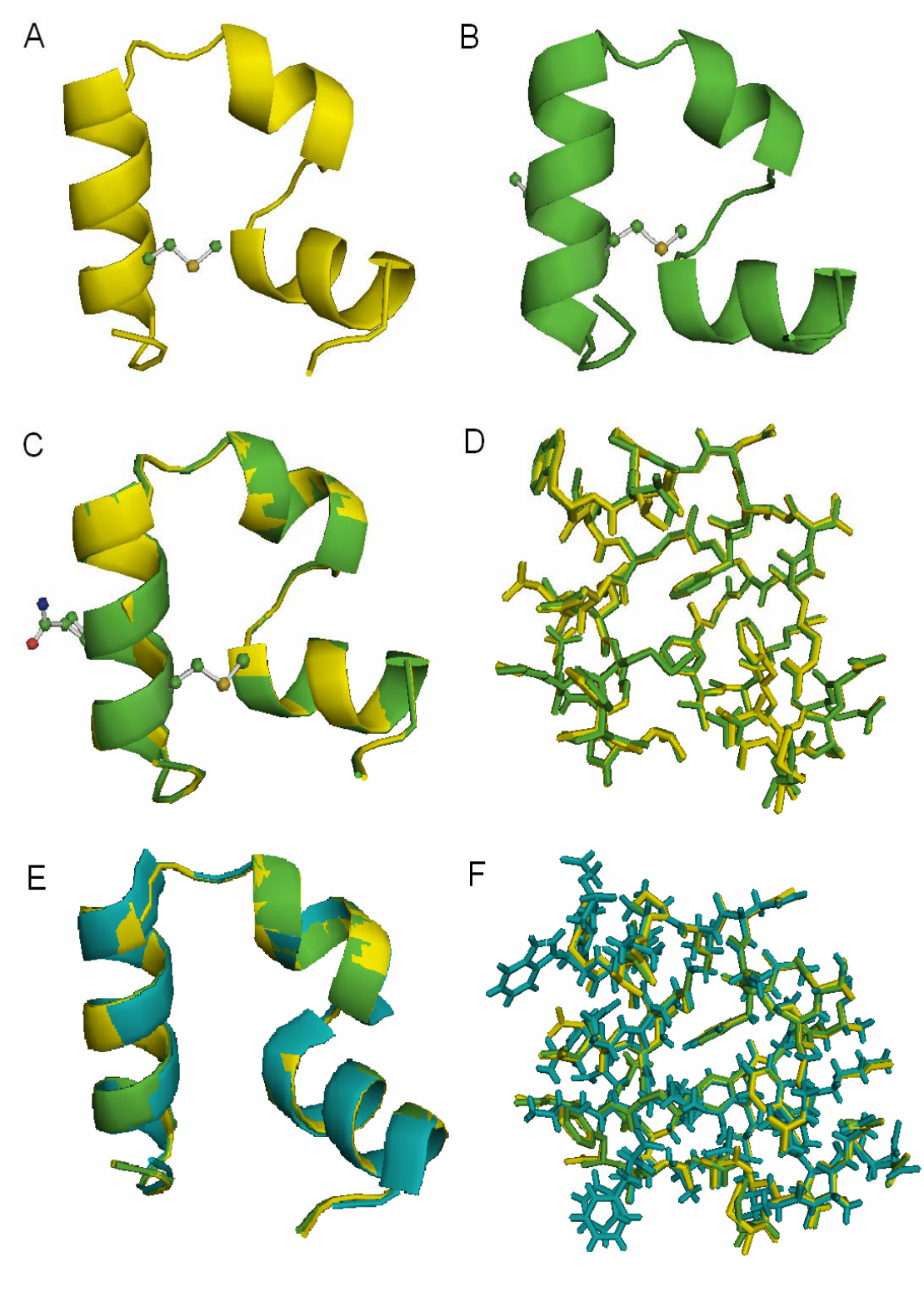


Figure 4.3: (A) Ribbon diagram of the crystal structure of HP36 K70M (yellow) and (B) HP36 N68AK70M (green), the side chains of residue 68 and 70 are shown in ball-and-stick representation; (C) Comparison of the X-ray structures of HP36 K70M

(yellow) and HP36 N68AK70M (green) shown in ribbon diagram representation and (D) All-bonds representation; (E) Comparison of the X-ray structure of HP35 N68H (1YRF, Cyan) and X-ray structures of HP36 K70M (yellow) and HP36 N68AK70M (green) shown in ribbon diagram representation and (F) All-bonds representation. The Figures were made using Pymol v0.99 (*83*). The X-ray strycture of HP35 shows two orientations for the Phe76 and Trp64 sidechains. Both are shown in the all bonds representation in panel F.

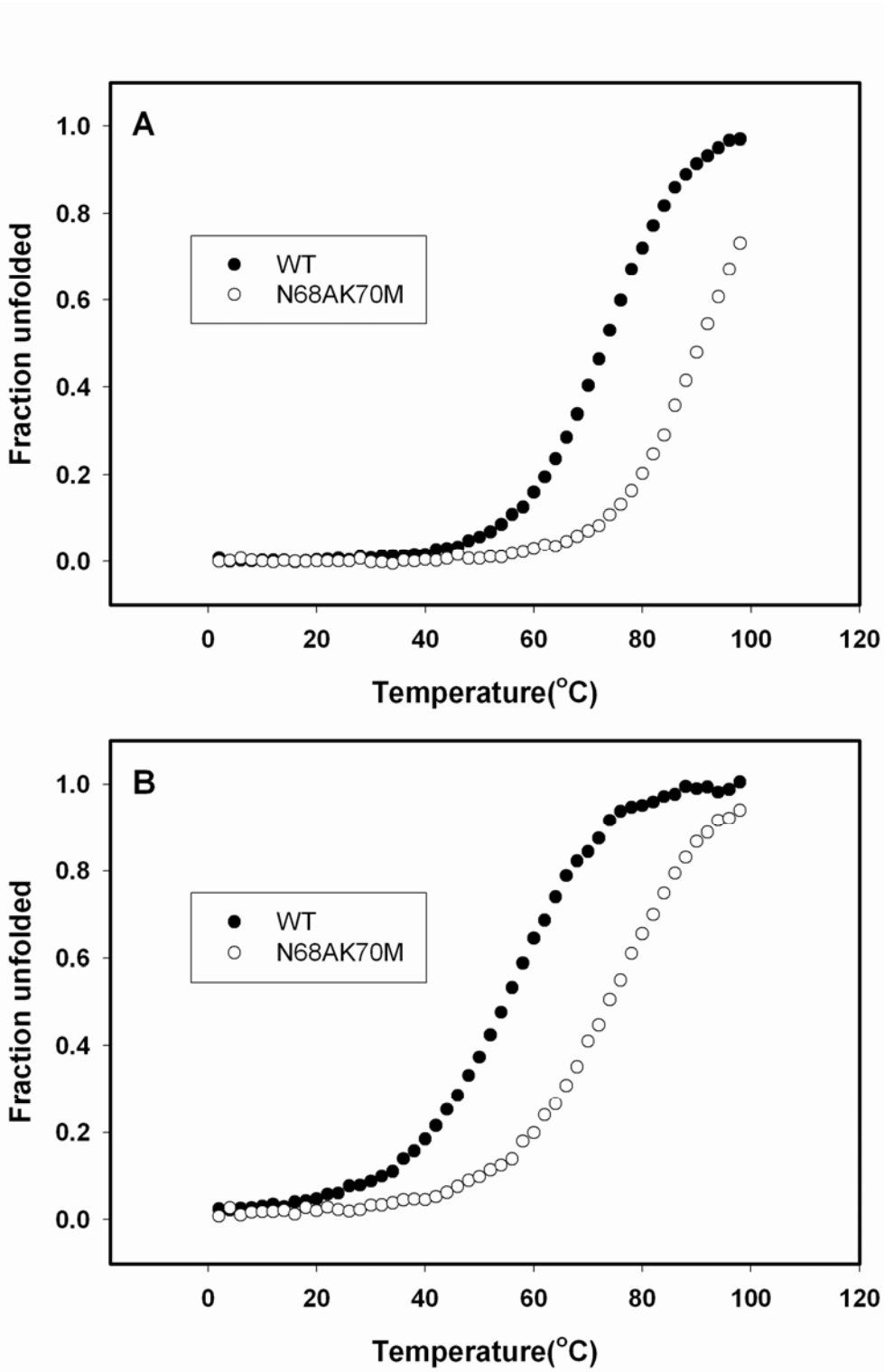


Figure 4.4: Thermal unfolding of HP36 (●) and N68AK70M (○). (A) With 0 M urea; (B) With 3M urea. All experiments were conducted in 10 mM sodium acetate buffer, at pH 5.0. Data is plotted as fraction unfolded.

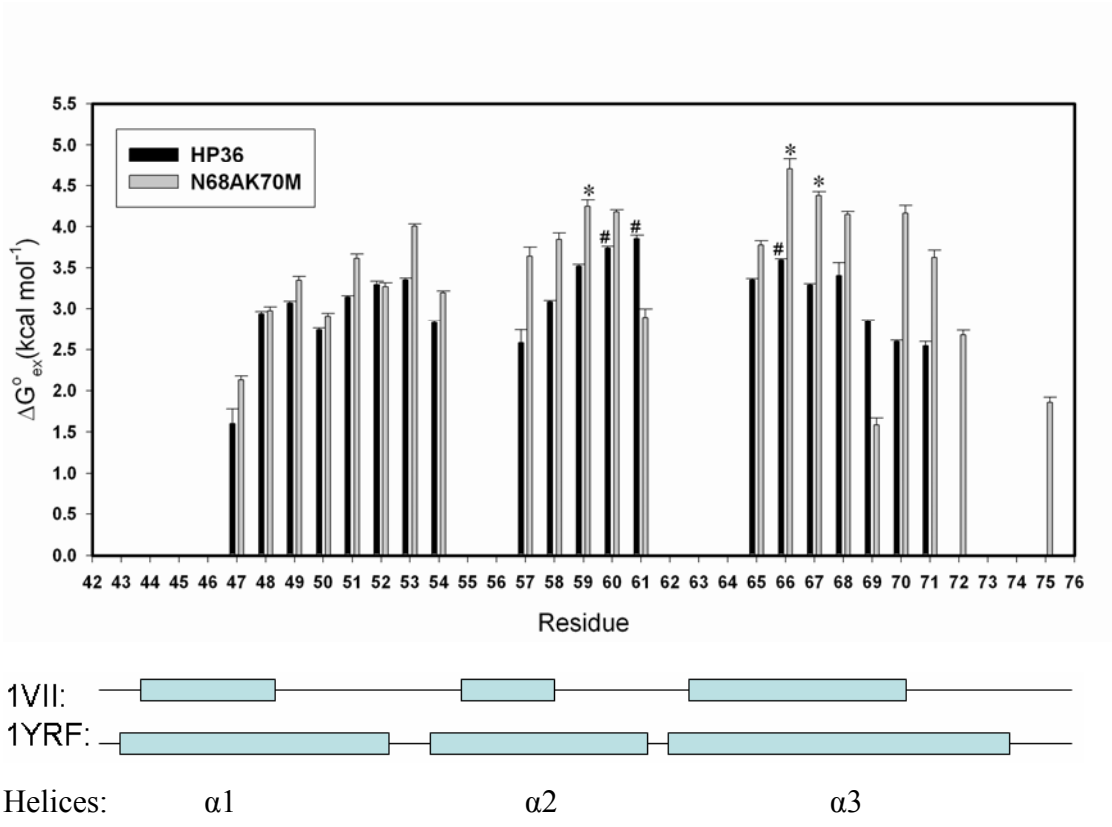


Figure 4.5: Histogram of the apparent free energy change for the opening reaction, ΔG^{o}_{ex} at 25 ^{o}C *versus* residue number for WT HP36 (Black bars) and N68AK70M (Grey bars). Samples were prepared in 10 mM sodium acetate buffer, pD 5.0, 25 ^{o}C. Error bars are shown. The three residues with the largest protection factor are labeled (#) for HP36 and (*) for HP36 N68AK70M. The blue bars represent the helices: 1VII is the NMR structure *(84)* and 1YRF is the x-ray structure *(87)*.

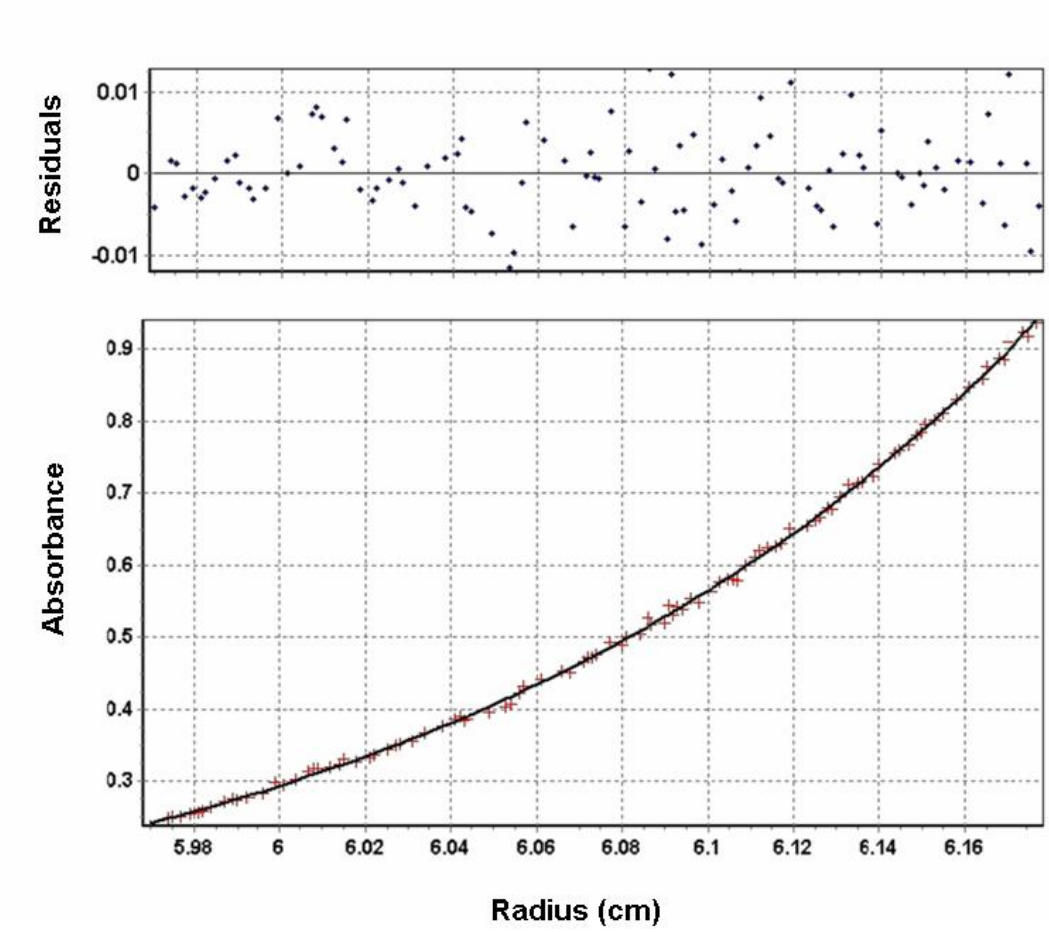


Figure 4.6: Sedimentation equilibrium studies of N68AK70M. The sample contained 100 uM N68AK70M in 10 mM sodium acetate buffer solution, pH 5.0, 25 ^{o}C. The rotor speed was 48,000 rpm. The observed molecular mass was 4278 determined from a single species fit; the expected mass is 4149.9.

5. 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

Abstract

The 36 residue helical subdomain of the villin headpiece, HP36, is one of the smallest cooperatively folded proteins, folding on the microsecond timescale. The domain is an extraordinarily popular model system for both experimental and computational studies of protein folding. The structure of HP36 has been solved using X-ray crystallography and NMR spectroscopy, with the resulting structures exhibiting differences in helix packing, van der Waals contacts and hydrogen bonding. It is important to determine the solution structure of HP36 with as much accuracy as possible since this structure is widely used as a reference for simulations and experiments. The results of the experiments are combined with the data from all-atom molecular dynamics simulations with explicit solvent to evaluate which of the experimental models is the better representation of HP36 in solution. The simulations were conducted by Ms Wickstrom and are discussed here because this was a joint project. After 50 ns of simulation initiated with the NMR structure, we observed that the protein spontaneously adopts structures with a backbone conformation, core packing and C-capping motif on the third helix that are more consistent with the crystal structure. We also examined hydrogen bonding and sidechain packing interactions between D44 and R55 and between F47 and R55 respectively, which were observed in the crystal structure but not present in the NMR-based solution structure. Experimental double mutant cycles confirmed that the F47/R55 pair has a larger energetic coupling than the D44/R55 interaction. Simulations showed large fluctuations in the distance between D44 and R55, while the distance between F47 and R55 remained stable, suggesting the formation of a cation-pi interaction between those residues. Overall, these combined experimental and computational studies show that the X-ray crystal structure is the better reference structure for HP36 in solution at neutral pH. Our analysis also shows how detailed molecular dynamics simulations can help bridge the gap between NMR and crystallographic methods.

Acknowledgments

The material presented in this chapter has been published (Lauren Wickstrom*, Yuan Bi*, Viktor Hornak, Daniel P. Raleigh and Carlos L. Simmerling 2007 *Biochemisty 46*, 3624-3634, "*" represents joint first authors). This chapter contains direct excerpts from the manuscript written by Yuan Bi and Lauren Wickstrom with suggestions and revisions from Professor Daniel P. Raleigh and Professor Carlos L. Simmerling. Ms Lauren

Wickstrom conducted the computational work. I thank Prof. Mcknight for his helpful discussions.

5.1. Introduction

The villin headpiece helical subdomain (HP36), the C-terminal portion of the villin headpiece, is the shortest naturally occurring sequence which has been shown to fold cooperatively (Figure 5.1). Infrared temperature jump (*154*), laser fluorescence (*158, 159*) and NMR lineshape analysis (*86*) techniques have measured the folding of HP36 to occur on the microsecond time scale. Its rapid folding, small size and simple topology of three helices have made this domain an extremely popular system for experimental (*21, 84-89, 110, 154, 158-161*) and computational studies (*90, 91, 94, 95, 97, 111-113, 136, 140, 141, 147, 148, 150, 162-164*). Much of this work relies on using the folded structure as a reference and thus the accuracy of the known HP36 structure is of particular importance.

Several structures have been solved for HP36, one by NMR and the others by X-ray crystallographic methods (*84, 87*). These structures vary in the hydrophobic core packing, interhelical H-bonds and in the length of the helices. In addition, two potentially important sidechain contacts differ significantly between the NMR and X-ray structures: F47/R55 (4.3 Å (X-ray) and 6.3 Å (NMR)) and D44/R55 (2.7 Å (X-ray) and 7.9 Å (NMR)) (Figure 5.2A and 5.2B**)**. In the X-ray structure, the F47/R55 pair forms a van der Waals contact which could be particularly stabilizing as a cation-pi interaction, while D44/R55 form a hydrogen bond (D44-Oδ1 and R55-Nε). Neither contact is present in the NMR structure. These differences may arise from changes in the HP36 sequence used in the two sets of experiments, although this seems unlikely. The crystallographic study employed the N68H mutant of HP36 and also lacks the N-terminal methionine incorporated by the expression system used for the NMR study (note that we adopt the typical numbering convention(*84, 85*) for HP36, in which L42 follows the N-terminal methionine). Another possible reason for the structural differences could be the variation of experimental conditions such as pH or temperature. There was significant deviation in the pH between structural determinations; the NMR structure was solved at pH 3.7 in contrast to the more neutral conditions of the crystallography experiment (pH 6.7). An alternate explanation for the observed structural differences is that they arise from methodological limitations conditions; these frequently give rise to differences in structures of the same protein solved using different techniques. In general, NMR structures are less precise than X-ray structures, particularly if only homonuclear methods are used. Nevertheless, X-ray structures can suffer from effects due to crystal packing; the resulting contacts may have a local influence on conformational preferences. The small size of HP36 and its correspondingly large surface area to volume ratio could make crystal contacts play an important role. On the other hand, crystallographic data is often collected at low temperatures which might result in the dampening of thermal motions that are present under physiological conditions.

Many computational studies have used HP36 as a model system for development

and validation of protein folding methods and for optimization of force field parameters (*90, 91, 94, 95, 97, 111-113, 136, 140, 141, 147, 148, 150, 162-164*). If the native reference structure is not correct, the basis of these studies may not be valid. For example, the structure of the first helix and the C-terminus vary in the ensemble of NMR structures (*84, 140*) and many MD studies have therefore neglected these regions of the experimental structure when evaluating their success. Nevertheless, most simulations are performed at neutral pH and thus it is not clear if the simulations should be compared to the NMR structure from pH 3.7. A better structural model for neutral conditions would be invaluable for further work in understanding the folding and stability of this important model system for protein folding.

Accurate computational studies can provide an alternate method to study conformational behavior and alleviate the uncertainty about which structure is the better representation of the folded state in solution. In principle, molecular dynamics (MD) simulations can supply detailed information with spatial and time resolution that exceed the ability of NMR and X-ray experiments, providing insight into the role of specific interactions that may not be readily accessible through experiments that probe averages over rapidly interconverting ensembles.

Here, we conducted all-atom MD simulation in explicit solvent using the NMR structure of HP36 in order to gain insight into the details of the folded state in solution. The simulation diverges from the initial NMR structure and spontaneously adopts a structure with much greater similarity to the X-ray structure, arguing that the X-ray structure is a more accurate representation of the structure in solution at neutral pH. In addition, two residue pairs, D44/R55 and F47/R55, spontaneously formed contacts during the simulation, with the F47/R55 pair appearing to be more stable. These interactions were reported in the crystal structure but were not present in the ensemble of structures generated by the NMR studies. Thus we conclude that the F47/R55 may play an important role in stabilizing HP36 in solution. We acknowledge that simulation models can be limited in accuracy and any predictions should be tested through direct experimentation. In order to validate our computational observations, we employed an experimental double mutant cycle analysis. The results are consistent with our simulation data, and suggest that the interaction between F47 and R55 plays a role in stabilizing the native state through a cation-pi interaction. Overall, the results show how properly validated MD simulations can provide an avenue to test the stability and validity of structural models that were derived from experimental data.

5.2. Materials and methods

5.2.1. Computational.

The numbering system corresponds to that used for the full length villin headpiece, with the sequence M41–F76 (MLSDE DFKAV FGMTR SAFAN LPLWK QQNLK KEKGLF). HP36 has free N and C-termini that were modeled in the charged state. This sequence and termini correspond to those used in the experimental studies. All sidechains for Asp, Glu, Lys, and Arg were charged during the simulation. All calculations employed Amber version 8 (*165*) and used the ff99SB modification (*166*) of the Amber ff99 force field (*167, 168*). SHAKE (*169*) was used to constrain bonds

involving hydrogen. The time step was 2 fs. The temperature was maintained using the weak coupling algorithm (*170*) with a thermostat of 37 °C (310 K) and the pressure was equilibrated to 1 atm. All production simulations were performed using the NVT ensemble. An independent simulation using the NPT ensemble provided similar results (Data not shown).

Solvation plays a key role in biomolecular structural preferences and thus accurate treatment of solvation is essential for the investigation of structural propensities in simulations. Explicit solvent models can be highly effective, particularly when water has non-bulk properties and interacts directly with the solute (*171*). Implicit models such as the semi-analytical Generalized Born model (GB)(*172*) are attractive because they are computationally less expensive and can converge more rapidly than simulations in explicit water due to lack of solvent viscosity. While GB has been widely used for protein folding studies by a number of groups, other investigators have reported poor results including secondary structural bias and ion pairing issues (*173-175*). Our previous studies on fragments of HP36 have shown that the use of explicit water produced results which were much more consistent with experimental trends than those obtained with implicit solvent (*90*). Consequently, we used explicit solvent in our simulations of HP36, in a truncated octahedral box using periodic boundary conditions with Particle Mesh Ewald (PME) (*176*) and a direct space cutoff of 8 Å. In order to investigate the influence of long-range periodicity, two additional simulations were run: one with the Isotropic Periodic Sum (IPS) (*177*) non-lattice method with a cutoff of 8 Å, and another with an atom-based nonbonded cutoff of 12 Å with no smoothing function. Simulations were initiated from the NMR structure (PDB ID 1VII) surrounded by 2327 TIP3P (*178*) waters molecules and equilibrated at 310 K for 50 ps with harmonic restraints on solute atoms, followed by minimization with gradually reduced positional restraints and three 5 ps MD simulations with gradually reduced restraints. The production simulations of the NMR structure were 50 ns in length for two PME simulations with different random seeds for assignment of velocities, and 30 ns for the IPS and cutoff simulations respectively. As a control, the X-ray structure (PDB ID 1YRF) was setup with the same amount of waters and equilibrated in a similar fashion. This simulation was run for 30 ns using PME.

5.2.2. Data analysis.

The last 5 ns of the simulation were used for cluster analysis and DSSP calculations. Cluster analysis was performed with Moil-view using all atoms as a similarity criterion with average linkage. Clusters were formed with the bottom-up approach using a similarity cutoff of 2.5 Å. DSSP analysis and calculation of distances, RMSD values, and radius of gyration were done using the ptraj module in Amber. Distances between sidechains were calculated using selected heavy atoms as indicated in the text. PMFs were calculated according to equation (5.1):

$$\Delta G = -RT \ln (N_i/N_0) \quad (5.1)$$

where N_i is the population of a particular histogram bin along the reaction coordinates that were employed and N_0 is the most populated bin. Error bars were estimated for

the PMF by averaging two independent simulations and subsequently subtracting the PMF of an individual simulation from the average PMF.

5.2.3. Cloning, expression and purification.

The plasmid (pET3a-NTL9-FXa-HP36) containing the gene for HP36 was prepared as described previously (*120, 179*). The primers for generation of the mutants D44N, F47L and R55M were from Operon. The fusion proteins with coded amino acid substitutions were expressed and purified as described previously (*120*). For the purification of the fusion protein of F47L and F47LR55M, ethylene diamine tetraacetic acid and phenylmethyl sulfonyl fluoride were added to final concentration of 1 mM to prevent degradation by proteases during ion exchange chromatography. Cleavage of WT HP36, D44N, F47L, R55M, D44NR55M and F47LR55M from fusion proteins were carried out by incubating with Factor Xa in buffer (50 mM Tris-HCl, 100 mM NaCl and 5 mM $CaCl_2$, pH 8.0) for 16 hours. The conditions were as follows: WT HP36, D44N, R55M and D44NR55M, 2 units of Factor Xa at 23 ^{o}C; F47L, 1 unit of Factor Xa at 4 ^{o}C; F47LR55M, 0.5 units of Factor Xa at 4 ^{o}C. The solutions were purified by RP-HPLC as described. All proteins were more than 95% pure. The identities of the proteins were analyzed by electro-spray ionization mass spectrometry (ESI MS). The expected and observed molecular weights were as follows: WT HP36, expected 4190.9, observed 4190.1; D44N, expected 4188.9, observed 4189.0; F47L, expected 4155.8, observed 4157.0; R55M, expected 4164.9, observed 4166.4; D44NR55M, expected 4163.9, observed 4164.3; F47LR55M, expected 4130.8, observed 4130.2.

5.2.4. Circular Dichroism (CD) Spectroscopy.

All Circular Dichroism (CD) experiments were carried out on an Aviv 62A DS and a 202SF Circular Dichroism spectrophotometer. All samples for CD experiments were prepared in 10 mM sodium acetate, 150 mM sodium chloride buffer solution at pH 5.0. The protein concentrations were determined from absorbance measurements using the method of Pace & coworkers (*122*). The absorbance was measured at 280 nm in 6M guanidine hydrochloride, 20 mM sodium phosphate pH 6.5, with an extinction coefficient of 5690 $M^{-1}cm^{-1}$. Thermal unfolding and urea denaturation experiments were carried out in a 1 cm quartz cuvette by monitoring the signal at 222 nm. Thermal unfolding experiments were performed from 2 ^{o}C to 98 ^{o}C with a 2 ^{o}C interval. The concentrations of urea were measured by measuring the refractive index. The concentration of urea was increased from 0 M to 10 M by ~0.25 M each step. Urea denaturation experiments were performed at 25 ^{o}C. Urea denaturation experiments in the osmolyte trimethylamine *N*-oxide (TMAO) for F47L and F47LR55M were also performed at 25 ^{o}C. The concentrations of TMAO were measured by measuring the refractive index (*180*).

All thermal and urea denaturations were analyzed by a non-linear least squares curve fitting equation using SigmaPlot (SPSS Inc.):

$$y(T) = \frac{a_n + b_n T + (a_d + b_d T) e^{-\Delta G^{\circ}_U(T)/RT}}{1 + e^{-\Delta G^{\circ}_U(T)/RT}} \tag{5.2}$$

$$y([urea]) = \frac{a_n + b_n[urea] + (a_d + b_d[urea])e^{-\Delta G°_U([urea])/RT}}{1 + e^{-\Delta G°_U([urea])/RT}} \quad (5.3)$$

The free energy of unfolding is assumed to be linearly proportional to the concentration of denaturant:

$$\Delta G^o{}_U([urea]) = \Delta G^o{}_U(H_2O) - m[urea] \quad (5.4)$$

Where $\Delta G^o{}_U([urea])$ is the apparent free energy for the native to denatured transition; $\Delta G^o{}_U(H_2O)$ is the free energy of unfolding in the absence of denaturant; m is the slope, equation (5.3) shows the linear dependence of $\Delta G^o{}_U([urea])$ on the denaturant concentration and the m-value reveals quantitative information about the buried surface area change upon unfolding; T is the temperature; R is the gas constant; a_n and b_n represents the intercept and slope of the pre-transition; a_d and b_d represent intercept and slope of the post-transition. The fraction of unfolded state is:

$$Pd = \frac{y(T) - (a_n + b_n T)}{(a_d + b_d T) - (a_n + b_n T)} \quad (5.5)$$

$$Pd = \frac{y([urea]) - (a_n + b_n[urea])}{(a_d + b_d[urea]) - (a_n + b_n[urea])} \quad (5.6)$$

5.3. Results

5.3.1. Simulations of the NMR structure.

Figure 5.3 shows the backbone RMSD *versus* time and RMSD distributions calculated during the last 5 ns for selected regions of HP36 during the simulation. The RMSD is shown relative to both the NMR and X-ray structures. At the end of the equilibration period, the backbone RMSD (residues L42-L75) to each experimental structure was ~ 2.0 Å (Figure 5.3A). At 8ns, a structural transition occurred causing the overall backbone RMSD (X-ray) to drop 1.0 Å below the RMSD (NMR). This greater similarity to the X-ray structure persisted throughout the remainder of the simulation.

In Figure 5.3B, the RMSD relative to the X-ray structure of the region containing helix-3 (residues P62-F76) demonstrates even more clearly a switch during the simulation from similarity to the initial NMR structure to a greater similarity to the X-ray structure, as indicated by a reduction in the RMSD to the X-ray structure from 3 - 4 Å to 0.5 – 1.0 Å. Clearly, the simulation shows the inclination of HP36 to sample structures with a backbone similar to the X-ray structure despite being initiated with the NMR solution structure. The RMSD values for the two other helices remained stable and also showed a clear preference for the X-ray structure (Figure 5.11A,B).

In order to investigate the source of the large reduction in RMSD relative to the X-ray structure, a best fit alignment was performed on residues 61 to 74 to compare the differences before and after the structural transition. In Figure 5.4A, the NMR, X-ray and simulation structures are shown. The conformations of the C-terminus differ significantly between the X-ray and the NMR structure. The simulation structure spontaneously converts from the conformation in the NMR structure to that in the X-ray structure, concomitant with formation of three hydrogen bonds that stabilize the observed conformation. G74 forms a C-capping interaction with K70 and

K71 at the end of helix-3, along with an additional hydrogen bond formed between K70 and L75. Figure 5.4B shows the time evolution of these hydrogen bond distances. In the beginning of the simulation, all three distances are 4 - 9 Å. At 8 ns, the distances are reduced to 2 - 3 Å, indicating formation of the hydrogen bonds that may play an important role in stabilizing the C-terminal helix. Importantly, all three hydrogen bonds are present in the X-ray structure but absent in the NMR structure (Figure 5.4A).

Dictionary of secondary structural prediction (DSSP) (*181*) analysis was employed to characterize the secondary structure in the simulation in order to facilitate comparisons with the X-ray and NMR structures (Figure 5.5). In the simulations, helix-1 spans the same 8 residues as found in the X-ray structure (D44 to F51), while the NMR structure contained only a five residue helix from D44-K48. Thus the simulation significantly extends the length of the first helix, in agreement with the X-ray structure. Overall, the locations of the sequence of helices 2 and 3 are similar in the NMR and X-ray structures, although helix-2 is one residue shorter in the NMR structure, (residues R55 to F58 for the NMR vs. R55 to A59 for the X-ray). In the simulation, helix-2 appears consistent with both experimental structures; full α-helical content is sampled for residues 55 through 58, with partial helical content (~50%) observed for A59. This may indicate that the C-terminus of the longer helix in the X-ray structure frays at the temperature of the NMR experiment. In both the NMR and X-ray structures, the α-helical content is the same for helix-3 (L63-K72). The simulations sample the same helix, with residue K73 sampling a partial population of helical structures. As noted above, the simulation spontaneously adopts a C-capping motif for this helix that is present in the X-ray structure. Overall, the alpha helical structural content of the structures in the simulation is in much better agreement with the X-ray structure, particularly in helix-1.

All-atom cluster analysis was used to generate a representative simulation structure using the last 5 nanoseconds of the trajectory. This structure has backbone and all-atom RMSD values relative to the X-ray structure of 1.5 and 2.7 Å (residues 42 to 75), while the RMSD values relative to the initial NMR structure were higher (2.3 Å (backbone) and 3.3 Å (all-atom)). Figure 5.6 shows all three structures after best-fit of the backbone from residues 42 to 62 (helices 1 and 2). Notably, the X-ray and simulation structure have a very similar spatial arrangement of their phenylalanine cores. In contrast to the X-ray and simulation structures, the NMR structure has F51 shifted more into the core. Thus, the backbone and core of the protein in the solution simulation possesses structural features that are much more similar to the X-ray structure despite being initiated from the NMR structure.

Structural similarities to the NMR Family

Given the diversity among the family of structures solved using the NMR data, it is reasonable to expect that some of them may be more similar than others to the X-ray structure. Figure 5.12 shows the backbone RMSD as compared to the X-ray, simulation and NMR average structures for each structure in the NMR family. Overall, the individual NMR structures are all more similar to the NMR average than to the X-ray structure (average RMSD values of 1.7 and 2.4 Å respectively). The RMSD of

the three individual helices demonstrate similar differences. However, some of the individual members of the NMR family are similar to the X-ray and simulation structures, especially in helix 1. According the DSSP, 7 out of the 29 members of the NMR family sample alpha helical conformations at V50 (data not shown) which is outside of the helical region in the average NMR structure. This suggests that extension of helix-1 beyond the range seen in the average structure remains consistent with the NMR family. However, the overall backbone of the X-ray and the simulation structure differs from all of the structures in the NMR family (Figure 5.12).

5.3.2. Specific sidechain interactions.

There are several specific sidechain interactions which differ in the NMR and X-ray structures. In the X-ray structure, R55 forms a van der Waals interaction with F47 and an interhelical sidechain-sidechain hydrogen bond with D44 (D44-Oδ1 and R55-Nε); both interactions are absent in the NMR structure. In Figure 5.7A and 5.7B, the simulation structure was aligned with the X-ray structure to highlight the similarities in the interaction of those particular sidechains. Since the simulation structure is a single snapshot, we also investigated the behavior of these contacts as a function of time during the MD run, observing fluctuations in both cases (Figure 5.8). In both the X-ray and the simulation structure, the H-bond distance between D44 and R55 is 2.7 Å, in contrast with the much longer distance of 7.9 Å in the NMR structure. This specific contact also samples a range of distances from 6.7 Å to 11.6 Å in the family of NMR structures (Figure 5.13). During the simulation, this hydrogen bond is broken and re-formed multiple times, suggesting that a reasonable description of the equilibrium distance distribution has been sampled (Figure 5.8A**)**. We used histogram analysis to calculate the potential mean force (PMF) for the pair to quantify the stability of the contact in the native state. While two free energy minima are located at the hydrogen bonding distance, two other local minima at 5.0 and 7.0 Å have relative energies of less than 0.6 kcal/mol compared to the contact pair (Figure 5.8B). Thus breaking this contact is expected to be a readily accessible thermal fluctuation. The stability of the contact between F47 and R55 was evaluated by measuring the distance from the Cγ of F47 to the Nε of R55 (Figure 5.8C). This distance had comparable values in the simulation and X-ray structures (4.7 and 4.3 Å, respectively), while a much longer distance of 6.3 Å is observed in the average NMR structure. Only 2 structures in the entire NMR family sample a contact distance of less than 5.5 Å (Figure 5.13). In contrast with the D44/R55 pair, the PMF for formation of the F47/R55 contact shows only a single minimum at 5.5 Å (Figure 5.8D**)**. Overall, this suggests that R55 has a much more stable interaction with F47 than the salt bridge that it forms with D44.

5.3.3. Simulations of the X-ray structure.

Figure 5.14 shows the backbone RMSD *versus* time and RMSD distributions calculated during the simulation starting from the X-ray structure. The RMSD is shown relative to the X-ray, NMR and simulation (from NMR) structures. After equilibration, the simulation samples backbone conformations (S43-L75) with an

average RMSD relative to the X-ray structure of 1.5 Å and remains quite stable through the 30 ns duration. Overall, there is a preference to adopt structures comparable to the simulation structure discussed above rather than the NMR structure (RMSD compared to the the simulation-equilibrated NMR structure is 1.5 Å below the RMSD to the original NMR structure). Individual helices demonstrate comparable preferences for the X-ray and simulation structures (Data not shown). Hence, the simulations starting from the NMR and X-ray structures both converge to a common simulation structure that is much closer to the X-ray structure than the NMR structure.

5.3.4. Experimental investigation of the putative sidechain interactions.

While simulations can provide a detailed view of molecule structure and dynamics, many approximations are involved, necessitating validation through experimentation. A set of single mutants and double mutants were prepared in order to probe the putative sidechain interactions involving D44/R55 and F47/R55. D44 was mutated to Asn, F47 to Leu and R55 to Met. Thermal unfolding experiments were performed for wildtype HP36 (WT HP36) and for each of the mutants at pH 5.0 (Figure 5.9A, Table 5.1). The WT HP36 has a transition midpoint (T_m) of 73.0 ^oC, while all the variants show a lower melting temperature. The T_m of D44N, F47L, R55M, D44NR55M and F47LR55M are 57.8 ^oC, 45.6 ^oC, 67.3 ^oC, 55.4 ^oC and 35.3 ^oC respectively. From the thermal unfolding curves, at 25 ^oC, 22 % of the population of F47L and 40 % of the population of F47LR55M are unfolded.

Urea denaturation experiments were also carried out in 10 mM sodium acetate and 150 mM sodium chloride at 25 ^oC to determine the free energy of unfolding. The estimated free energy for unfolding (ΔG^o_U) was 3.22 kcal mol^{-1} for WT HP36, 2.48 kcal mol^{-1} for D44N, 2.19 kcal mol^{-1} for R55M and 1.74 kcal mol^{-1} for D44NR55M (Figure 5.9B, Table 5.1). The F47L and F47LR55M mutants were so unstable that the native baseline was not observed (Figure 5.9B) and the unfolding free energy could not be accurately measured by urea denaturation. Thermal and urea denaturation experiments showed that F47L and F47LR55M are partially unfolded in the absence of urea at 25 °C. Previous studies have shown that TMAO can stabilize partially or completely unfolded proteins (*182*). Therefore the combination of urea denaturation and TMAO stabilization can be utilized to estimate the stability of marginally stable proteins. In order to determine the unfolding free energy of F47L and F47LR55M, we performed urea denaturation experiments in increasing TMAO concentrations. For F47L, the titration curves show good pre- and post-transitions in different TMAO concentrations (Figure 5.10A). With increasing TMAO concentrations, the urea denaturation curves shifted to higher urea concentrations. The free energy of unfolding at each TMAO concentration was measured: ΔG^o_U ranges from 1.27 kcal mol^{-1} in 1.62 M TMAO to 1.67 kcal mol^{-1} in 2.50 M TMAO (Table 5.2). Mello and coworkers(*182*) have shown that the free energy of unfolding depends linearly on TMAO concentration. The extrapolated ΔG^o_U of F47L at 0 M TMAO was estimated to be 0.52 kcal mol^{-1} at 25^{o}C (Figure 5.10B), which is in reasonable agreement with the value estimated from the thermal unfolding curve.

Unfortunately the same strategy could not be applied to the F47LR55M double

mutant. High TMAO concentrations are necessary to stabilize the protein to detect the pre-transition, but comparatively high urea concentrations are needed to observe the post-transition. Therefore it is very difficult to find conditions where full unfolding curves could be measured. Thus, we extrapolated from the thermal unfolding data using the Gibbs-Helmholtz equation:

$$\Delta G^{o}{}_{U}(T) = \Delta H^{o}(T_m)\left(1 - \frac{T}{T_m}\right) - \Delta C^{o}{}_{p}\left[T_m - T + T\ln\left(\frac{T}{T_m}\right)\right] \quad (5.7)$$

This calculation requires knowledge of the heat capacity change, $\Delta C^{o}{}_{P}$. HP36 is small, resulting in a very broad differential scanning calorimetry (DSC) transition, which makes it very difficult to calculate the heat capacity accurately by DSC. From the literature, the value of $\Delta C^{o}{}_{P}$ of unfolding is expected to be about 0.012 kcal mol^{-1} K^{-1} per residue of protein (*42*). To a first approximation, the $\Delta C^{o}{}_{P}$ for HP36 can be calculated to be 0.43 kcal mol^{-1} K^{-1}. Another small 41-residue helical protein, the peripheral subunit-binding domain, has a $\Delta C^{o}{}_{P}$ value of 0.43 kcal mol^{-1} K^{-1} (*124*), suggesting that the estimate for HP36 is reasonable. In order to check whether the value of $\Delta C^{o}{}_{P}$ significantly affects the results, we use heat capacities ranging from 0.30 to 0.70 kcal mol^{-1} K^{-1} to calculate the $\Delta G^{o}{}_{U}$. The F47LR55M has a measured T_m of 35.3 ^{o}C and $\Delta H^{o}(T_m)$ of 9.5 kcal mol^{-1}, and the resulting calculated $\Delta G^{o}{}_{U}$ of F47LR55M at 25 ^{o}C ranged from 0.19 to 0.28 kcal mol^{-1} depending on the value of $\Delta C^{o}{}_{P}$ used (Table 5.3). The value of $\Delta G^{o}{}_{U}$ estimated from the Gibbs-Helmholtz equation is in good agreement with the fraction unfolded determined directly from the fit to the thermal melt.

The coupling free energy between the D44 or F47 sidechain and the R55 sidechain, $\Delta\Delta G^{o}{}_{coupling}$, was calculated using equation (8), where $\Delta G^{o}{}_{WT}$ and $\Delta G^{o}{}_{R55M}$ are the free energies of unfolding for wild type protein and R55M single mutant; and $\Delta G^{o}{}_{Single}$ represent D44N or F47L single mutants; and $\Delta G^{o}{}_{Double}$ represents the D44NR55M or F47LR55M double mutants.

$$\Delta\Delta G^{o}{}_{coupling} = (\Delta G^{o}{}_{WT} - \Delta G^{o}{}_{Double}) - [(\Delta G^{o}{}_{Single} - \Delta G^{o}{}_{Double}) + (\Delta G^{o}{}_{R55M} - \Delta G^{o}{}_{Double})] \quad (5.8a)$$

The relationship can be rearranged to a simpler form:

$$\Delta\Delta G^{o}{}_{coupling} = \Delta G^{o}{}_{WT} - \Delta G^{o}{}_{Single} - \Delta G^{o}{}_{R55M} + \Delta G^{o}{}_{Double} \quad (5.8b)$$

Using the $\Delta G^{o}{}_{U}$ values (Table 5.1) measured from experiments, the coupling free energy between the D44 sidechain and the R55 sidechain was close to zero (0.29±0.20 kcal mol^{-1}). In contrast, the coupling free energy between the F47 sidechain and R55 sidechain ranged from 0.70±0.20 to 0.79±0.20 kcal mol^{-1}. The different estimates arise from using different $\Delta C^{o}{}_{P}$ values to calculate $\Delta G^{o}{}_{Double}$. The analysis shows that there is a non-zero coupling between the F47 and R55.

5.4. Discussion

The explicit water MD simulation starting from the NMR structure showed a clear preference to sample structures with much greater similarity to the X-ray

structure, as indicated by RMSD values, DSSP analysis, packing of the phenyalanine core, formation of a C-capping motif on helix-3 and adopting of specific contacts between side chains. Double mutant cycle experiments were performed and demonstrated clear coupling between F47 and R55. It is apparent that these residues are not interacting in the NMR structure but appear to do so in the X-ray structure. Based on free energies calculated from MD simulations and obtained experimentally through double mutant cycles, the F47/R55 contact appears to be a stronger interaction than the proposed salt bridge between D44 and R55. Hence, the van der Waals interaction seen in the X-ray structure appears to play an important role in stabilizing the solution structure of HP36. The coupling free energy between the D44 sidechain and the R55 sidechain is small, only 0.29±0.20 kcal mol^{-1}. The F47 sidechain to R55 sidechain coupling free energy is 0.70±0.20 to 0.79±0.20 kcal mol^{-1}. These results are consistent with the simulation results showing that F47/R55 interacts strongly and that the stability of the D44/R55 pair is lower than the thermal energy.

Previous studies by Frank et al.(*21*) have shown the importance of each phenylalanine in stabilizing the core of the protein. Interestingly, the F47LR55M double mutant is even less stable than these single Phe mutants, which suggests that the sidechain of R55 also plays a key role in stabilizing the structure. It is likely that the optimum packing of the three phenylalanines in the core is enhanced by R55 because it helps to shield the core with its long sidechain and also forms a cation-pi interaction. Cation-pi interactions can be important for folding and thermostability of various proteins and protein ligand systems (*183-185*). For the single mutant, R55M, the Tm dropped 6 ^{o}C in thermal stability, showing that more than just a bulky sidechain it is required at position 55. In the majority of villin sequences, Lys is found as a conservative mutation in place of R55 (*98*). This suggests that the charge is important for stabilizing the structure, but as the simulations and double mutant experiments indicate, the importance of this charge at position 55 does not arise from formation of an ion pair with D44 as observed in the crystal structure. It is worth noting, however that in the X-ray structure D44 appears to be involved in a network of interactions including a hydrogen bond to the backbone carbonyl L42. Backbone sidechain interactions can not be probed by double mutant cycle analysis.

These simulations are models and as with any model there are limitations, especially in the interpretation of results. Realistic, detailed simulations come at a high computational cost that must often be balanced against the need for obtaining extensive conformational sampling. Computational models continue to improve; the Amber and CHARMm force fields have been used extensively enough to identify weaknesses (*166, 186*) such as overstabilization of secondary structure elements. In the simulations that we report here, this type of systematic error might contribute to theextension of alpha-helices that we observed, although we specifically addressed secondary structure bias in the development of the parameter set that was used for all of the present simulations (*166*). It has also been noted that the use of PME to calculate long range electrostatics imposes long-range periodicity that that may result in artifacts from a crystal-like environment (*187-189*). In the present case, simulations with two alternate treatments of long-range interactions (including undesirable direct

truncation) provided essentially the same conclusion, that the simulations adopt a structure in better agreement with the crystal structure than with the NMR structure. Thus there is no evidence that the present results are an artifact of PME.

Previous work by van der Spoel and Lindahl (*163*) reported a series of simulations of the NMR structure of HP36. These authors noted a modest degree of sensitivity to force field, water models, and protonation states. In their simulations, they noted larger structure fluctuations in the region connecting helices 1 and 2 as compared to the rest of the molecule. This observation is consistent with our results, which indicate this linker as one region in which the simulations spontaneously adopt a conformation more consistent with the crystal structure. At that time, there was no way for the van der Spoel and Lindahl to determine whether this larger fluctuation resulted from a conversion toward the crystal structure which was reported two years later. Importantly, van der Spoel and Lindahl also noted the importance of taking into account the pH of the experiment when running simulations of HP36. Upon protonation of the glutamic acids disde chains in the starting structure, the resulting simulation displayed a greater correlation to the chemical shift and j-coupling results which were originally measured at a pH of 3.7. This further suggests that one must be cautious in the quantitative comparison of simulations at neutral pH to experimental data obtained at low pH.

In summary, the results from our simulations and experiments show that the recently published X-ray structure is a more accurate representation of the structure in solution at neutral pH than the NMR structure at low pH. Importantly, the simulations also indicated that a salt bridge between R55 and D44 observed in the low-temperature crystal structure was thermally unstable, in contrast to the stable interaction between R55 and F47 in the simulation. Experimental double mutant analysis confirmed that the interaction free energy of the salt bridge was small, and that the F47-R55 pair likely plays an important role in stabilizing the protein via a cation-pi interaction. The analysis presented here shows how the combination of molecular dynamics simulations and experimental measurements can be used to develop a better understanding of the structural properties of proteins in solution.

Table 5.1: Summary of equilibrium stability measurements for WT HP36 and its mutants in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0 at 25°C.

protein	T_m(°C)	$\Delta H^o(T_m)$ (kcal mol^{-1})	$\Delta G^o_U(H_2O)$ (kcal mol^{-1})	M (kcal mol^{-1} M^{-1})
WT HP36	73.0	31.8	3.22	-0.52
D44N	57.8	32.1	2.48	-0.55
F47L	45.6	15.8	0.52[a]	-0.45[b]
R55M	67.3	26.3	2.19	-0.43
D44NR55M	55.4	27.4	1.74	-0.44
F47LR55M	35.3	9.8	0.19-0.28[c]	N/A

[a] $\Delta G^o_U(H_2O)$ of F47L is extrapolated from urea denaturation in different TMAO concentrations; [b] m is the average value of the m from urea denaturation in different TMAO concentrations; [c] $\Delta G^o_U(H_2O)$ of F47LR55M is calculated from Gibbs-Helmholtz equation using ΔC^o_P values ranging from 0.30-0.70 kcal mol^{-1} K^{-1}.

Table 5.2: Summary of urea denaturation measurements in different TMAO concentrations for F47L in 10 mM sodium acetate, 150 mM sodium chloride, pH 5.0 at 25°C.

[TMAO] (M)	ΔG^{o}_{U}(TMAO) (kcal mol^{-1})	M (kcal mol^{-1} M^{-1})
1.62	1.27	-0.48
1.88	1.38	-0.44
2.15	1.57	-0.45
2.50	1.67	-0.44

Table 5.3: Calculation of $\Delta G^{o}_{U}(H_2O)$ of F47LR55M from Gibbs-Helmholtz equation using different ΔC^{o}_{P} values.

ΔC^{o}_{P} (kcal mol^{-1} K^{-1})	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70
$\Delta G^{o}_{U}(H_2O)$ (kcal mol^{-1})	0.28	0.25	0.24	0.23	0.22	0.214	0.206	0.197	0.19
$\Delta\Delta G^{o}_{coupling}$ [a] (kcal mol^{-1})	0.79	0.76	0.75	0.74	0.73	0.724	0.716	0.707	0.70
[a] The coupling free energy for WT HP36, F47L, R55M and F47LR55M double mutant cycle using different calculated the values of $\Delta G^{o}_{U}(H_2O)$ of F47LR55M.									

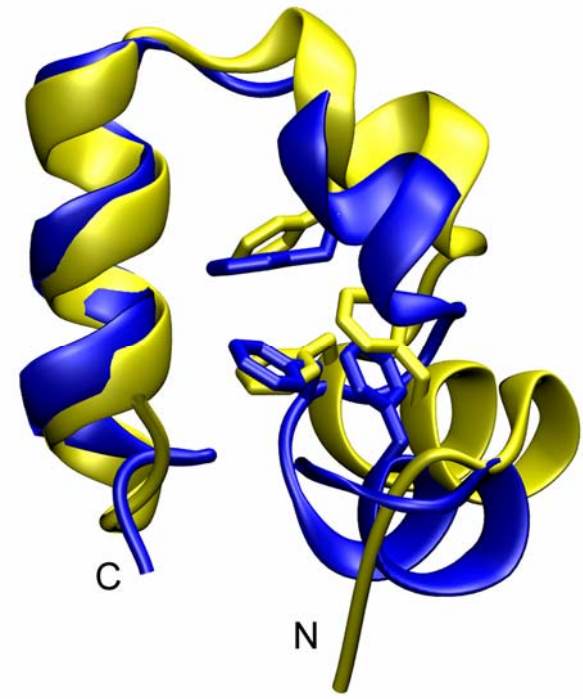


Figure 5.1: Two experimental structures of the villin helical subdomain, showing only the backbone (ribbons) and heavy atoms for the 3 phenylalanines in the core (F47, F51 and F58). The NMR structure of HP36 (pdb code -1VII) is colored blue and the X-ray structure (pdb code – 1YRF) is colored yellow. Differences in the backbone and the phenylalanine core packing are highlighted using a best fit alignment on the backbone residues L62 to F76.

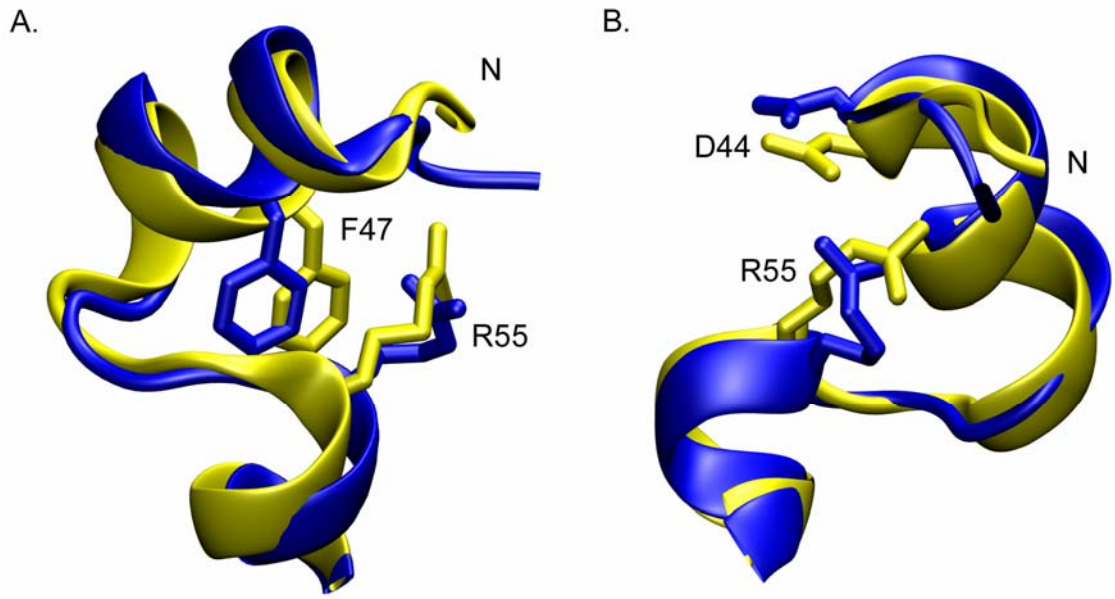


Figure 5.2: Comparison of sidechain interactions in the X-ray and NMR structures, using a best fit alignment on residues L42 to P62. (A) The R55 and F47 sidechains are shown in both the NMR (blue) and X-ray structure (yellow). In the X-ray structure, R55 is involved with a van der Waals contact with F47 and a hydrogen bond with D44. (B) In the X-ray structure, R55-Nε forms a hydrogen bond with D44-Oδ1 in contrast to the NMR structure where the atoms are almost 8 Å apart. The N-terminus is labeled.

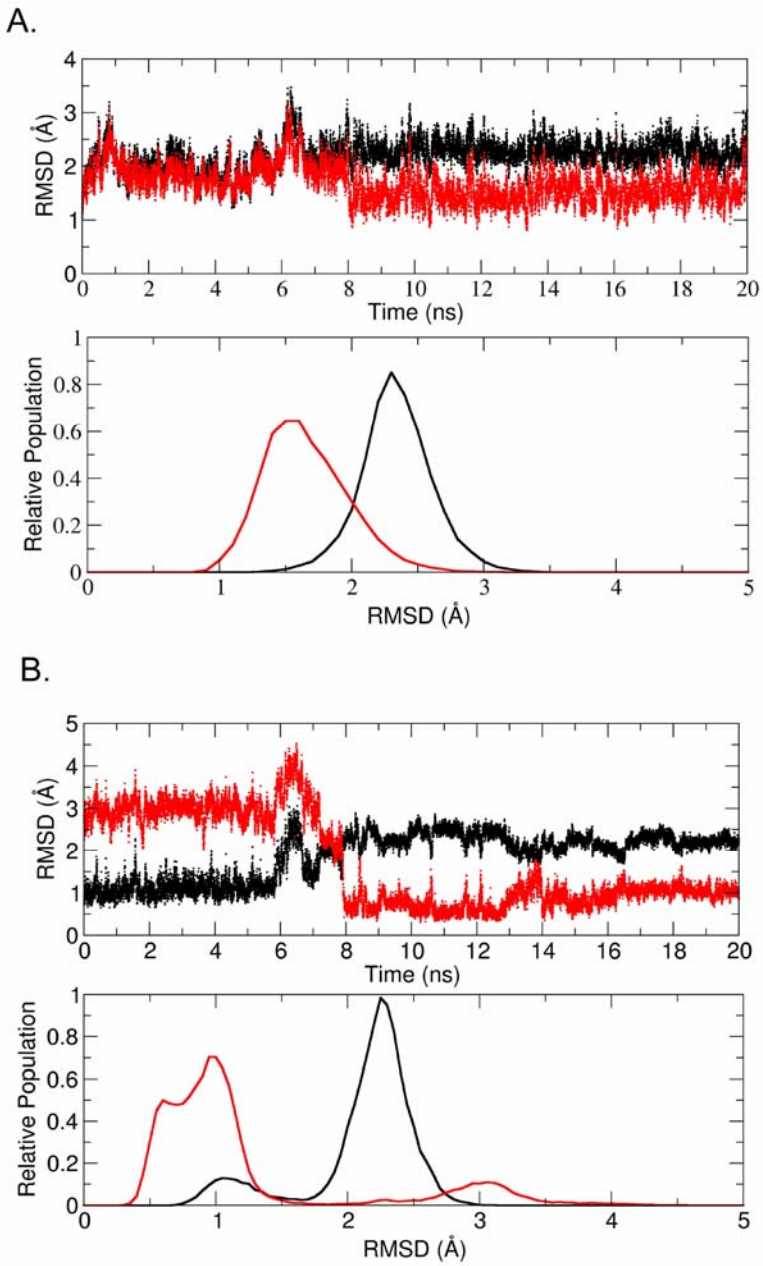


Figure 5.3: Time evolution and histogram distributions of the heavy atom backbone RMSD of (A) residues L42 to F76; and (B) residues P62 to F76 during the simulation of the NMR structure. Each calculation was performed using both the NMR (black) and X-ray (red) structures as the reference. A transition occurs near 8 ns, resulting in lower RMSD values compared to the X-ray structure. The C-terminal region (B) shows a particularly dramatic change from the initial NMR structure to one that much more closely matches the X-ray structure.

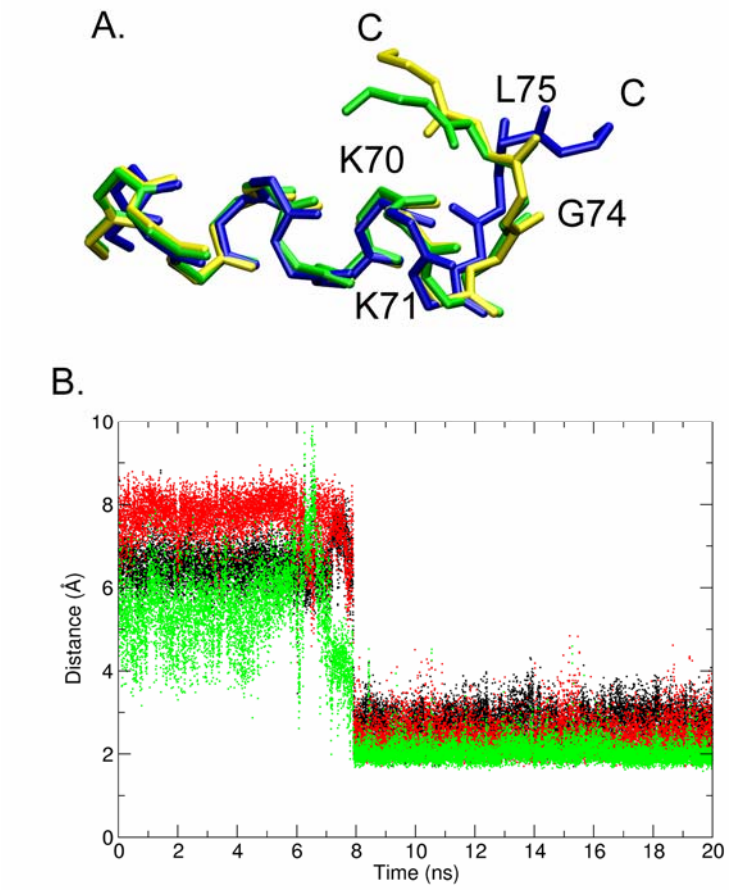


Figure 5.4: (A) Comparison of the C-terminal region (P62-F76) in the X-ray (yellow), NMR (blue) and simulation (green) structures. A key difference between the NMR and X-ray structures is the absence in the NMR structure of a C-capping motif on helix-3 observed in the X-ray structure. This motif is spontaneously adopted in the simulation. (B) The C-capping motif involves three backbone hydrogen bonds (black: K70-G74, red: K71-G74, green: K70-L75) that are formed at ~8 ns and stable throughout the remainder of the simulation.

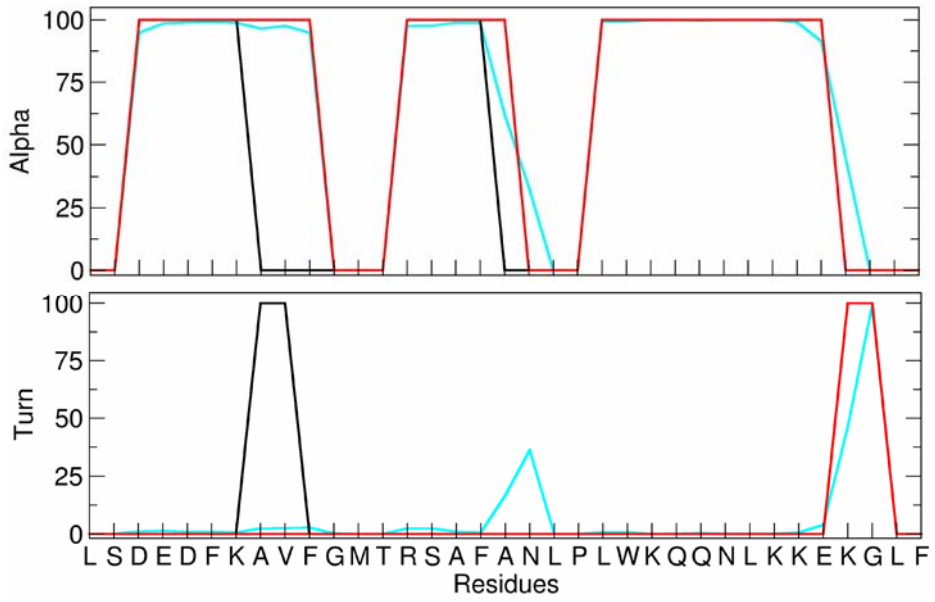


Figure 5.5: DSSP analysis of the NMR (black), X-ray (red); and simulation (cyan) structures of HP36. (a) Alpha helical content per residue. (b) Turn content per residue. Overall, helix-2 and helix-3 are nearly the same length in the X-ray and NMR structures, but helix-1 is 3 residues longer in the X-ray structure than in the NMR structure. The alpha helical content of the MD simulation is in very good agreement with the X-ray structure even though it was initiated from the NMR structure.

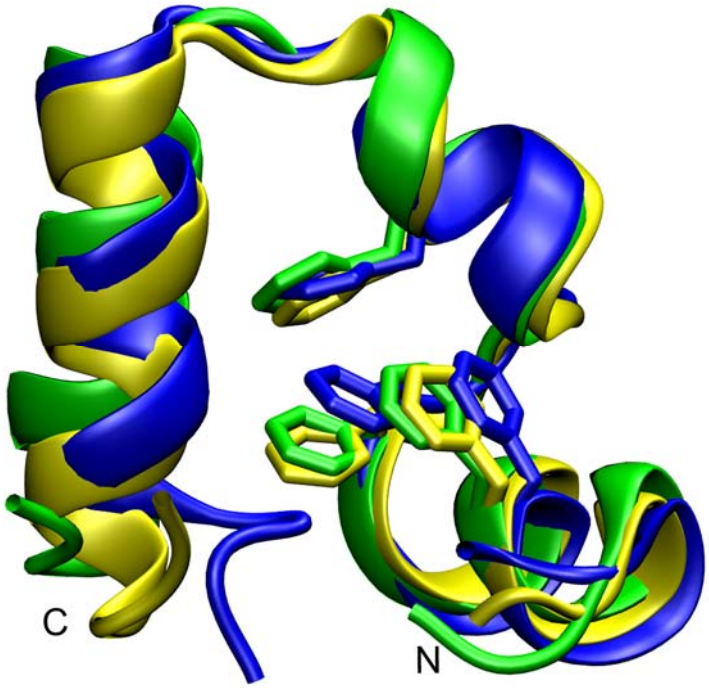


Figure 5.6: Comparison of backbone and core packing in the simulation (green), NMR (blue), and the X-ray (yellow) structure, highlighting the differences in the core packing. A best fit alignment was performed on residues L42 through P62. The packing of the phenylalanine core in the structure from the simulation structure is in much better agreement with the X-ray structure than with the NMR structure.

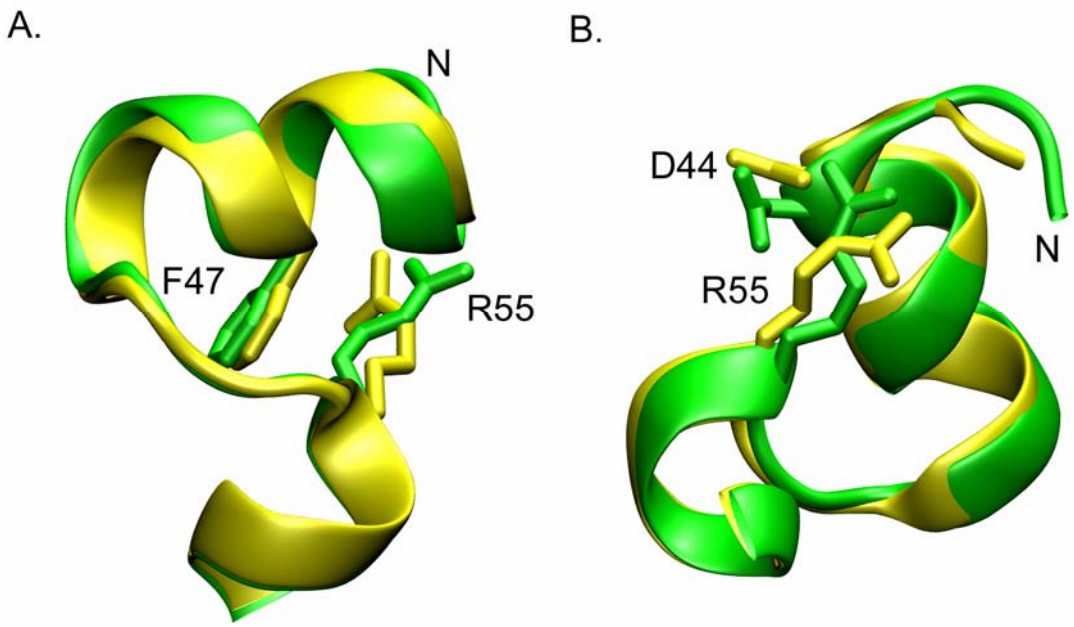


Figure 5.7: Comparison of selected sidechain interactions in the simulation structure (green) and the X-ray structure (yellow). A best fit alignment was performed on residues L42 through P62. In the simulation structure, R55 is 4.7 Å away from the base of the phenylalanine ring (A) and 2.7 Å away from the Oδ1 of D44 (B). This suggests that both contacts may play a role in the stability of the protein.

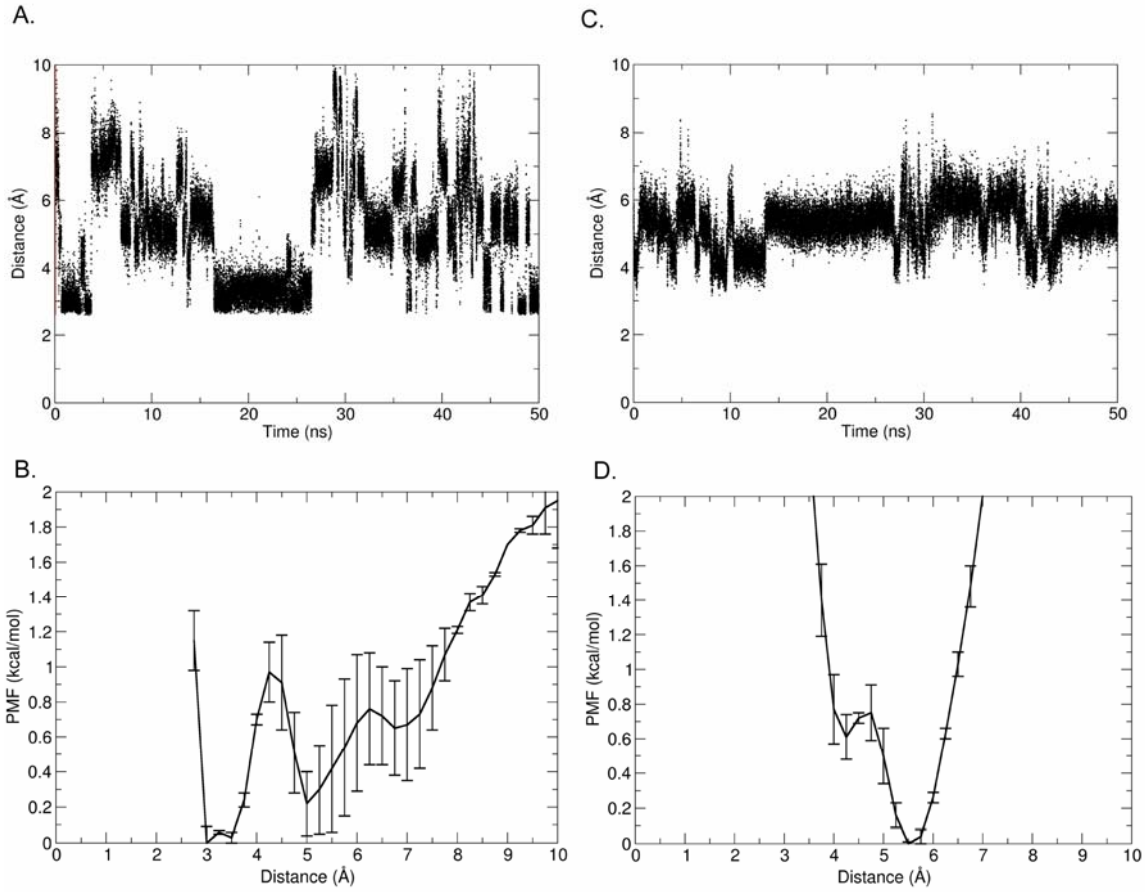


Figure 5.8: Time evolution and PMFs of specific contact distances involving (A and C) R55 and D44 and (B and D) R55 and F47. The distance between R55 and D44 fluctuates throughout the trajectory and shows multiple shallow free energy minima. In contrast, the distance measuring the contact between R55 and F47 is stable during the entire trajectory, with a single free energy minimum at 5.5 Å. The results indicate that the R55/F47 contact is the more stable of these 2 residue pairs.

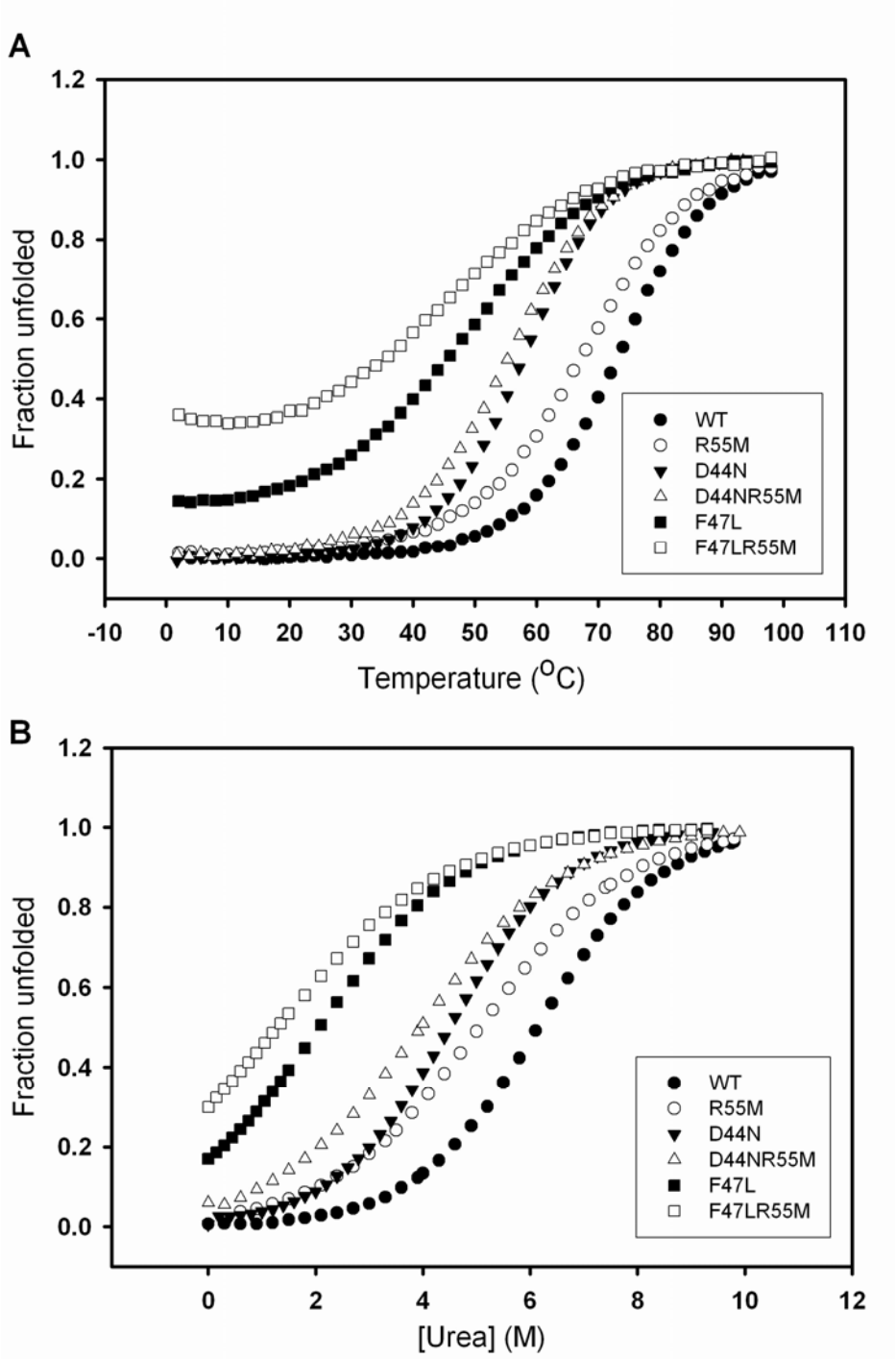


Figure 5.9: (A) Thermal unfolding curves for WT HP36 and its mutants; (B) Urea unfolding curves for WT HP36 and its mutants. Closed circles (●) represents the WT HP36, R55M is represented by open circles (○), D44N by closed triangle (▼), D44NR55M by open triangle (Δ), F47L by closed squares (■) and F47lR55M by open squares (□).

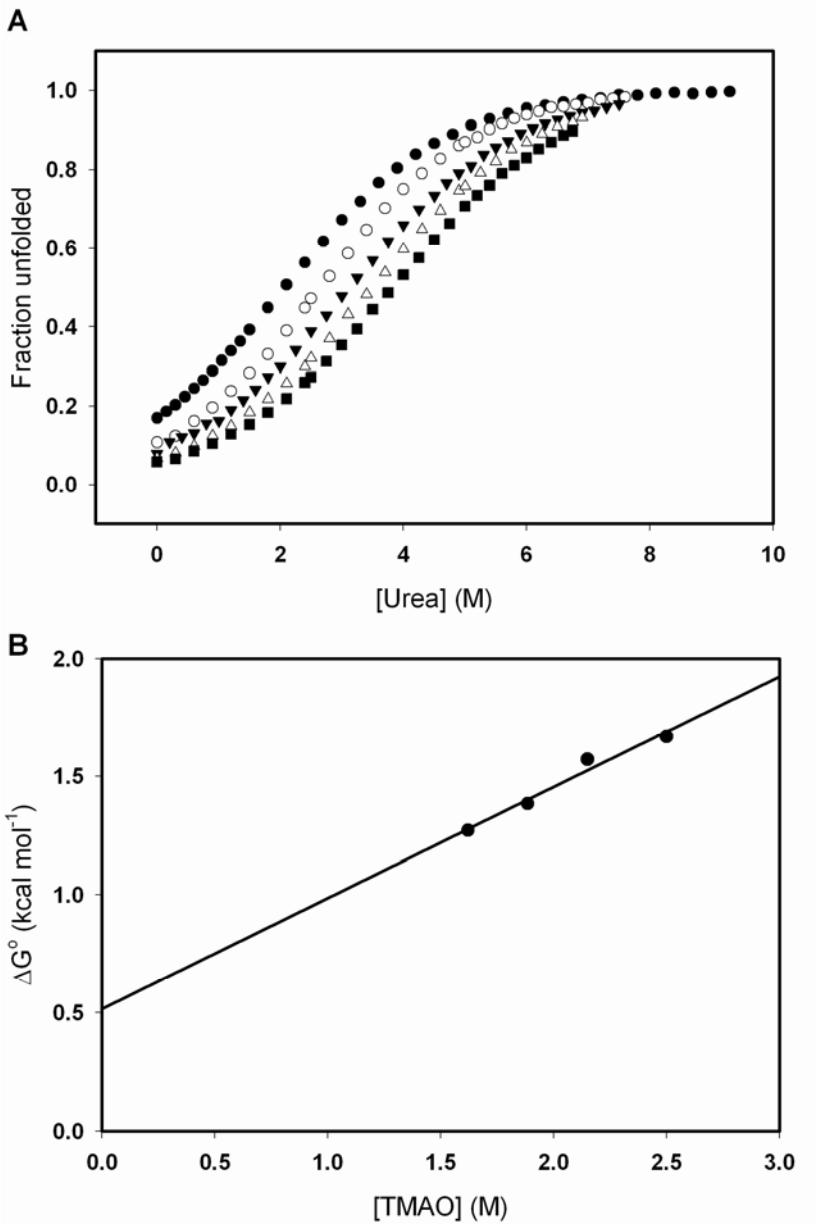
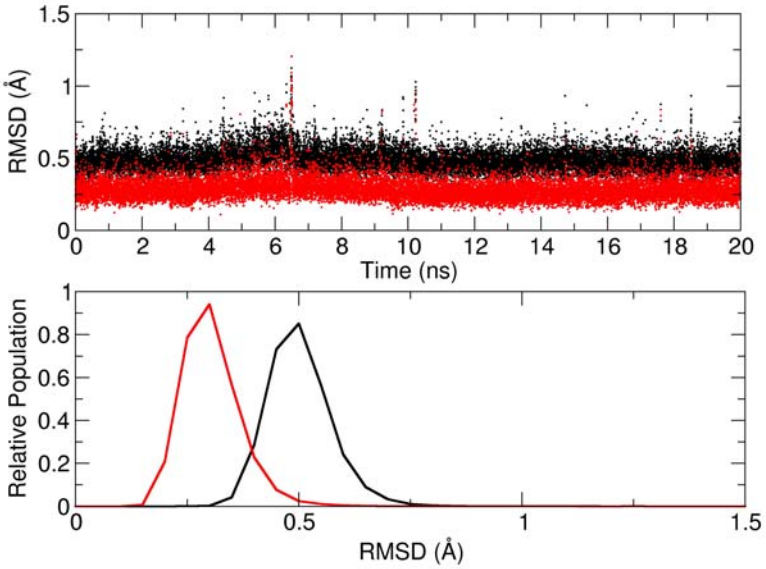


Figure 5.10: (A) Unfolding transitions of the F47L mutant in a mixed urea/ TMAO cosolvent monitored by circular dichroism at 222 nm. Urea denaturation in increasing TMAO concentrations (from left to right at 0, 1.60, 1.88, 2.15, 2.50M TMAO); (B) Dependence of unfolding free energy on TMAO concentration for the F47L mutant; parameters are obtained by fitting urea unfolding curves in different TMA004F concentrations. The straight line is the result of linear regression to each parameter.

(A)



(B)

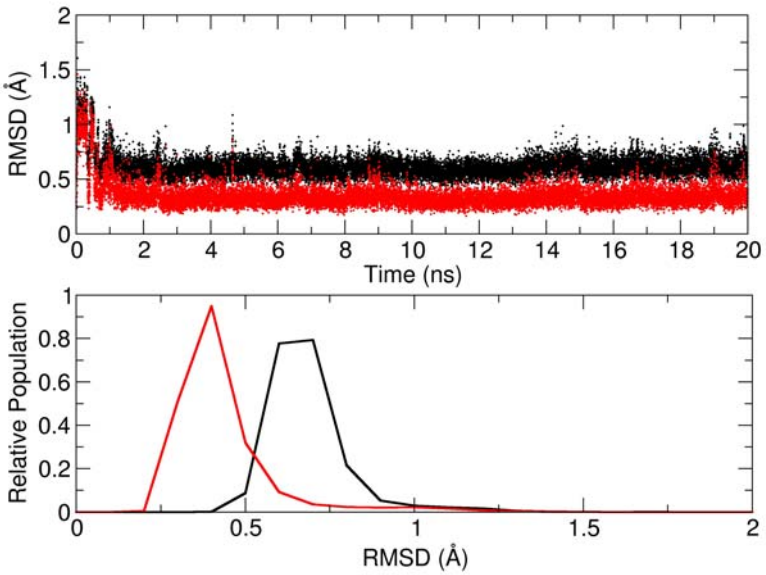


Figure 5.11: Time evolution and distributions of the heavy atom backbone RMSD of the (A) residues 43 to 49 (helix 1); and (B) residues 54 to 59 (helix 2). Each calculation was performed using the NMR (black) and X-ray (red) as a reference structure. The first and second helix remain quite stable during the simulation. Both helices have backbone structures are more structurally similar to the X-ray structure despite being initiated in the NMR structure.

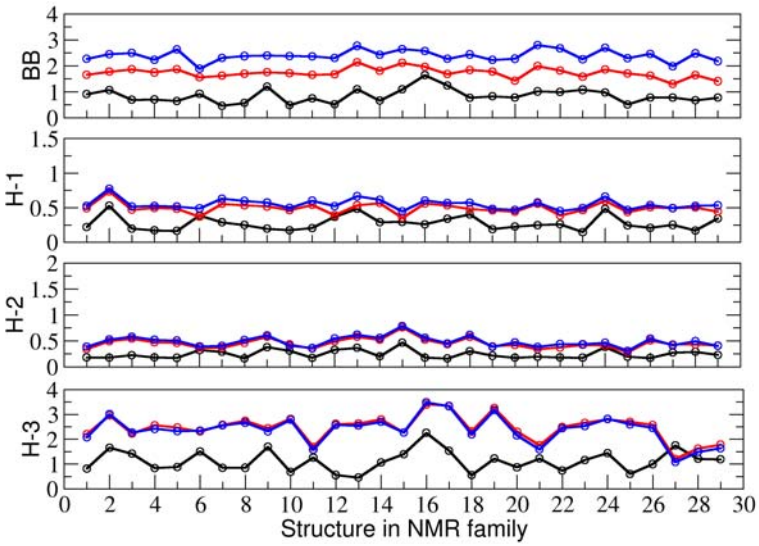


Figure 5.12: RMSD values of each structure in the NMR family, for backbone (BB – residues 42 to 75), helix-1 (H-1 - residues 3 to 9), helix-2 (H-2 – residues 54 to 59), and helix-3 (H-3 – residues 62 to 76). Each calculation was performed using the NMR (black), X-ray (red) and the simulation (Blue) as a reference structure. Overall, the NMR structures are more similar to the NMR average structure than to the simulation or X-ray structure, although individual secondary structure elements are in good agreement with the X-ray structure for a few of the NMR models (e.g. H1 for structure #6).

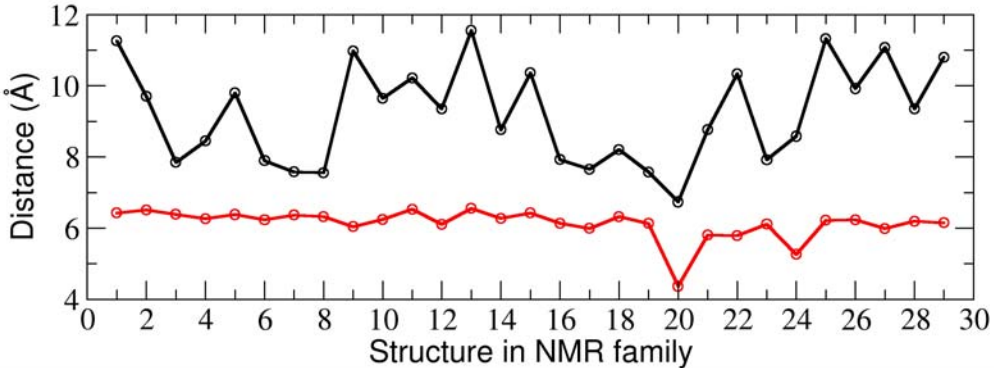


Figure 5.13: Specific contact distances involving R55 and D44 (black) and R55 and F47 (red) for each structure in the NMR family. The contact involving R55 and D44 ranged from 6.7 Å to 11.6 Å and the contact involving R55 and F47 ranged from 4.4 Å to 6.5 Å. However, the later contact has 2 structures with distances between the two residues less than 5.5 Å. For the most part, the NMR family does not contain the hydrogen bond and the van der Waals contact seen in both the X-ray and simulation structure.

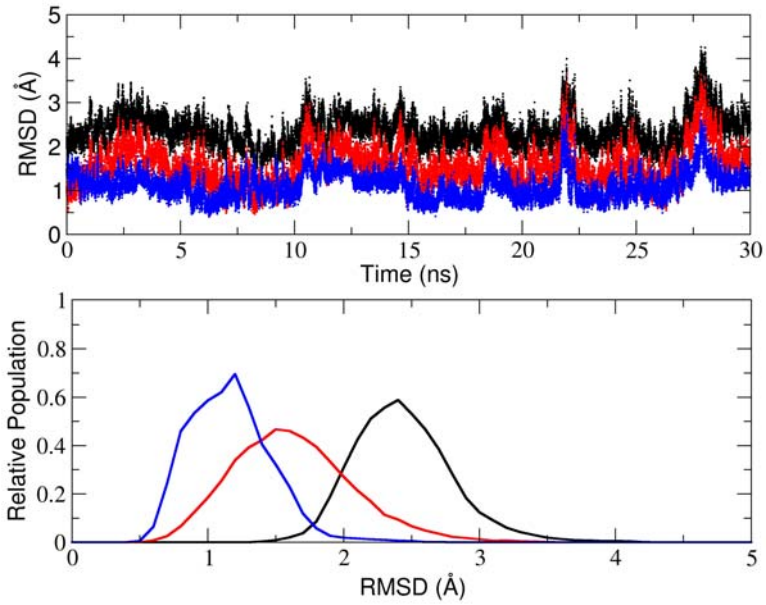
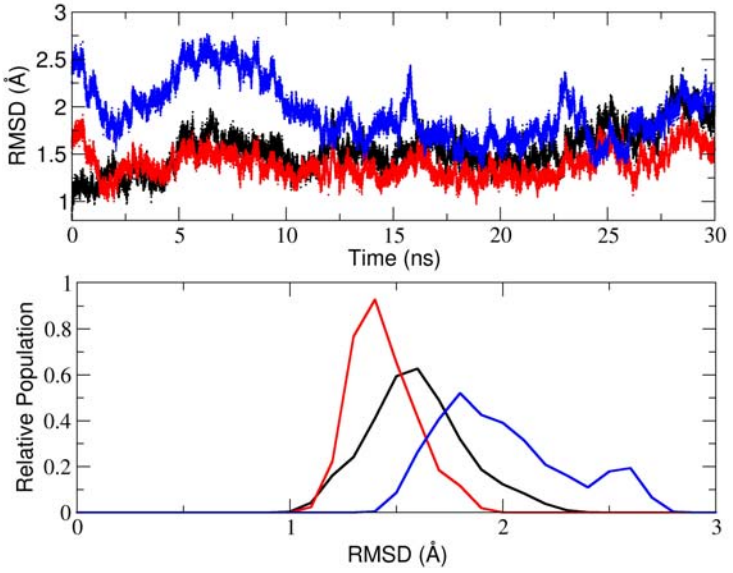


Figure 5.14: Time evolution and distributions of the heavy atom backbone RMSD of the residues 43 to 75 (heavy atoms of the backbone) during the simulation of the X-ray structure. Each calculation was performed using the NMR (black), X-ray (red) and the Simulation (Blue) as a reference structure. The simulation shows a preference for adopting a backbone structure similar to simulation and X-ray structure.

(A)



(B)

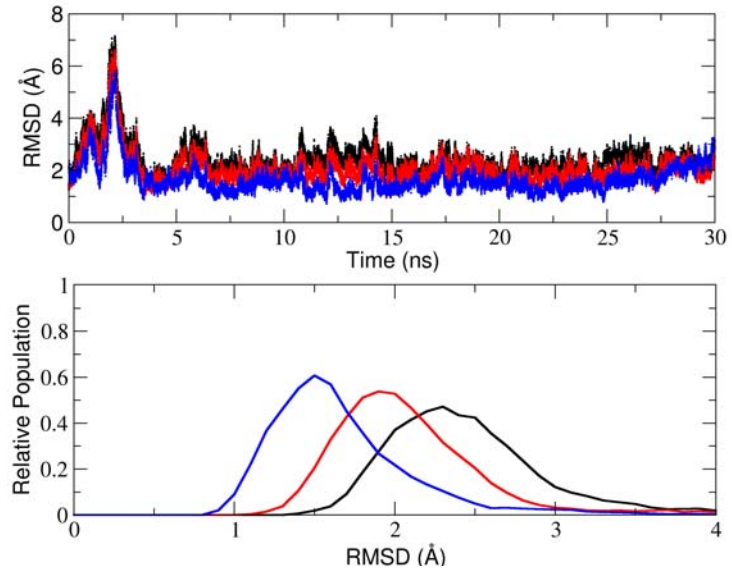


Figure 5.15: Time evolution and distributions of the heavy atom backbone RMSD of the (a) residues 42 to 75 (heavy atoms of the backbone) during the simulation of the NMR structure using (A) a cutoff of 12 Å with no smoothing and (B) IPS for the electrostatic treatment. Each calculation was performed using the NMR (black), X-ray (red) and the Simulation (Blue) as a reference structure. In both cases, the resulting structures are in better agreement with the X-ray structure than NMR structure (red histograms are shifted to lower RMSD values as compared to black).

References

1. Aguzzi, A., and Haass, C. (2003) Games played by rogue proteins in prion disorders and Alzheimer's disease, *Science 302*, 814-818.
2. Chakraborty, C., Nandi, S., and Jana, S. (2005) Prion disease: a deadly disease for protein misfolding, *Current Pharmaceutical Biotechnology 6*, 167-177.
3. Malolepsza, E., Boniecki, M., Kolinski, A., and Piela, L. (2005) Theoretical model of prion propagation: a misfolded protein induces misfolding, *Proceedings of the National Academy of Sciences of the United States of America 102*, 7835-7840.
4. Sigurdson, C., Polymenidou, M., and Aguzzi, A. (2005) Reconstructing prions: fibril assembly from simple yeast to complex mammals, *Neuro-degenerative Diseases 2*, 1-5.
5. Chaudhuri, T. K., and Paul, S. (2006) Protein-misfolding diseases and chaperone-based therapeutic approaches, *The FEBS Journal 273*, 1331-1349.
6. Harrison, R. S., Sharpe, P. C., Singh, Y., and Fairlie, D. P. (2008) Amyloid peptides and proteins in review, *Reviews of Physiology, Biochemistry and Pharmacology 159*, 1-77.
7. Lee, S., and Tsai, F. T. (2005) Molecular chaperones in protein quality control, *Journal of Biochemistry and Molecular Biology 38*, 259-265.
8. Dobson, C. M. (2003) Protein folding and misfolding, *Nature 426*, 884-890.
9. Dobson, C. M. (2003) Protein folding and disease: a view from the first Horizon Symposium, *Nature Reviews 2*, 154-160.
10. Baldwin, R. L. (1999) Protein folding from 1961 to 1982, *Nature Structural Biology 6*, 814-817.
11. Chan, H. S., Bromberg, S. and Dill, K. A. . (1995) Models of co-operativity in protein folding, *Philosophical Transactions: Biological Sciences 348*, 61-70.
12. Matthews, C. R., Crisanti, M. M. (1981) Urea-induced unfolding of the .alpha. subunit of tryptophan synthase: evidence for a multistate process, *Biochemistry 20*, 784-792.
13. Bai, Y., and Englander, S. W. (1996) Future directions in folding: the multi-state nature of protein structure, *Proteins 24*, 145-151.
14. Leopold, P. E., Montal, M., and Onuchic, J. N. (1992) Protein folding funnels: a kinetic approach to the sequence-structure relationship, *Proceedings of the National Academy of Sciences of the United States of America 89*, 8721-8725.
15. Bryngelson, J. D., Onuchic,J. N. , Socci, N. D. and Wolynes,P. G. . (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis, *Proteins: Structure, Function, and Genetics 21*, 167-195.
16. Dobson, C. M., Sali, A. and Karplus,M. (1998) Protein Folding: A Perspective from Theory and Experiment, *Angewandte Chemie International Edition 37*, 868-893.
17. Kauzmann, W. (1959) Some factors in the interpretation of protein

denaturation., *Advances in Protein Chemistry 14*, 1-63.
18. Dill, K. A. (1990) Dominant forces in protein folding, *Biochemistry 29*, 7133-7155.
19. Tanford, C. (1962) Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins, *Journal of the American Chemical Society 84*, 4240-4247.
20. Prevost, M., Wodak, S. J., Tidor, B., and Karplus, M. (1991) Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile-96----Ala mutation in barnase, *Proceedings of the National Academy of Sciences of the United States of America 88*, 10880-10884.
21. 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 Science 11*, 680-687.
22. Pace, C. N. (2001) Polar group burial contributes more to protein stability than nonpolar group burial, *Biochemistry 40*, 310-313.
23. Mirsky, A. E., and Pauling, L. (1936) On the Structure of Native, Denatured, and Coagulated Proteins, *Proceedings of the National Academy of Sciences of the United States of America 22*, 439-447.
24. Pauling, L., Corey, R. B., and Branson, H. R. (1951) The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain, *Proceedings of the National Academy of Sciences of the United States of America 37*, 205-211.
25. Stickle, D. F., Presta, L. G., Dill, K. A., and Rose, G. D. (1992) Hydrogen bonding in globular proteins, *Journal of Molecular Biology 226*, 1143-1159.
26. Shirley, B. A., Stanssens, P., Hahn, U., and Pace, C. N. (1992) Contribution of hydrogen bonding to the conformational stability of ribonuclease T1, *Biochemistry 31*, 725-732.
27. Takano, K., Scholtz, J. M., Sacchettini, J. C., and Pace, C. N. (2003) The contribution of polar group burial to protein stability is strongly context-dependent, *The Journal of Biological Chemistry 278*, 31790-31795.
28. Pace, C. N., Trevino, S., Prabhakaran, E., and Scholtz, J. M. (2004) Protein structure, stability and solubility in water and other solvents, *Philosophical Transactions of the Royal Society of London 359*, 1225-1234; discussion 1234-1225.
29. Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme, *Biochemistry 29*, 2403-2408.
30. Hendsch, Z. S., and Tidor, B. (1999) Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding, *Protein Science 8*, 1381-1392.
31. Kajander, T., Kahn, P. C., Passila, S. H., Cohen, D. C., Lehtio, L., Adolfsen, W., Warwicker, J., Schell, U., and Goldman, A. (2000) Buried charged surface in proteins, *Structure 8*, 1203-1214.
32. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X., and Thomas, S.

T. (2003) Contribution of surface salt bridges to protein stability: guidelines for protein engineering, *Journal of Molecular Biology 327*, 1135-1148.
33. 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 Science 9*.
34. Bosshard, H. R., Marti, D. N., and Jelesarov, I. (2004) Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings, *Journal of Molecular Recognition 17*, 1-16.
35. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) Forces contributing to the conformational stability of proteins, *The Faseb Journal 10*, 75-83.
36. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K., and Baker, D. (2005) Progress in modeling of protein structures and interactions, *Science 310*, 638-642.
37. 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.
38. Marshall, S. A., Morgan, C. S., and Mayo, S. L. (2002) Electrostatics significantly affect the stability of designed homeodomain variants, *Journal of Molecular Biology 316*, 189-199.
39. Spector, S., Wang, M., 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.
40. Strickler, S. S., Gribenko, A. V., 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.
41. Tsai, C. J., Ma, B., Sham, Y. Y., Kumar, S., and Nussinov, R. (2001) Structured disorder and conformational selection, *Proteins 44*, 418-427.
42. Fersht, A. (1998) Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding, W. H. Freeman and company, New York.
43. Englander, S. W., Sosnick, T. R., Englander, J. J., and Mayne, L. (1996) Mechanisms and uses of hydrogen exchange, *Current Opinion in Structural Biology 6*, 18-23.
44. Li, R., and Woodward, C. (1999) The hydrogen exchange core and protein folding, *Protein Science 8*, 1571-1590.
45. Bai, Y., Milne, J. S., Mayne, L., and Englander, S. W. (1994) Protein stability parameters measured by hydrogen exchange, *Proteins 20*, 4-14.
46. Cole, J. L., Hansen, J. C. (1999) Analytical Ultracentrifugation as a Contemporary Biomolecular Research Tool, *Journal of Biomolecular Techniques 10*, 163-176.
47. Condino, J. (1998) Determination of Molecular Weights by Sedimentation

Equilibrium, in *Technical information DS-820*, Beckman Coulter, Inc.
48. Van Holde, K. E. (1985) Physical biochemistry, 2nd ed., pp 124-129, Prentice-Hall, Inc.New Jersey.
49. Voelker, P., McRorie, D. (1994) alpha-Chymotrypsin: Characterization of a Self-Associating System in the Analytical Ultracentrifuge in *Technical information T-1872A.*, Beckman Coulter, Inc.
50. Marshall, S. A., Lazar, G. A., Chirino, A. J., and Desjarlais, J. R. (2003) Rational design and engineering of therapeutic proteins, *Drug Discovery Today 8*, 212-221.
51. Balland, A., Courtney, M., Jallat, S., Tessier, L. H., Sondermeyer, P., de la Salle, H., Harvey, R., Degryse, E., and Tolstoshev, P. (1985) Use of synthetic oligonucleotides in gene isolation and manipulation, *Biochimie 67*, 725-736.
52. Garvey, E. P., and Matthews, C. R. (1990) Site-directed mutagenesis and its application to protein folding, *Biotechnology 14*, 37-63.
53. Wagner, C. R., and Benkovic, S. J. (1990) Site directed mutagenesis: a tool for enzyme mechanism dissection, *Trends in Biotechnology 8*, 263-270.
54. Desjarlais, J. R., and Clarke, N. D. (1998) Computer search algorithms in protein modification and design, *Current Opinion in Structural Biology 8*, 471-475.
55. Pokala, N., and Handel, T. M. (2001) Review: protein design--where we were, where we are, where we're going, *Jornal of Structure Biology 134*, 269-281.
56. Voigt, C. A., Gordon, D. B., and Mayo, S. L. (2000) Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design, *Journal of Molecular Biology 299*, 789-803.
57. Hellinga, H. W. (1997) Rational protein design: combining theory and experiment, *Proceedings of the National Academy of Sciences of the United States of America 94*, 10015-10017.
58. Chen, J., Lu, Z., Sakon, J., and Stites, W. E. (2000) Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability, *Journal of Molecular Biology 303*, 125-130.
59. Kumar, S., Tsai, C. J., and Nussinov, R. (2000) Factors enhancing protein thermostability, *Protein Engineering 13*, 179-191.
60. Lien, S., and Lowman, H. B. (2003) Therapeutic peptides, *Trends in Biotechnology 21*, 556-562.
61. Durrschmidt, P., Mansfeld, J., and Ulbrich-Hofmann, R. (2005) An engineered disulfide bridge mimics the effect of calcium to protect neutral protease against local unfolding, *Febs Journal 272*, 1523-1534.
62. Ghosh, T., Garde, S., and Garcia, A. E. (2003) Role of backbone hydration and salt-bridge formation in stability of alpha-helix in solution, *Biophysical Journal 85*, 3187-3193.
63. Malakauskas, S. M., and Mayo, S. L. (1998) Design, structure and stability of a hyperthermophilic protein variant, *Nature Structure Biology 5*, 470-475.
64. Lazar, G. A., Desjarlais, J. R., and Handel, T. M. (1997) De novo design of the hydrophobic core of ubiquitin, *Protein Science 6*, 1167-1178.

65. Kwok, S. C., Tripet, B., Man, J. H., Chana, M. S., Lavigne, P., Mant, C. T., and Hodges, R. S. (1998) Structural cassette mutagenesis in a de novo designed protein: proof of a novel concept for examining protein folding and stability, *Biopolymers 47*, 101-123.
66. Wisz, M. S., Garrett, C. Z., and Hellinga, H. W. (1998) Construction of a family of Cys2His2 zinc binding sites in the hydrophobic core of thioredoxin by structure-based design, *Biochemistry 37*, 8269-8277.
67. Ventura, S., Vega, M. C., Lacroix, E., Angrand, I., Spagnolo, L., and Serrano, L. (2002) Conformational strain in the hydrophobic core and its implications for protein folding and design, *Nature Structural Biology 9*, 485-493.
68. Cobos, E. S., Filimonov, V. V., Vega, M. C., Mateo, P. L., Serrano, L., and Martinez, J. C. (2003) A thermodynamic and kinetic analysis of the folding pathway of an SH3 domain entropically stabilised by a redesigned hydrophobic core, *Journal of Molecular Biology 328*, 221-233.
69. Adachi, M., Chunying, H., and Utsumi, S. (2004) Effects of designed sulfhydryl groups and disulfide bonds into soybean proglycinin on its structural stability and heat-induced gelation, *Journal of Agricultural and Food Chemistry 52*, 5717-5723.
70. Doucet, A., Williams, M., Gagnon, M. C., Sasseville, M., and Beauregard, M. (2002) Engineering nutritious proteins: improvement of stability in the designer protein MB-1 via introduction of disulfide bridges, *Journal of Agricultural and Food Chemistry 50*, 92-98.
71. Jeong, M. Y., Kim, S., Yun, C. W., Choi, Y. J., and Cho, S. G. (2007) Engineering a de novo internal disulfide bridge to improve the thermal stability of xylanase from Bacillus stearothermophilus No. 236, *Journal of Biotechnology 127*, 300-309.
72. Vaz, D. C., Rodrigues, J. R., Sebald, W., Dobson, C. M., and Brito, R. M. (2006) Enthalpic and entropic contributions mediate the role of disulfide bonds on the conformational stability of interleukin-4, *Protein Science 15*, 33-44.
73. Sanchez-Ruiz, J. M., and Makhatadze, G. I. (2001) To charge or not to charge?, *Trends in Biotechnology 19*, 132-135.
74. Zollars, E. S., Marshall, S. A., and Mayo, S. L. (2006) Simple electrostatic model improves designed protein sequences, *Protein Science 15*, 2014-2018.
75. Bretscher, A., and Weber, K. (1979) Villin: the major microfilament-associated protein of the intestinal microvillus, *Proceedings of the National Academy of Sciences of the United States of America 76*, 2321-2325.
76. Friederich, E., Vancompernolle, K., Huet, C., Goethals, M., Finidori, J., Vandekerckhove, J., and Louvard, D. (1992) An actin-binding site containing a conserved motif of charged amino acid residues is essential for the morphogenic effect of villin, *Cell 70*, 81-92.
77. Azim, A. C., Knoll, J. H., Beggs, A. H., and Chishti, A. H. (1995) Isoform cloning, actin binding, and chromosomal localization of human erythroid dematin, a member of the villin superfamily, *The Journal of Biological*

Chemistry 270, 17407-17413.

78. Glenney, J. R., Jr., and Weber, K. (1981) Calcium control of microfilaments: uncoupling of the F-actin-severing and -bundling activity of villin by limited proteolysis in vitro, *Proceedings of the National Academy of Sciences of the United States of America 78*, 2810-2814.
79. Pope, B., Way, M., Matsudaira, P. T., and Weeds, A. (1994) Characterisation of the F-actin binding domains of villin: classification of F-actin binding proteins into two groups according to their binding sites on actin, *FEBS letters 338*, 58-62.
80. Finidori, J., Friederich, E., Kwiatkowski, D. J., and Louvard, D. (1992) In vivo analysis of functional domains from villin and gelsolin, *The Journal of Cell Biology 116*, 1145-1155.
81. Vardar, D., Buckley, D. A., Frank, B. S., and McKnight, C. J. (1999) NMR structure of an F-actin-binding "headpiece" motif from villin, *Journal of Molecular Biology 294*, 1299-1310.
82. Vardar, D., Chishti, A. H., Frank, B. S., Luna, E. J., Noegel, A. A., Oh, S. W., Schleicher, M., and McKnight, C. J. (2002) Villin-type headpiece domains show a wide range of F-actin-binding affinities, *Cell Motility and the Cytoskeleton 52*, 9-21.
83. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.
84. McKnight, C. J., Matsudaira, P. T., and Kim, P. S. (1997) NMR structure of the 35-residue villin headpiece subdomain, *Nature Structural Biology 4*, 180-184.
85. McKnight, C. J., Doering, D. S., Matsudaira, P. T., and Kim, P. S. (1996) A thermostable 35-residue subdomain within villin headpiece, *Journal of Molecular Biology 260*, 126-134.
86. Wang, M., Tang, Y., Sato, S., Vugmeyster, L., McKnight, C. J., and Raleigh, D. P. (2003) Dynamic NMR line-shape analysis demonstrates that the villin headpiece subdomain folds on the microsecond time scale, *Journal of the American Chemical Society 125*, 6032-6033.
87. 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, *Proceedings of the National Academy of Sciences of the United States of America 102*, 7517-7522.
88. Tang, Y., 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.
89. Tang, Y., 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.
90. 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, *Journal of Molecular Biology 360*, 1094-1107.

91. 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, *Journal of Molecular Biology 339*, 915-925.
92. Butt, T. R., Edavettal, S. C., Hall, J. P., and Mattern, M. R. (2005) SUMO fusion technology for difficult-to-express proteins, *Protein Expression and Purification 43*, 1-9.
93. Sato, S. (2002) Folding of ribosomal protein L9 and its isolated N- and C-terminal domains, in *Department of Chemistry*, State University of New York at Stony Brook, New York.
94. 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.
95. Islam, S. A., Karplus, M., and Weaver, D. L. (2002) Application of the diffusion-collision model to the folding of three-helix bundle proteins, *Journal of Molecular Biology 318*, 199-215.
96. Jang, S., Kim, E., Shin, S., and Pak, Y. (2003) Ab initio folding of helix bundle proteins using molecular dynamics simulations, *Journal of the American Chemical Society 125*, 14841-14846.
97. Sullivan, D. C., and Kuntz, I. D. (2002) Protein Folding as Biased Conformational Diffusion, *The Journal of Physical Chemistry B 106*, 3255-3262.
98. Vermeulen, W., Vanhaesebrouck, P., Van Troys, M., Verschueren, M., Fant, F., Goethals, M., Ampe, C., Martins, J. C., and Borremans, F. A. (2004) Solution structures of the C-terminal headpiece subdomains of human villin and advillin, evaluation of headpiece F-actin-binding requirements, *Protein Science 13*, 1276-1287.
99. Bokor, M., Csizmok, V., Kovacs, D., Banki, P., Friedrich, P., Tompa, P., and Tompa, K. (2005) NMR relaxation studies on the hydrate layer of intrinsically unstructured proteins, *Biophysical Journal 88*, 2030-2037.
100. Bourhis, J. M., Receveur-Brechot, V., Oglesbee, M., Zhang, X., Buccellato, M., Darbon, H., Canard, B., Finet, S., and Longhi, S. (2005) The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded, *Protein Science 14*, 1975-1992.
101. Hamada, D., Kato, T., Ikegami, T., Suzuki, K. N., Hayashi, M., Murooka, Y., Honda, T., and Yanagihara, I. (2005) EspB from enterohaemorrhagic Escherichia coli is a natively partially folded protein, *The FEBS Journal 272*, 756-768.
102. Sanchez-Puig, N., Veprintsev, D. B., and Fersht, A. R. (2005) Human full-length Securin is a natively unfolded protein, *Protein Science 14*, 1410-1418.
103. Uversky, V. N. (2002) What does it mean to be natively unfolded?, *European Journal of Biochemistry / FEBS 269*, 2-12.
104. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins:

re-assessing the protein structure-function paradigm, *Journal of Molecular Biology 293*, 321-331.
105. Kuhlman, B., Boice, J. A., Fairman, R., and Raleigh, D. P. (1998) Structure and stability of the N-terminal domain of the ribosomal protein L9: evidence for rapid two-state folding, *Biochemistry 37*, 1025-1032.
106. 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 H2O and D2O. Determination of the thermodynamic parameters, deltaH(o), deltaS(o), and deltaC(o)p and evaluation of solvent isotope effects, *Protein Science 7*, 2405-2412.
107. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis, *Nucleic Acids Research 31*, 3784-3788.
108. Selkoe, D. J. (2003) Folding proteins in fatal ways, *Nature 426*, 900-904.
109. Ellisdon, A. M., and Bottomley, S. P. (2004) The role of protein misfolding in the pathogenesis of human diseases, *IUBMB Life 56*, 119-123.
110. Vugmeyster, L., Trott, O., McKnight, C. J., Raleigh, D. P., and Palmer, A. G., 3rd. (2002) Temperature-dependent dynamics of the villin headpiece helical subdomain, an unusually small thermostable protein, *Journal of Molecular Biology 320*, 841-854.
111. 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, *Proceedings of the National Academy of Sciences of the United States of America 95*, 9897-9902.
112. Sullivan, D. C., and Kuntz, I. D. (2001) Conformation spaces of proteins, *Proteins 42*, 495-511.
113. Shen, M. Y., and Freed, K. F. (2002) All-atom fast protein folding simulations: the villin headpiece, *Proteins 49*, 439-445.
114. 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.
115. Kubelka, J., Hofrichter, J., and Eaton, W. A. (2004) The protein folding 'speed limit', *Current Opinion in Structural Biology 14*, 76-88.
116. Wei, Z., and Song, J. (2005) Molecular mechanism underlying the thermal stability and pH-induced unfolding of CHABII, *Journal of Molecular Biology 348*, 205-218.
117. Loladze, V. V., and Makhatadze, G. I. (2002) Removal of surface charge-charge interactions from ubiquitin leaves the protein folded and very stable, *Protein Science 11*, 174-177.
118. Schwehm, J. M., Fitch, C. A., Dang, B. N., Garcia-Moreno, E. B., and Stites, W. E. (2003) Changes in stability upon charge reversal and neutralization substitution in staphylococcal nuclease are dominated by favorable electrostatic effects, *Biochemistry 42*, 1118-1128.

119. Dominy, B. N., Minoux, H., and Brooks, C. L., 3rd. (2004) An electrostatic basis for the stability of thermophilic proteins, *Proteins 57*, 128-141.
120. Bi, Y., Tang, Y., Raleigh, D. P., and Cho, J. H. (2006) Efficient high level expression of peptides and proteins as fusion proteins with the N-terminal domain of L9: application to the villin headpiece helical subdomain, *Protein Expression and Purification 47*, 234-240.
121. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. M. (1992) *Analytical Ultracentrifugation in Biochemistry and Polymer Science (Harding, S. E. , Rowe, A. J. , and Horton, J. C., eds)*, Royal Society of Chemistry, London
122. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein, *Protein Science 4*, 2411-2423.
123. Spector, S., Kuhlman, B., Fairman, R., Wong, E., Boice, J. A., and Raleigh, D. P. (1998) Cooperative folding of a protein mini domain: the peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex, *Journal of Molecular Biology 276*, 479-489.
124. 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.
125. Tang, Y., Grey, M. J., McKnight, J., Palmer, A. G., 3rd, and Raleigh, D. P. (2006) Multistate folding of the villin headpiece domain, *Journal of Molecular Biology 355*, 1066-1077.
126. Kay, L. E., Ikura, M., Tschudin, R. and Bax, A. (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins, *Journal of Magnetic Resonance 89*, 496-514.
127. Yamazaki, T., Nicholson, L.K., Torchia, D.A., Wingfield, P., Stahl, S.J., Kaufman, J.D., Eyermann, C.J., Hodge, C.N., Lam. P.Y.S., Ru, Y., Jadhav, P.K., Chang, C.H. and Weber, P.C. (1994) NMR and X-rays evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor, *Journal of the American Chemical Society 116*, 10791-10792.
128. Fraczkiewicz, R., and Braun, W. (1998) Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules, *Journal of Computational Chemistry 19*, 319-333.
129. 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.
130. Rees, D. C., and Robertson, A. D. (2001) Some thermodynamic implications for the thermostability of proteins, *Protein Science 10*, 1187-1194.
131. 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, *Journal of Molecular Biology 326*, 1261-1270.

132. Privalov, P. L. (1989) Thermodynamic problems of protein structure, *Annual Review of Biophysics and Biophysical Chemistry 18*, 47-69.
133. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *Journal of Molecular Graphics 14*, 51-55, 29-32.
134. Arpin, M., Pringault, E., Finidori, J., Garcia, A., Jeltsch, J. M., Vandekerckhove, J., and Louvard, D. (1988) Sequence of human villin: a large duplicated domain homologous with other actin-severing proteins and a unique small carboxy-terminal domain related to villin specificity, *The Journal of Cell Biology 107*, 1759-1766.
135. Meng, J., Vardar, D., Wang, Y., Guo, H. C., Head, J. F., and McKnight, C. J. (2005) High-resolution crystal structures of villin headpiece and mutants with reduced F-actin binding activity, *Biochemistry 44*, 11963-11973.
136. Bandyopadhyay, S., Chakraborty, S., and Bagchi, B. (2005) Secondary structure sensitivity of hydrogen bond lifetime dynamics in the protein hydration layer, *Journal of the American Chemical Society 127*, 16660-16667.
137. Faccioli, P., Sega, M., Pederiva, F., and Orland, H. (2006) Dominant pathways in protein folding, *Physical Review Letters 97*, 108101.
138. Jayachandran, G., Vishal, V., Garcia, A. E., and Pande, V. S. (2006) Local structure formation in simulations of two small proteins, *Journal of Structural Biology 157,*491-499.
139. Jang, S., Sreerama, N., Liao, V. H., Lu, S. H., Li, F. Y., Shin, S., Woody, R. W., and Lin, S. H. (2006) Theoretical investigation of the photoinitiated folding of HP-36, *Protein Science 15*, 2290-2299.
140. Jayachandran, G., Vishal, V., and Pande, V. S. (2006) Using massively parallel simulation and Markovian models to study protein folding: examining the dynamics of the villin headpiece, *The Journal of Chemical Physics 124*, 164902.
141. Trebst, S., Troyer, M., and Hansmann, U. H. (2006) Optimized parallel tempering simulations of proteins, *The Journal of Chemical Physics 124*, 174903.
142. 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, *Journal of the American Chemical Society 128*, 11742-11743.
143. Carr, J. M., and Wales, D. J. (2005) Global optimization and folding pathways of selected alpha-helical proteins, *The Journal of Chemical Physics 123*, 234901.
144. Fogolari, F., Tosatto, S. C., and Colombo, G. (2005) A decoy set for the thermostable subdomain from chicken villin headpiece, comparison of different free energy estimators, *BMC Bioinformatics 6*, 301.
145. Guner, U., Arkun, Y., and Erman, B. (2006) Optimum folding pathways of proteins: their determination and properties, *The Journal of Chemical Physics 124*, 134911.

146. Massi, F., and Palmer, A. G., 3rd. (2003) Temperature dependence of NMR order parameters and protein dynamics, *Journal of the American Chemical Society 125*, 11158-11159.
147. Zagrovic, B., and van Gunsteren, W. F. (2006) Comparing atomistic simulation data with the NMR experiment: how much can NOEs actually tell us?, *Proteins 63*, 210-218.
148. De Mori, G. M., Colombo, G., and Micheletti, C. (2005) Study of the Villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics, *Proteins 58*, 459-471.
149. Hansmann, U. H. (2004) Simulations of a small protein in a specifically designed generalized ensemble, *Physical Review 70*, 012902.
150. Herges, T., and Wenzel, W. (2004) An all-atom force field for tertiary structure prediction of helical proteins, *Biophysical Journal 87*, 3100-3109.
151. Mukherjee, A., and Bagchi, B. (2004) Contact pair dynamics during folding of two small proteins: chicken villin head piece and the Alzheimer protein beta-amyloid, *The Journal of Chemical Physics 120*, 1602-1612.
152. Srinivasan, R., Fleming, P. J., and Rose, G. D. (2004) Ab initio protein folding using LINUS, *Methods in Enzymology 383*, 48-66.
153. Zagrovic, B., and Pande, V. S. (2004) How does averaging affect protein structure comparison on the ensemble level?, *Biophysical Journal 87*, 2240-2246.
154. 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, *Proceedings of the National Academy of Sciences of the United States of America 102*, 16662-16667.
155. 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.
156. O'Neil, K. T., and DeGrado, W. F. (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids, *Science 250*, 646-651.
157. Rath, A., and Davidson, A. R. (2000) The design of a hyperstable mutant of the Abp1p SH3 domain by sequence alignment analysis, *Protein Science 9*, 2457-2469.
158. Kubelka, J., Eaton, W. A., and Hofrichter, J. (2003) Experimental tests of villin subdomain folding simulations, *Journal of Molecular Biology 329*, 625-630.
159. Buscaglia, M., Kubelka, J., Eaton, W. A., and Hofrichter, J. (2005) Determination of ultrafast protein folding rates from loop formation dynamics, *Journal of Molecular Biology 347*, 657-664.
160. Havlin, R. H., and Tycko, R. (2005) Probing site-specific conformational distributions in protein folding with solid-state NMR, *Proceedings of the National Academy of Sciences of the United States of America 102*, 3284-3289.
161. Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A., and Hofrichter, J. (2006)

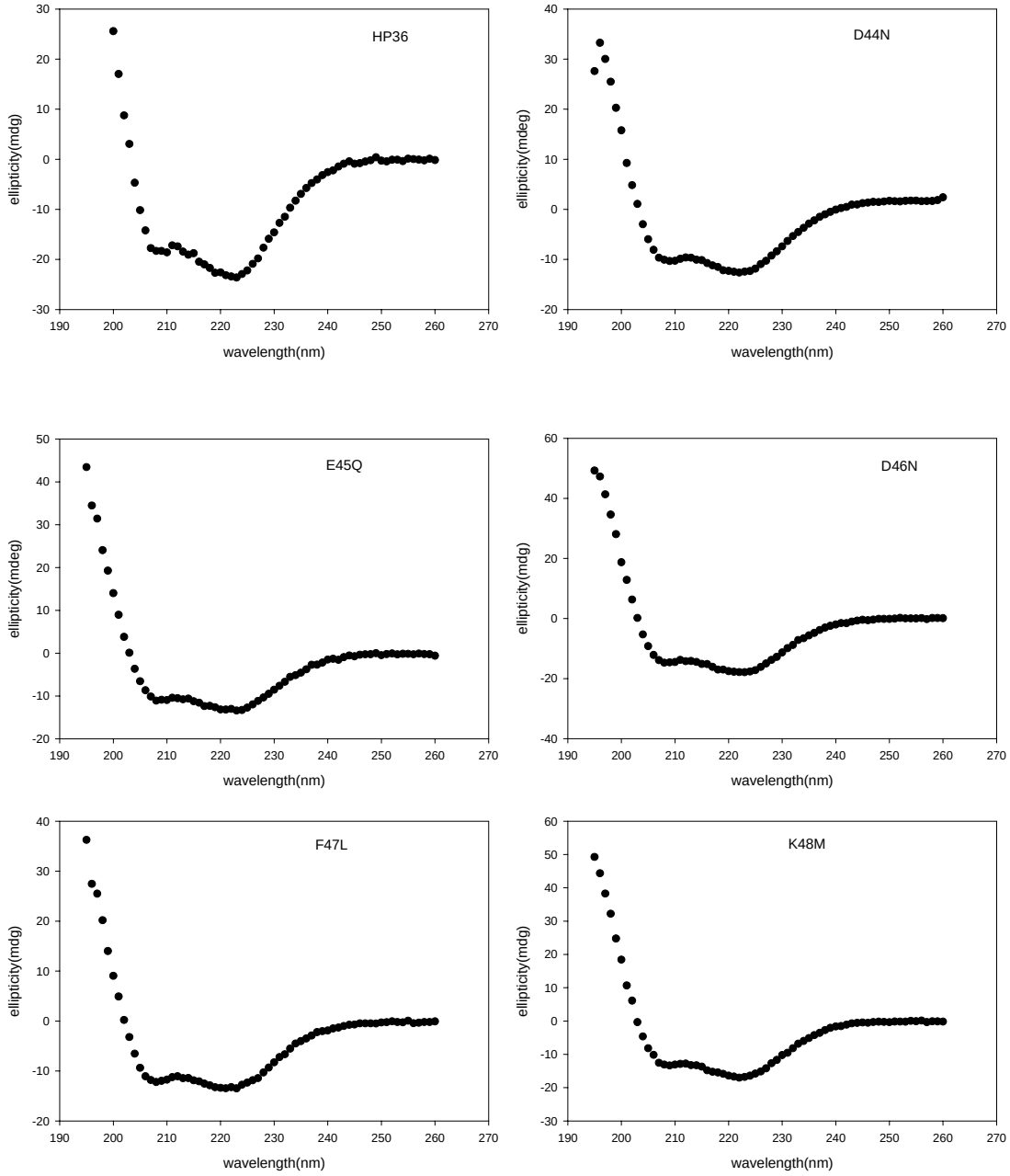
Sub-microsecond protein folding, *Journal of Molecular Biology 359*, 546-553.
162. Duan, Y., and Kollman, P. A. (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution, *Science 282*, 740-744.
163. van der Spoel, D., Lindahl, E. (2003) Brute-Force Molecular Dynamics Simulations of Villin Headpiece: Comparison with NMR Parameters, *The Journal of Physical Chemistry B 107*, 11178-11187.
164. 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, *Journal of Molecular Biology 323*, 927-937.
165. Case, D., Cheatham, T. I. I. I., Darden, T., G., G., Luo, R., Merz Jr, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *Journal of Computational Chemistry 26*, 1668-1688.
166. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins 65*, 712-725.
167. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz Jr, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *Journal of the American Chemical Society 117*, 5179-5197.
168. Wang, J., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, *Journal of Computational Chemistry 21*, 1049-1074.
169. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *Journal of Computational Physics 23*, 327-341.
170. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *Journal of Chemical Physics 81*, 3684-3690.
171. Shirts, M. R., and Pande, V. S. (2005) Solvation free energies of amino acid side chain analogs for common molecular mechanics water models, *The Journal of Chemical Physics 122*, 134508.
172. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendrickson, T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics, *Journal of the American Chemical Society 112*, 6127-6129.
173. Zhou, R. (2003) Free energy landscape of protein folding in water: Explicit vs. implicit solvent, *Proteins: Structure, Function, and Genetics 53*, 148-161.
174. Pitera, J. W., and Swope, W. (2003) Understanding folding and design: Replica-exchange simulations of ``Trp-cage'' miniproteins, *Proceedings of the*

National Academy of Sciences of the United States of America 100, 7587-7592.

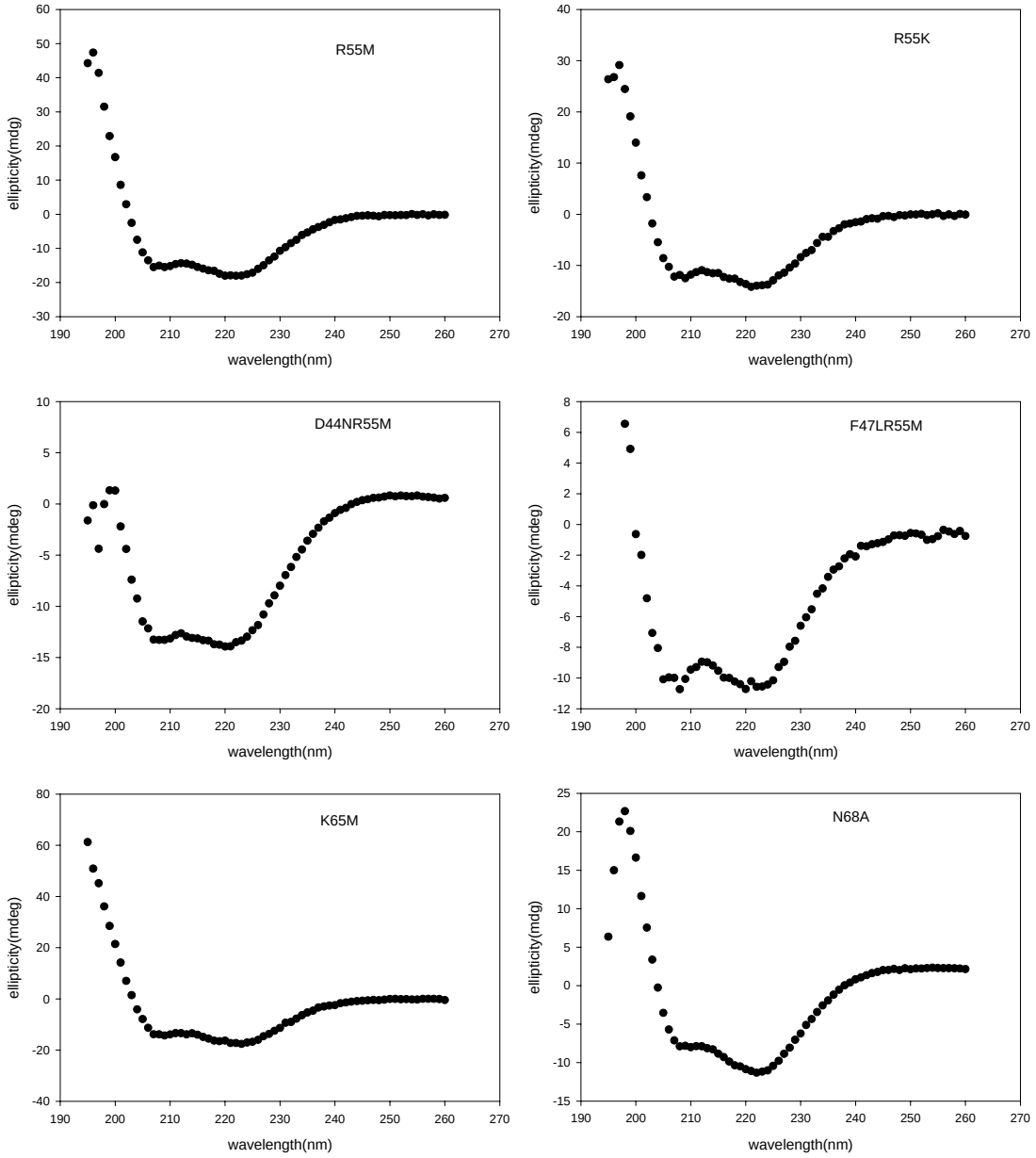
175. Zhou, R., and Berne, B. J. (2002) Can a continuum solvent model reproduce the free energy landscape of a beta -hairpin folding in water?, *Proceedings of the National Academy of Sciences of the United States of America 99*, 12777-12782.
176. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems, *Journal of Chemical Physics 98*, 10089-10092.
177. Wu, X. W., and Brooks, B. R. (2005) Isotropic periodic sum: A method for the calculation of long-range interactions, *Journal of Chemical Physics 122*, -.
178. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *Journal of Chemical Physics 79*, 926-935.
179. Simmerling, C., Elber, R. and Zhang, J., . (1995) MOIL-View - A Program for Visualization of Structure and Dynamics of Biomolecules and STO- A Program for Computing Stochastic Paths, in Modelling of Biomolecular Structure and Mechanisms, in *Modeling of Biomolecular Structure and Mechanism* (Pullman, A., Ed.), pp 241-265, Kluwer, Netherlands.
180. Wang, A., and Bolen, D. W. (1997) A naturally occurring protective system in urea-rich cells: mechanism of osmolyte protection of proteins against urea denaturation, *Biochemistry 36*, 9101-9108.
181. Kabsch, W., and Sander, C. . (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577-2637.
182. Mello, C. C., and Barrick, D. (2003) Measuring the stability of partly folded proteins using TMAO, *Protein Science 12*, 1522-1529.
183. Dougherty, D. A. (1996) Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp, *Science 271*, 163-168.
184. Gallivan, J. P., and Dougherty, D. A. (1999) Cation-pi interactions in structural biology, *Proceedings of the National Academy of Sciences of the United States of America 96*, 9459-9464.
185. Shi, Z., Olson, C. A., Bell, A. J. J., and Kallenbach, N. R. . (2001) Stabilization of alpha-helix structure by polar side-chain interactions: Complex salt bridges, cation-pi interactions, and C-H...O H-bonds, *Biopolymers 60*, 366-380.
186. Okur, A., Strockbine, B., Hornak, V., and Simmerling, C. (2003) Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins, *Journal of Computational Chemistry 24*, 21-31.
187. Hunenberger, P. H., and McCammon, J. A. (1999) Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: a continuum electrostatics study, *Biophysical Chemistry 78*, 69-88.
188. Hunenberger, P. H., and McCammon, J. A. (1999) Ewald artifacts in computer simulations of ionic solvation and ion-ion interaction: A continuum

electrostatics study, *Journal of Chemical Physics 110*, 1856-1872.
189. Weber, W., Hunenberger, P. H., and McCammon, J. A. (2000) Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: Influence of artificial periodicity on peptide conformation, *Journal of Physical Chemistry B 104*, 3668-3675.

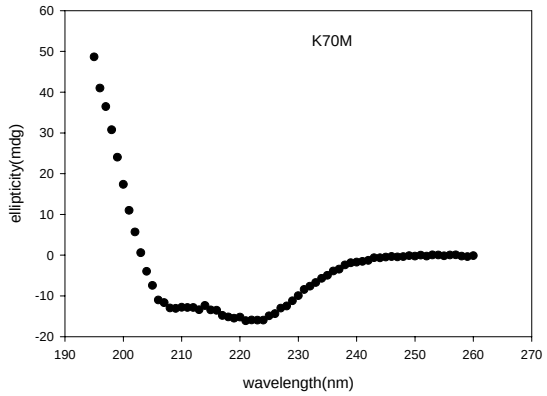
Appendix 1. Wavelength scan curves for WT HP36 and its mutants in 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0 and 25^{o}C.

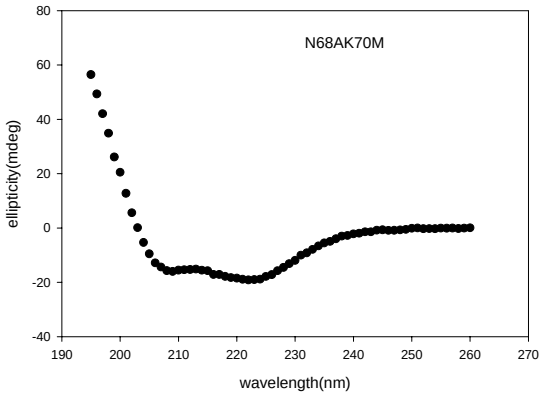


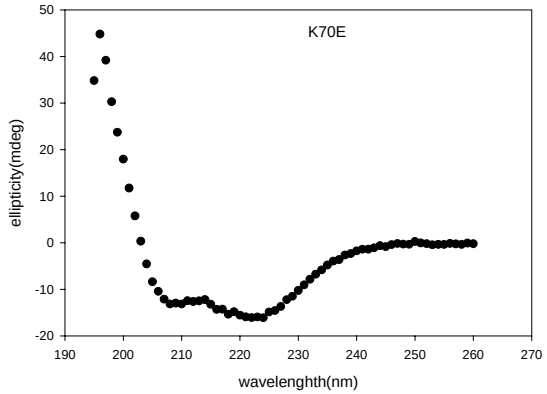
Appendix 1. *continued.*

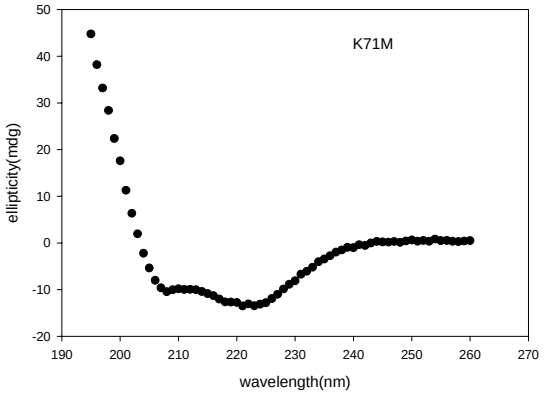


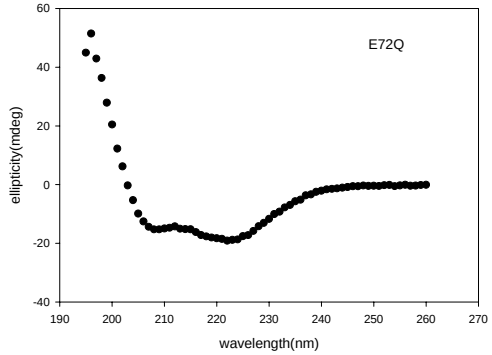
Appendix 1. *continued.*

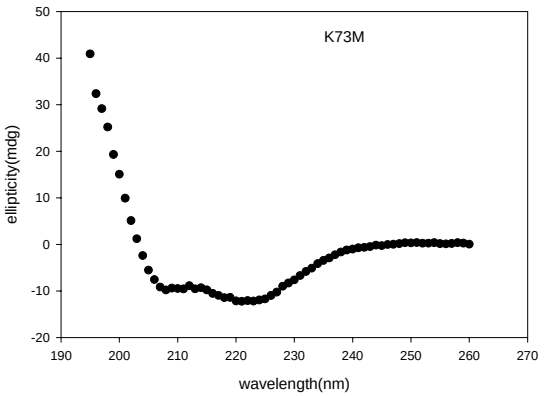




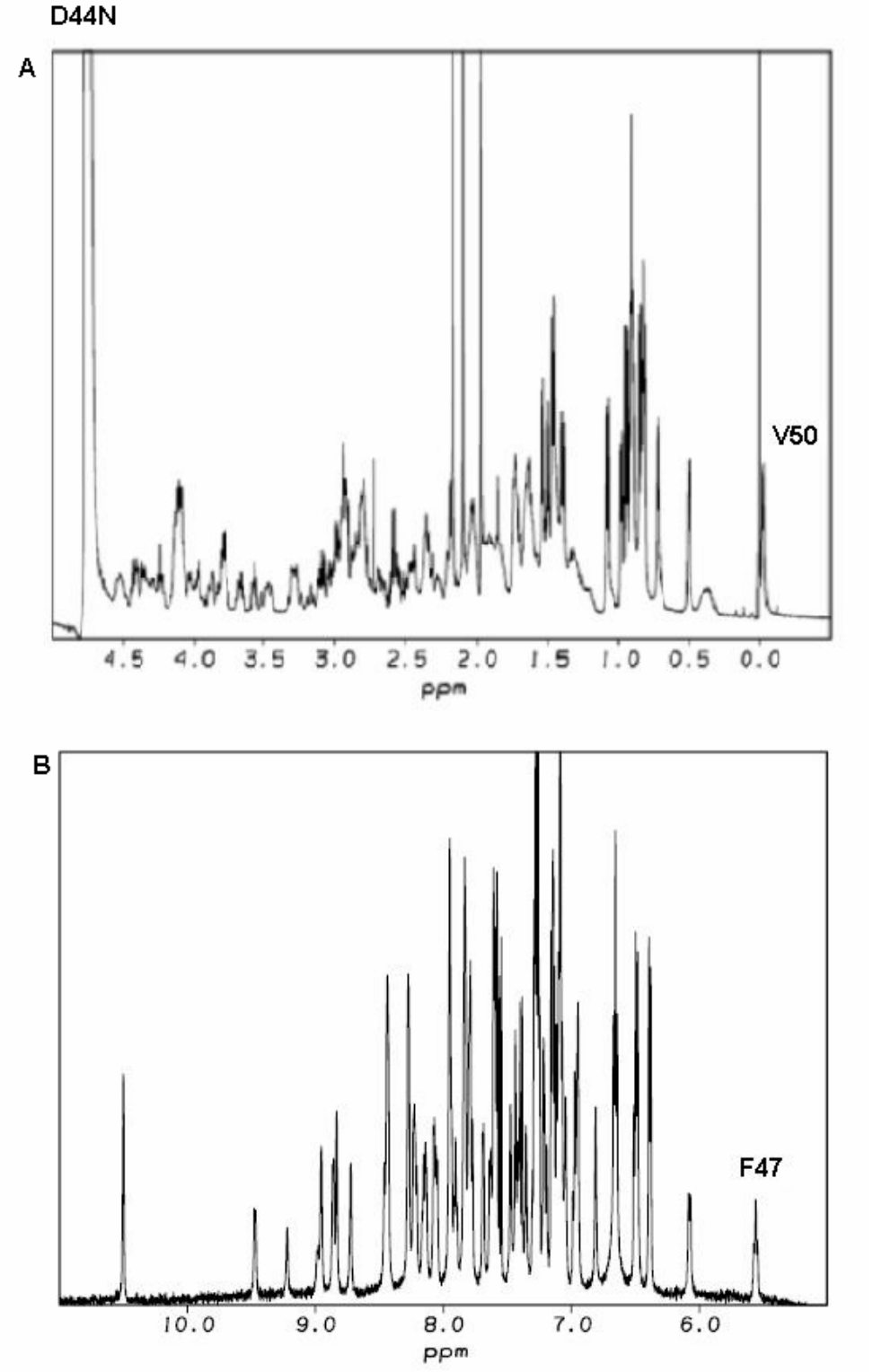




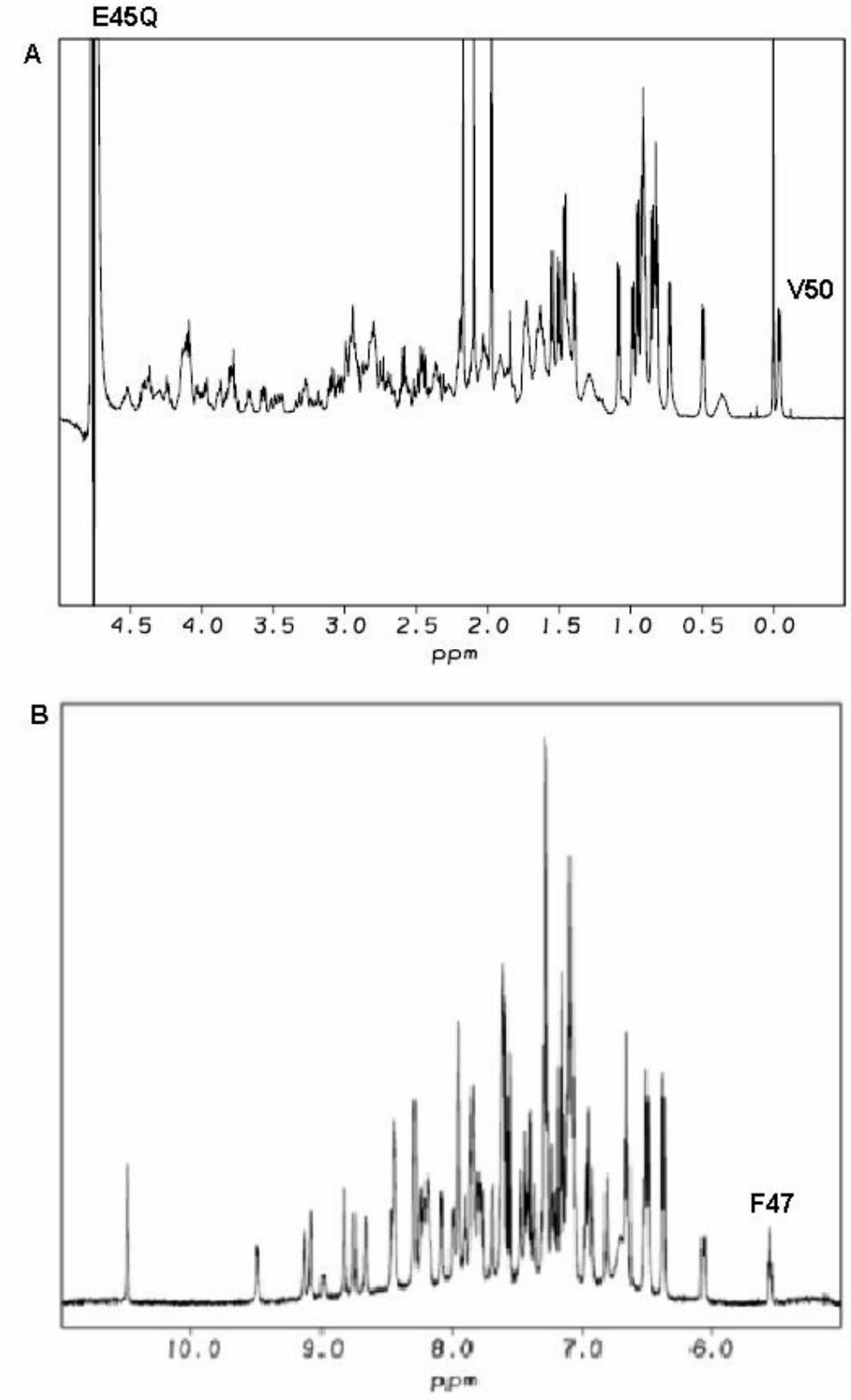




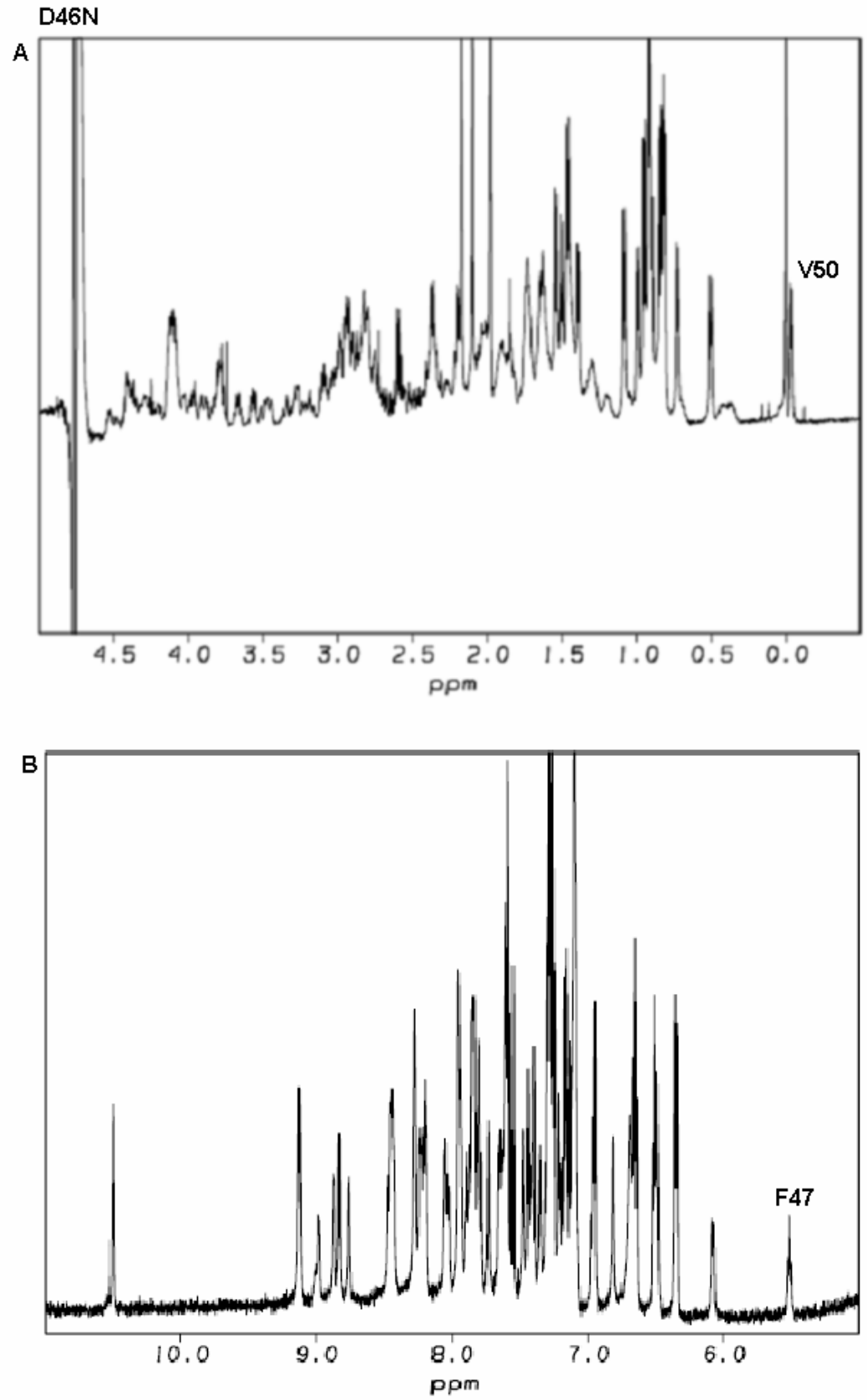
Appendix 2. 1D ^{1}H NMR spectrum of HP36 mutants in 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0 and 25^{o}C. (A) Upfield region (-0.5-5.0 ppm); (B) Downfield region (5.0-11.0 ppm).



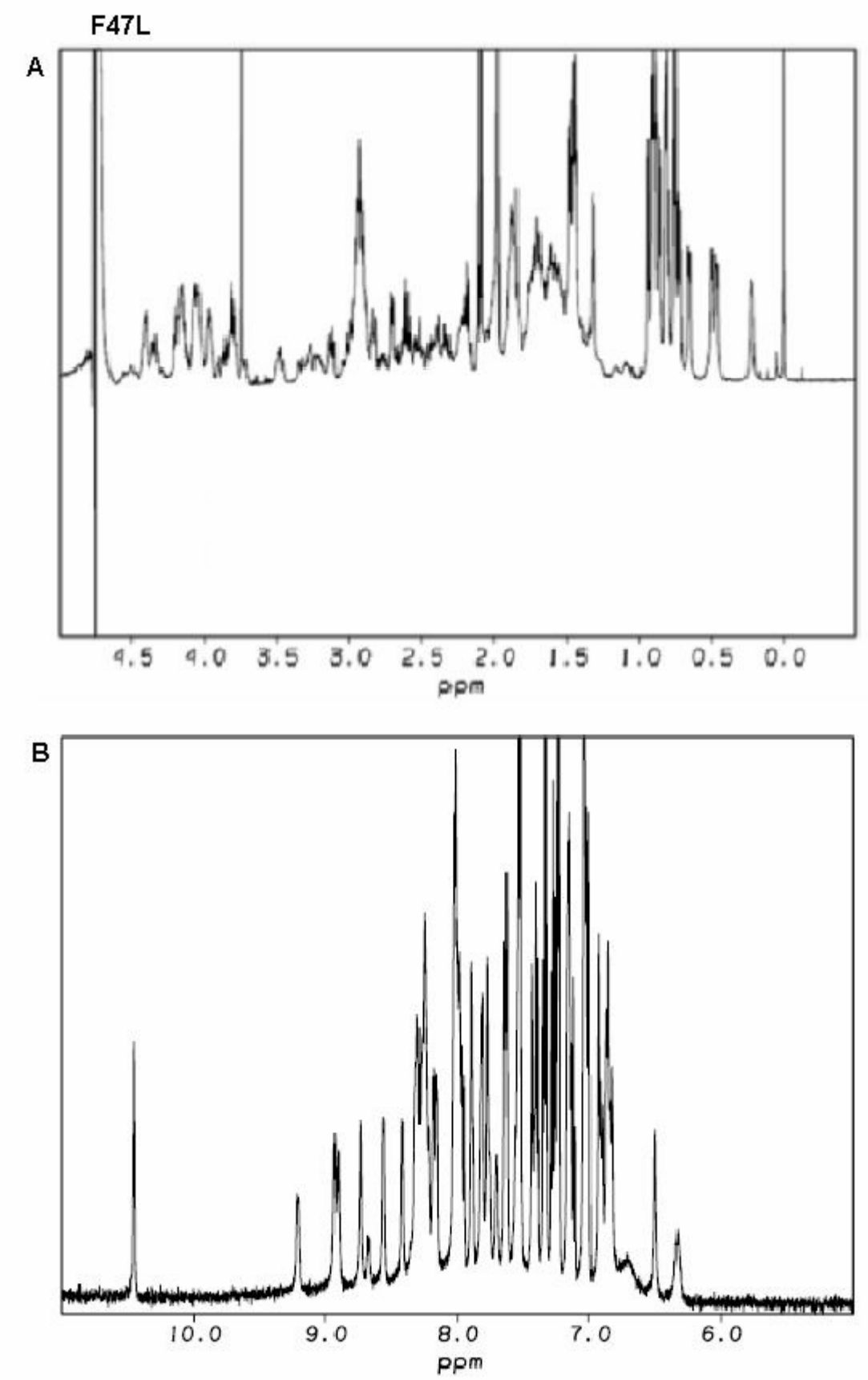
Appendix 2. *continued.*



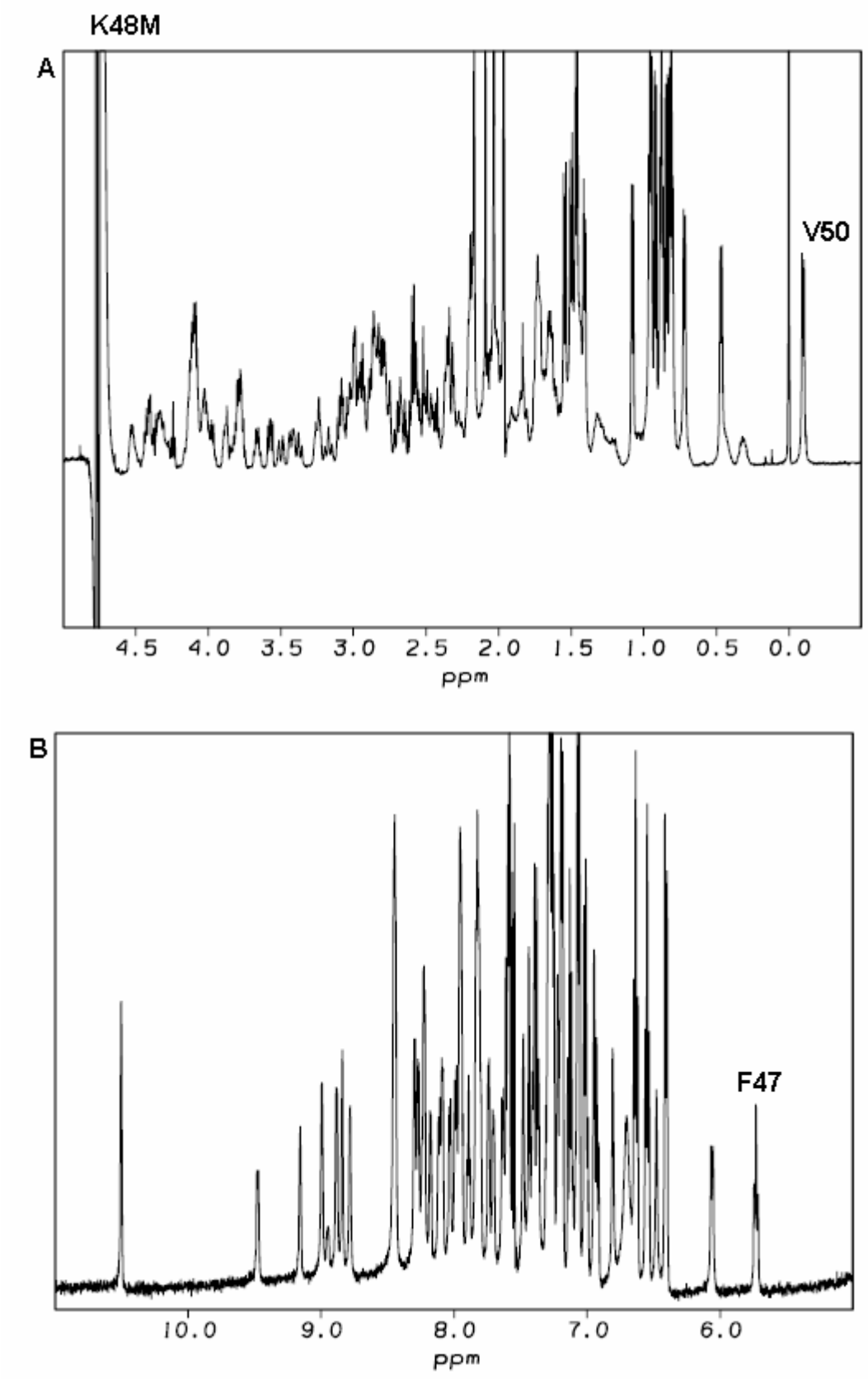
Appendix 2. *continued.*



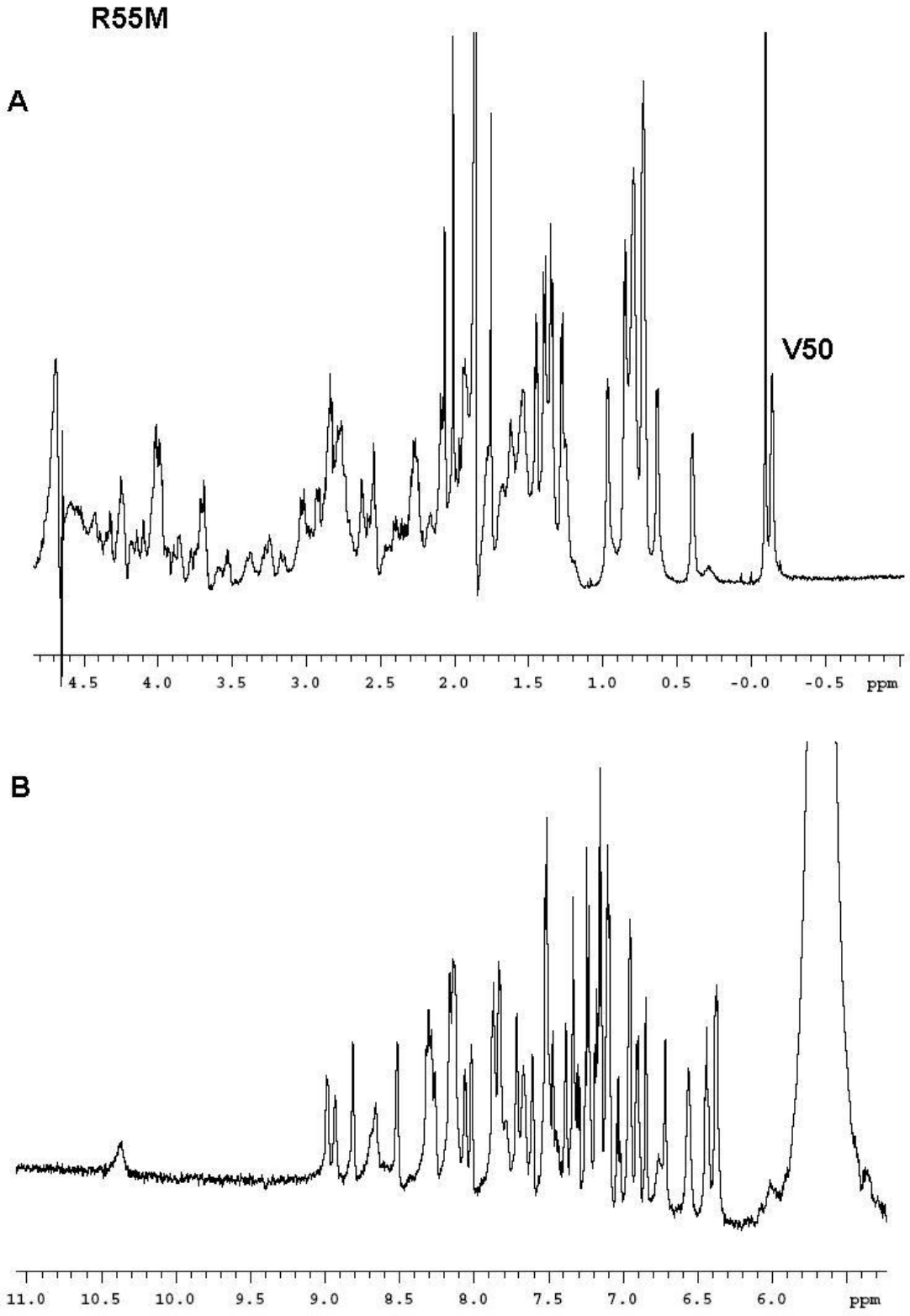
Appendix 2. *continued.*



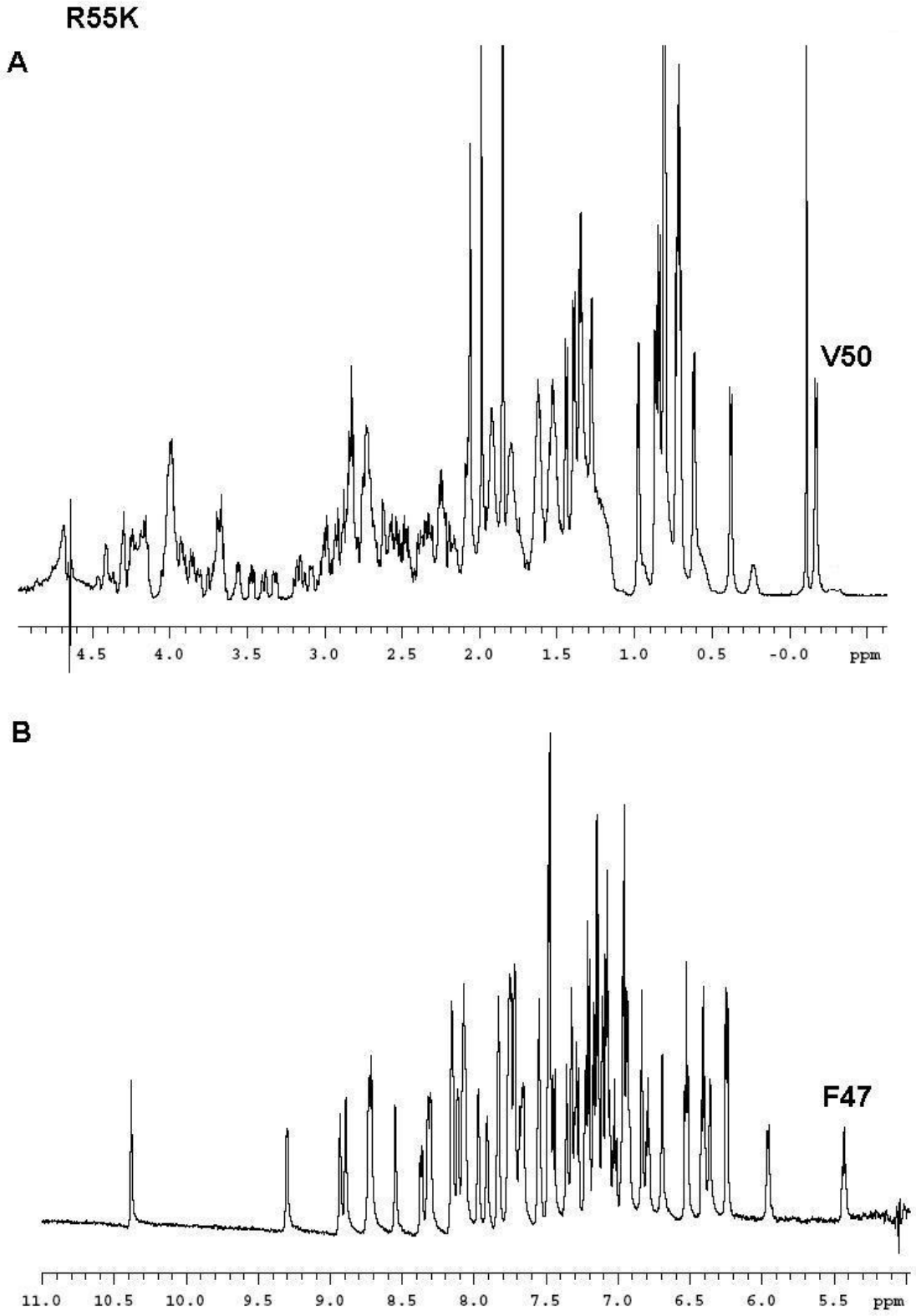
Appendix 2. *continued.*



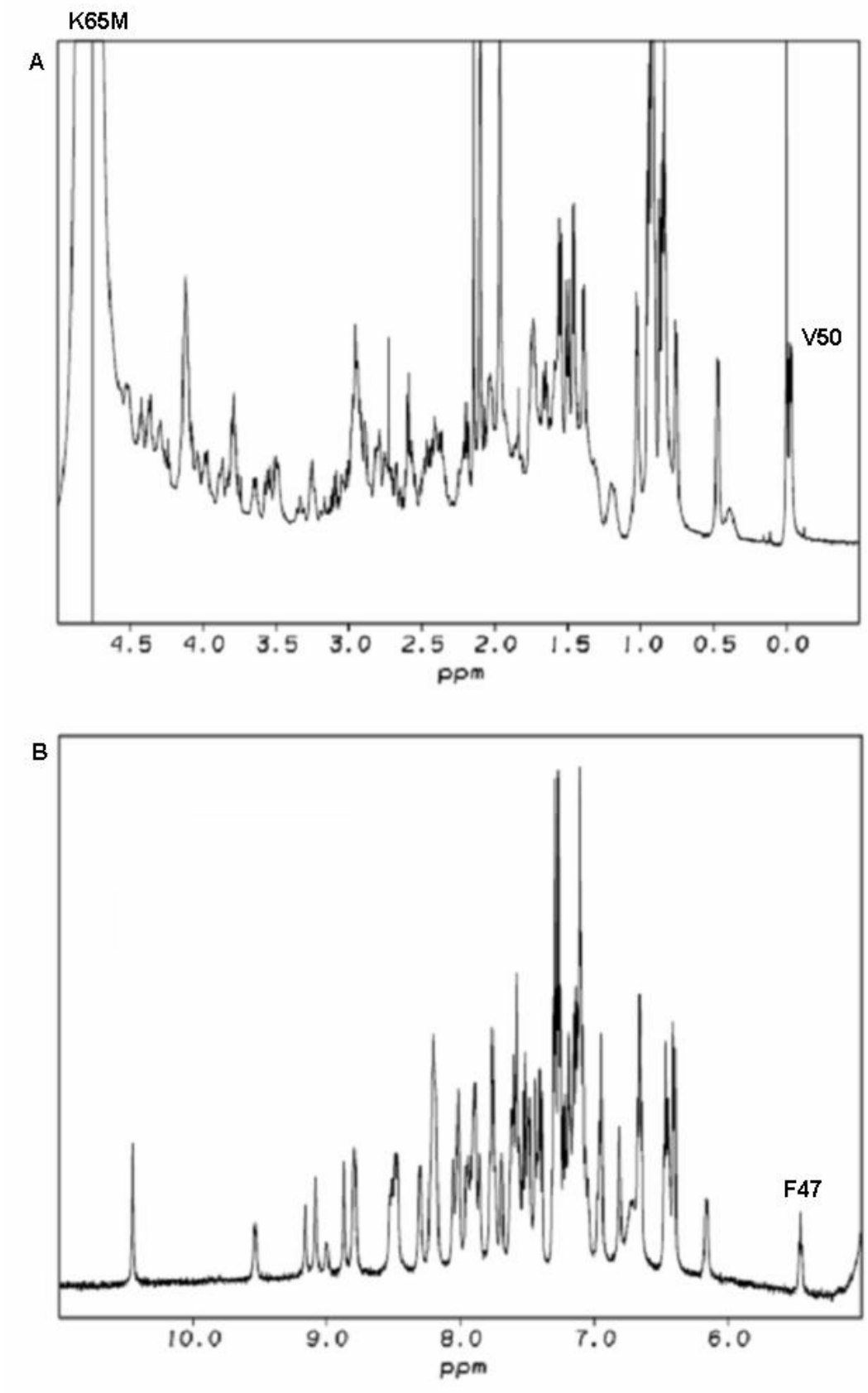
Appendix 2. *continued.*



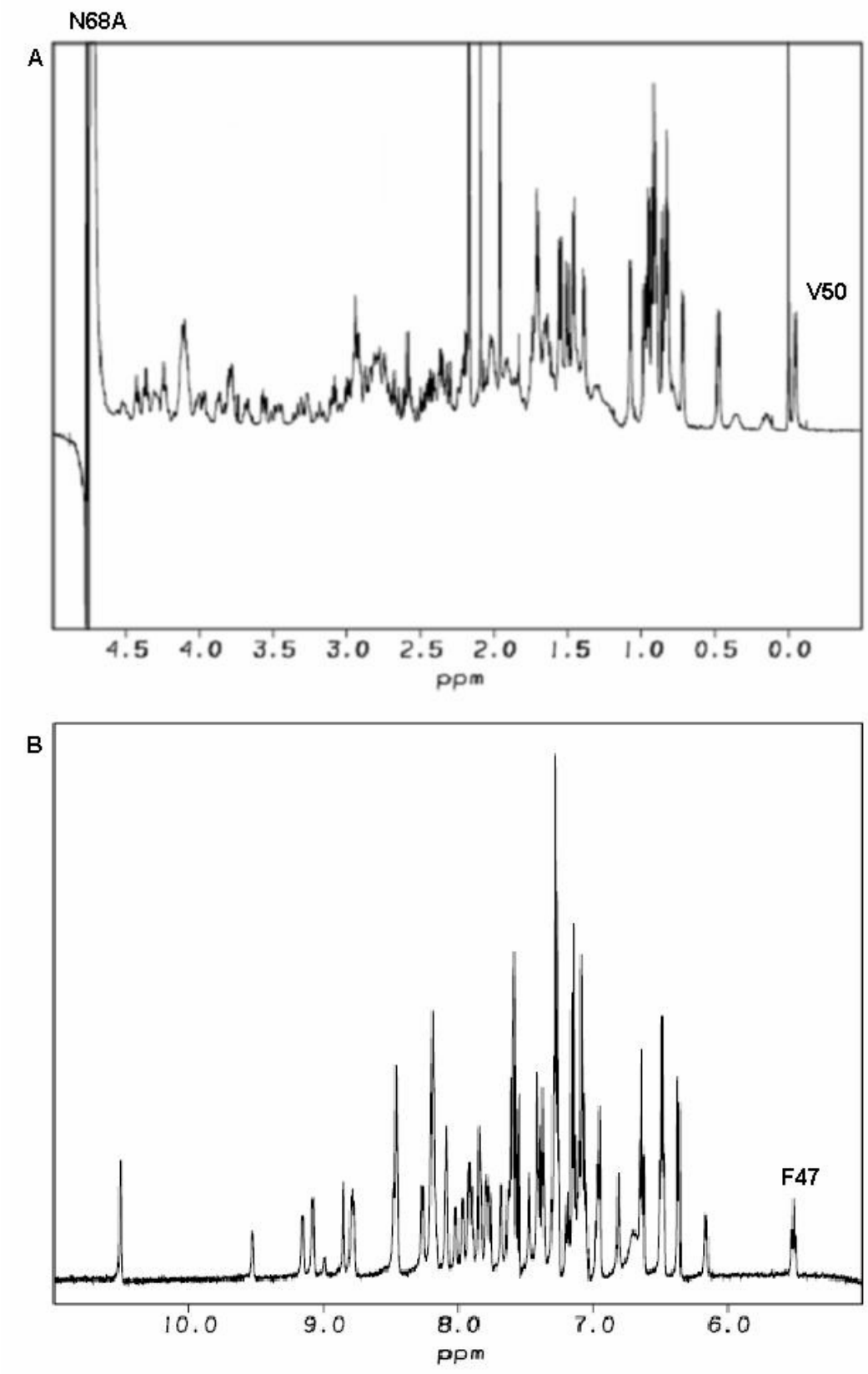
Appendix 2. *continued.*



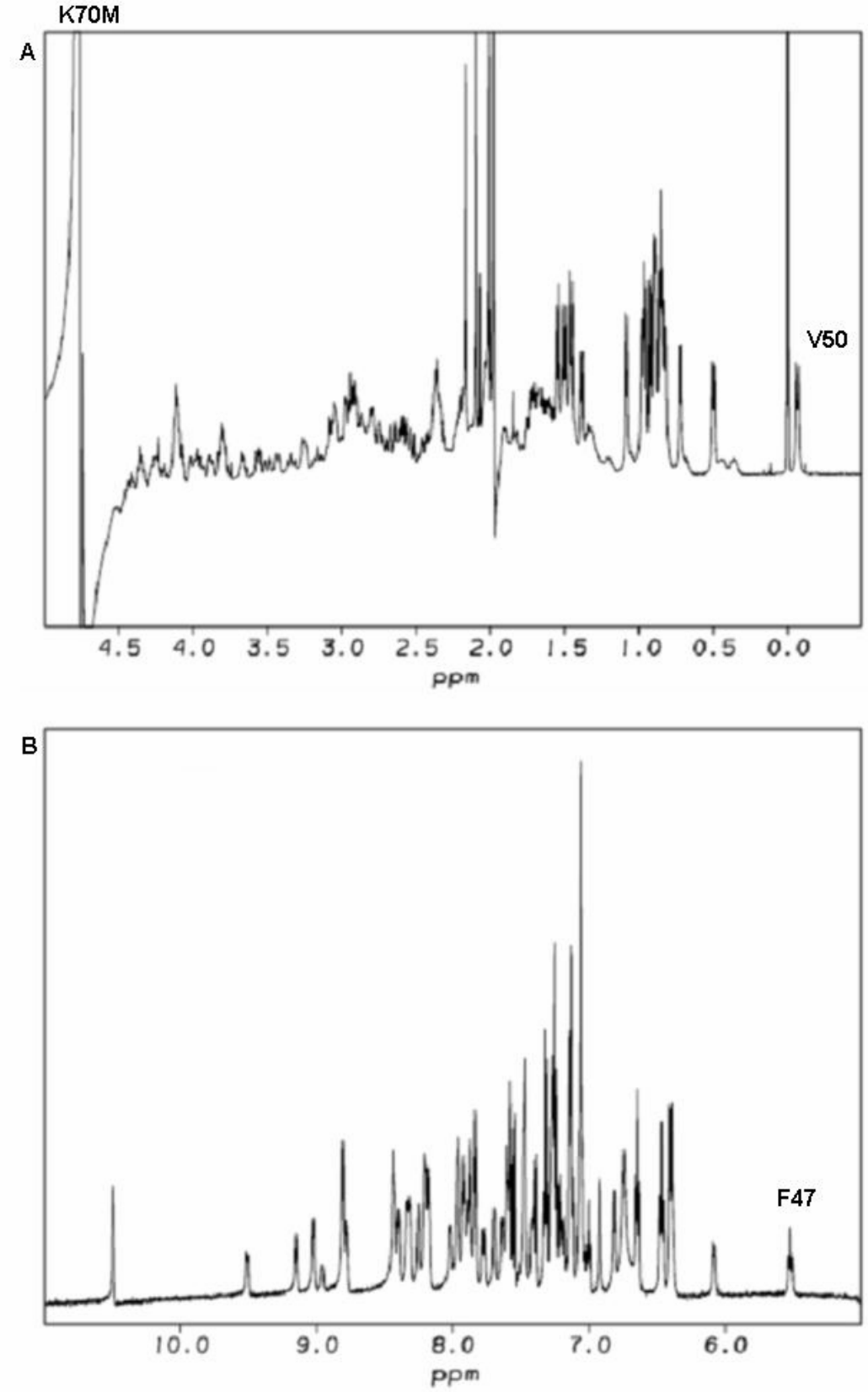
Appendix 2. *continued.*



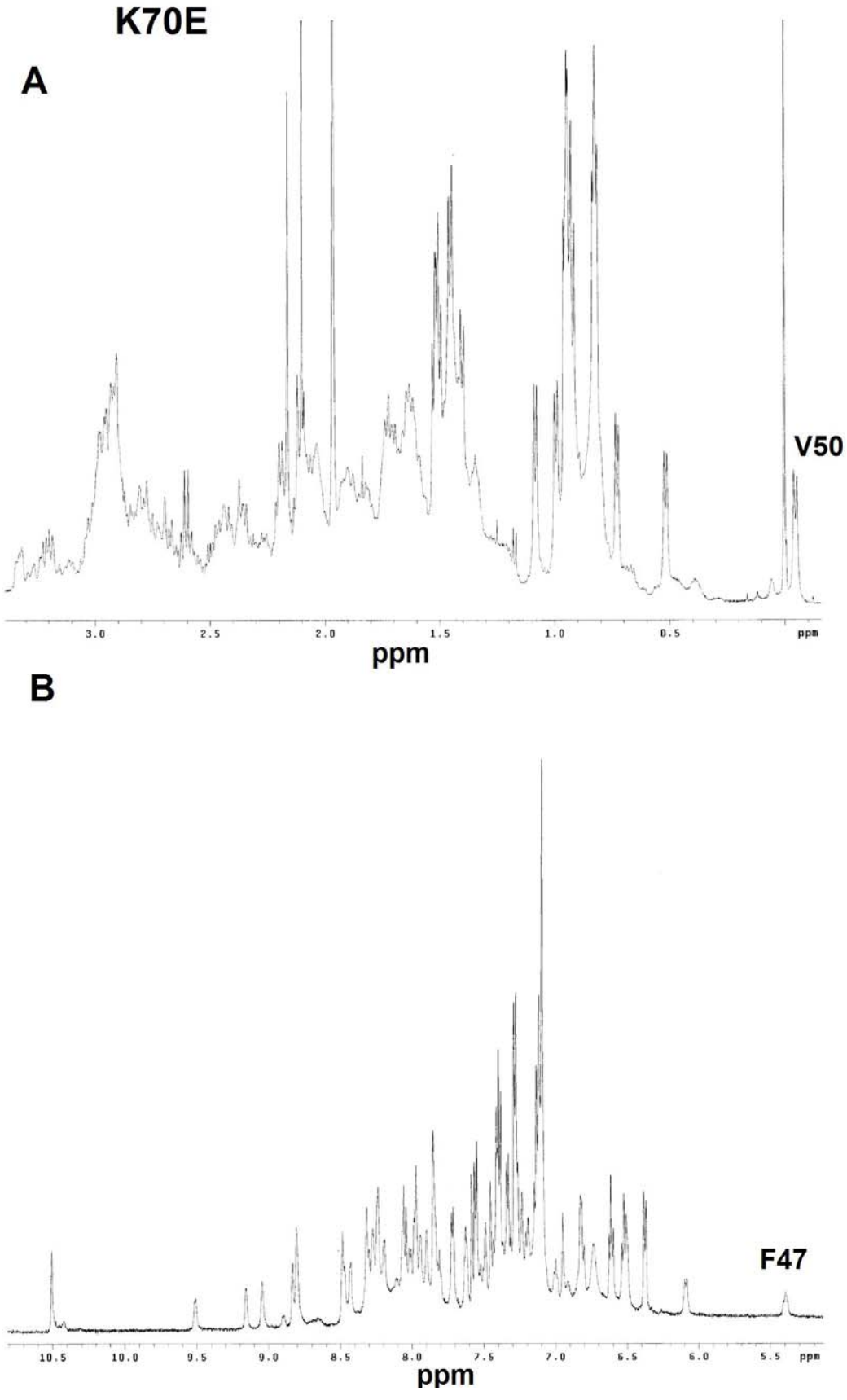
Appendix 2. *continued.*



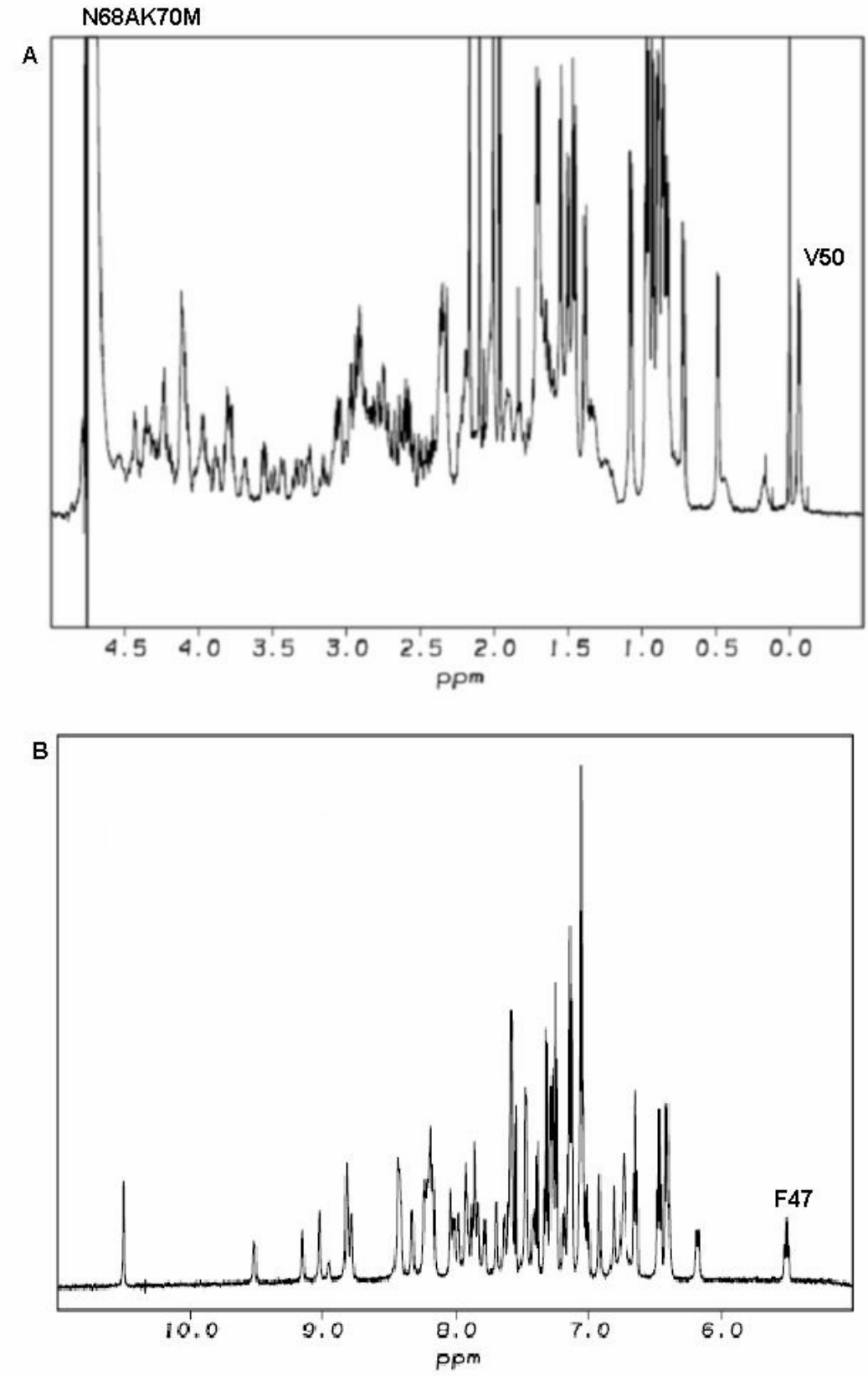
Appendix 2. *continued.*



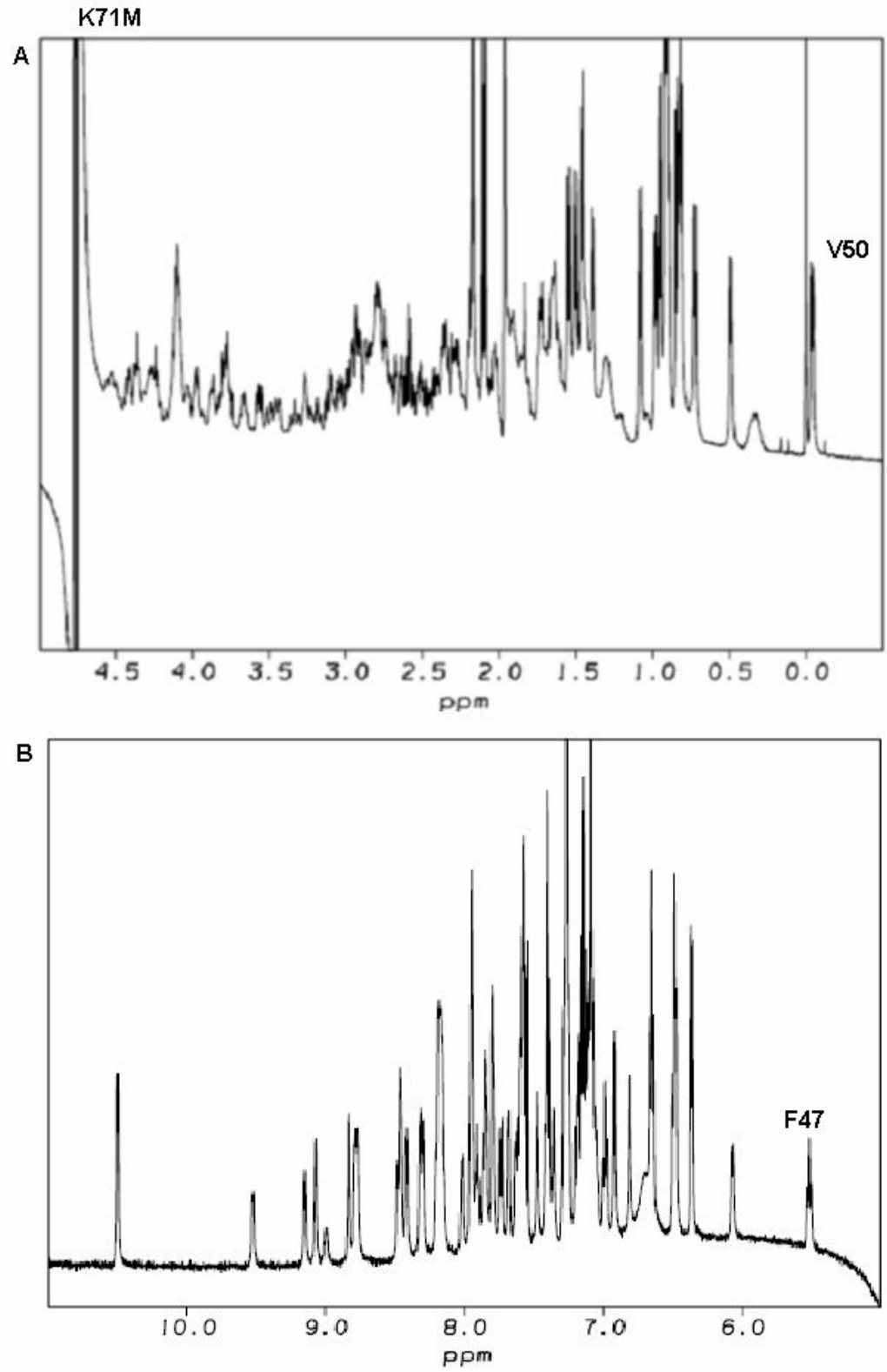
Appendix 2. *continued.*



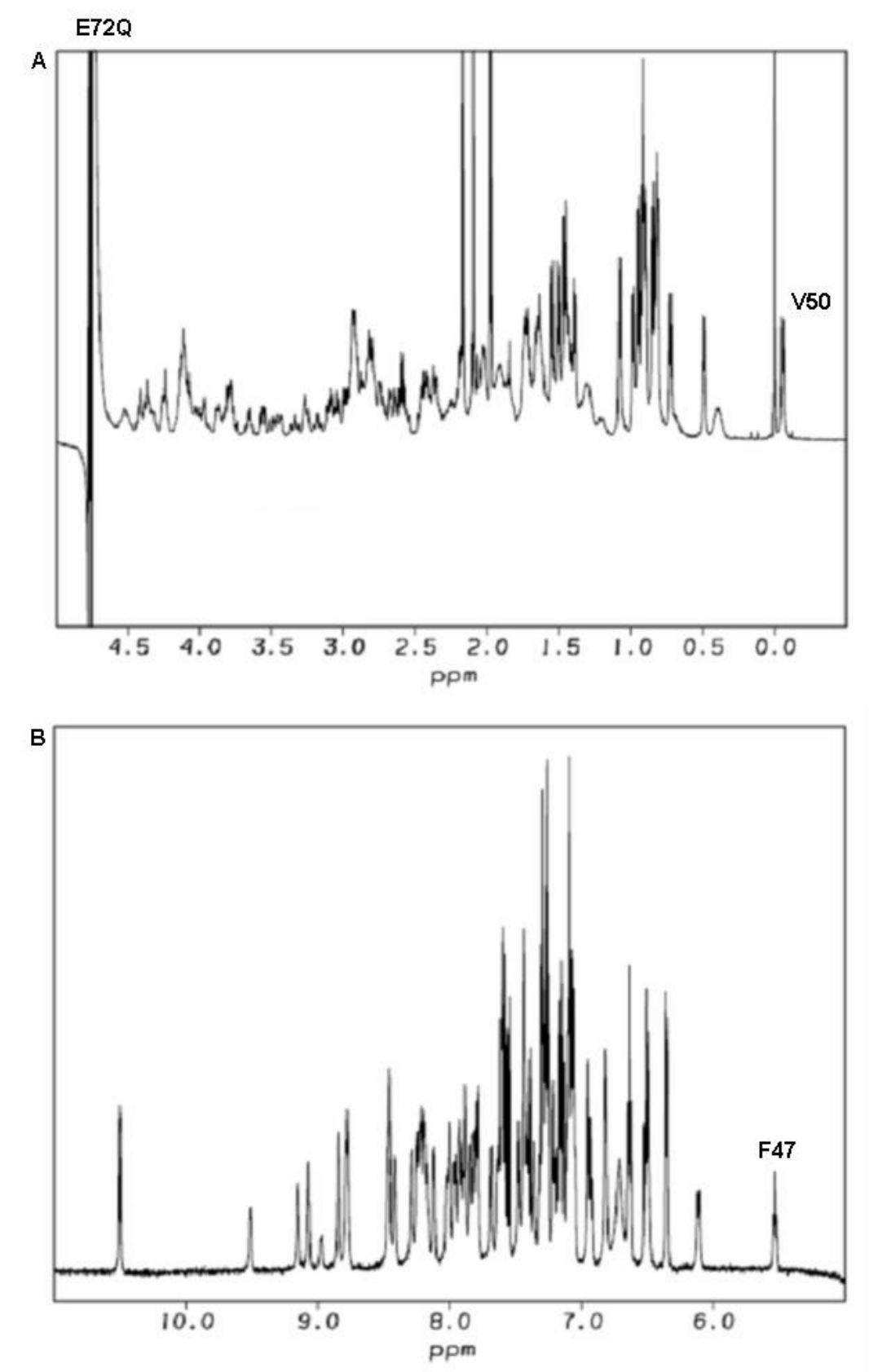
Appendix 2. *continued.*



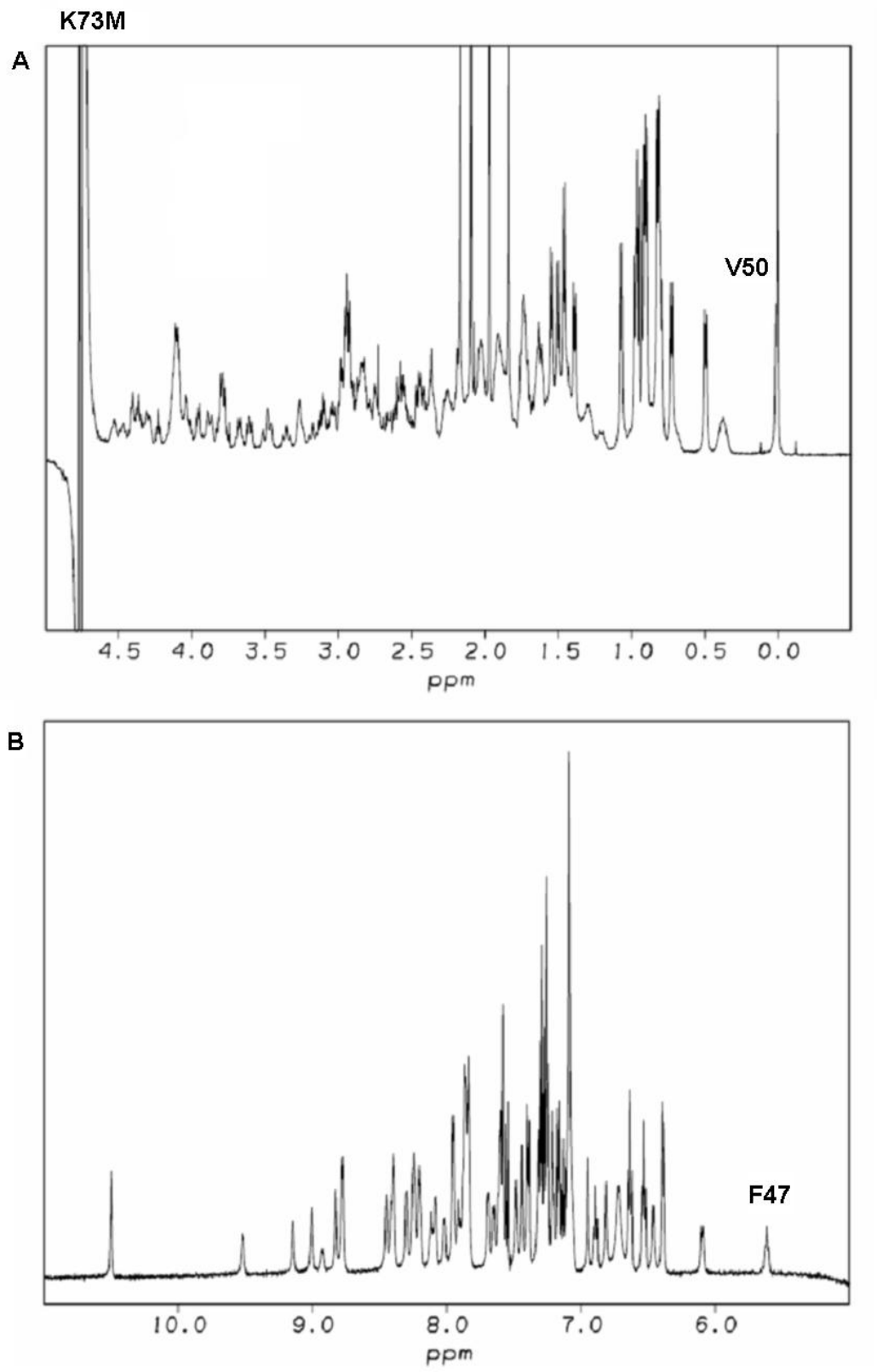
Appendix 2. *continued.*



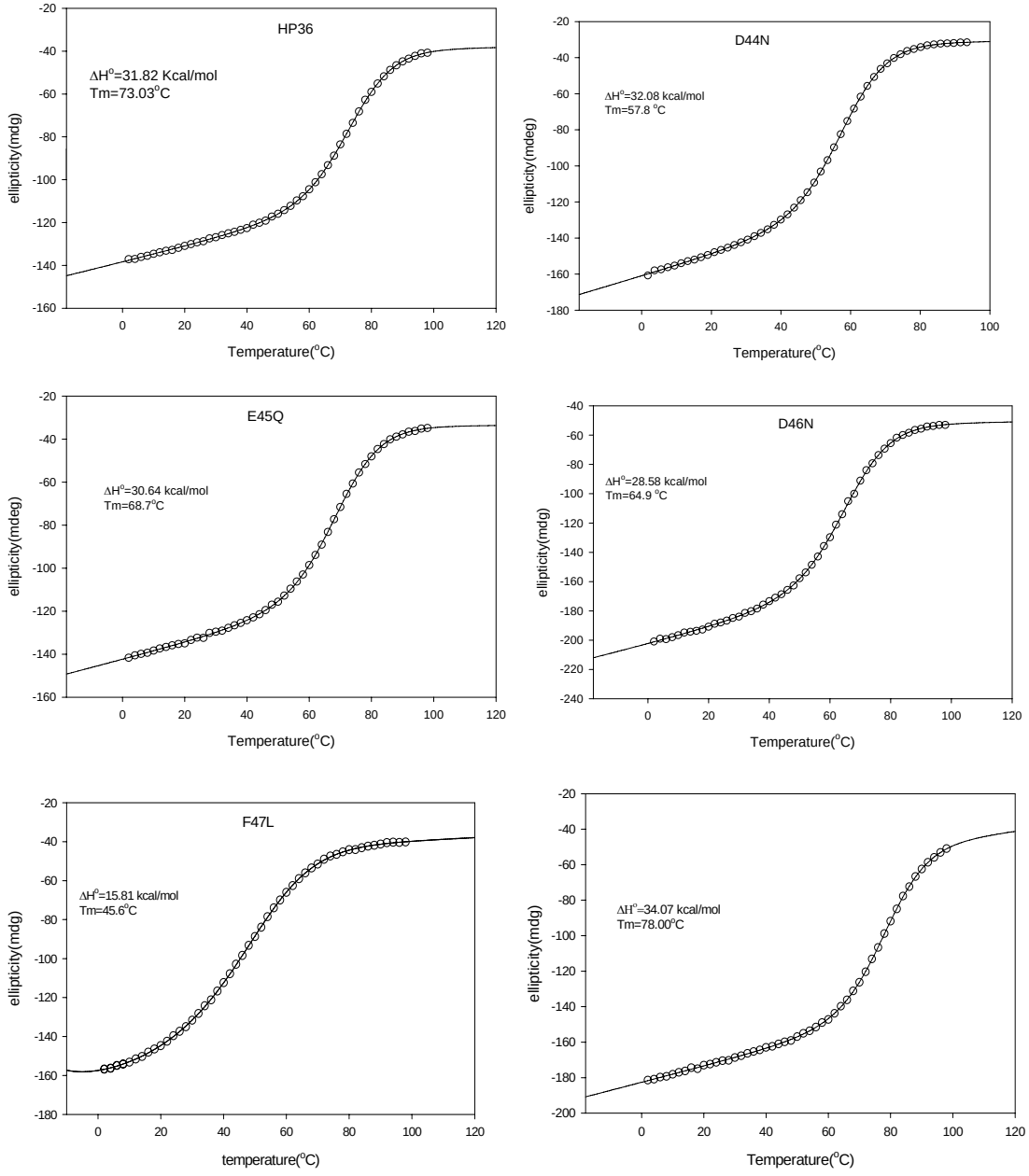
Appendix 2. *continued.*



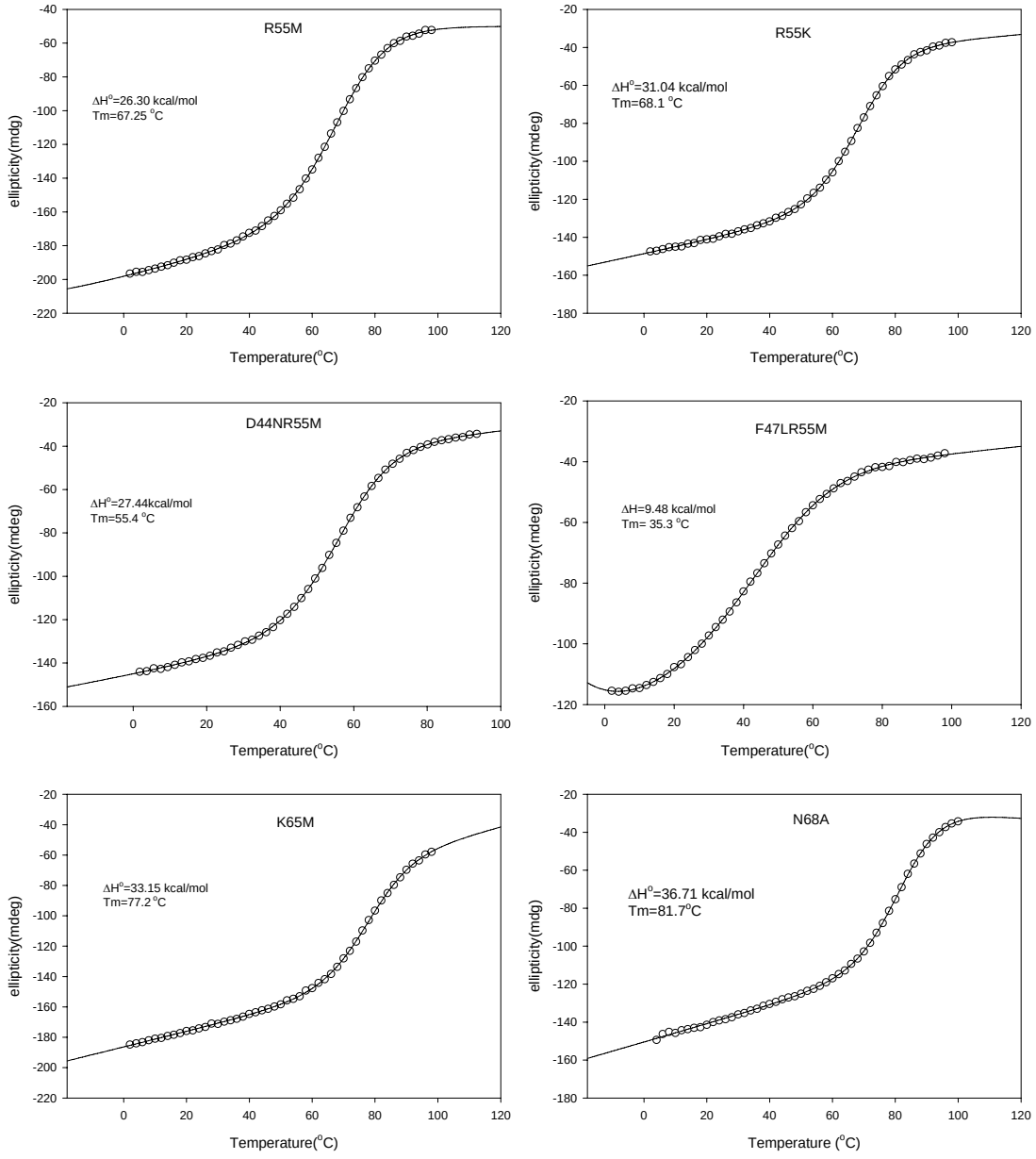
Appendix 2. *continued.*



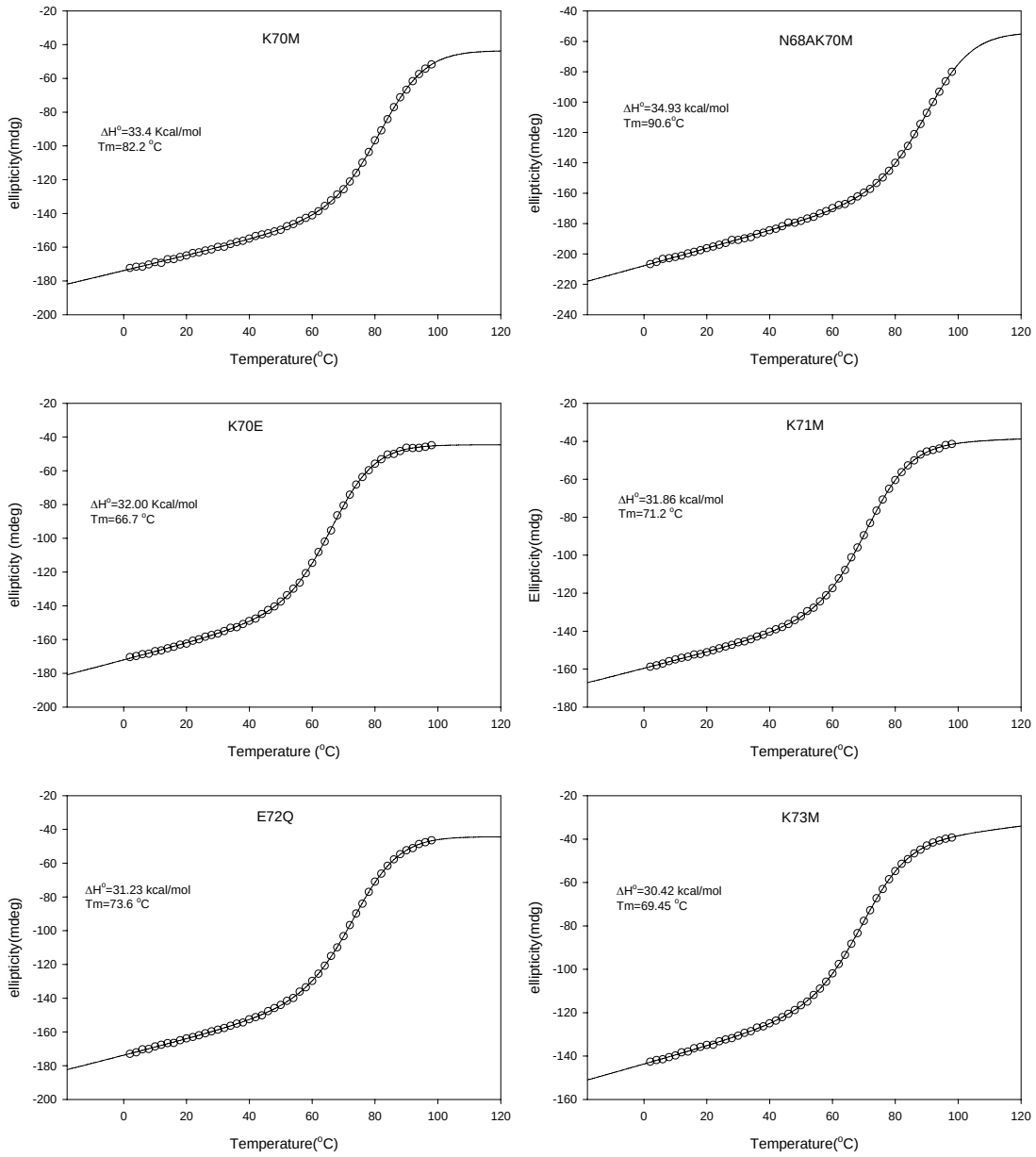
Appendix 3. Thermal unfolding curves for WT HP36 and its mutants in 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0 and 25°C.



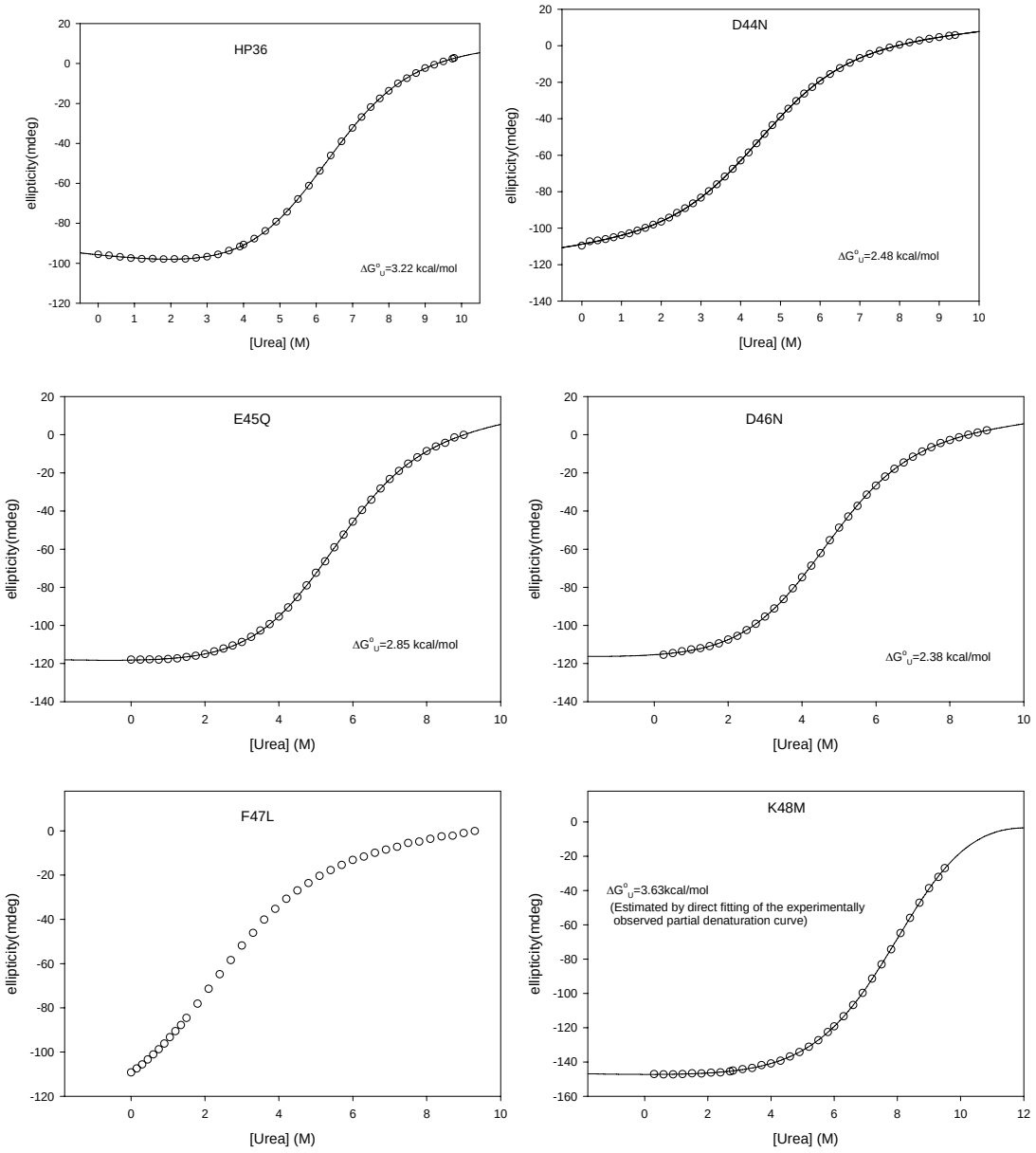
Appendix 3. *continued.*



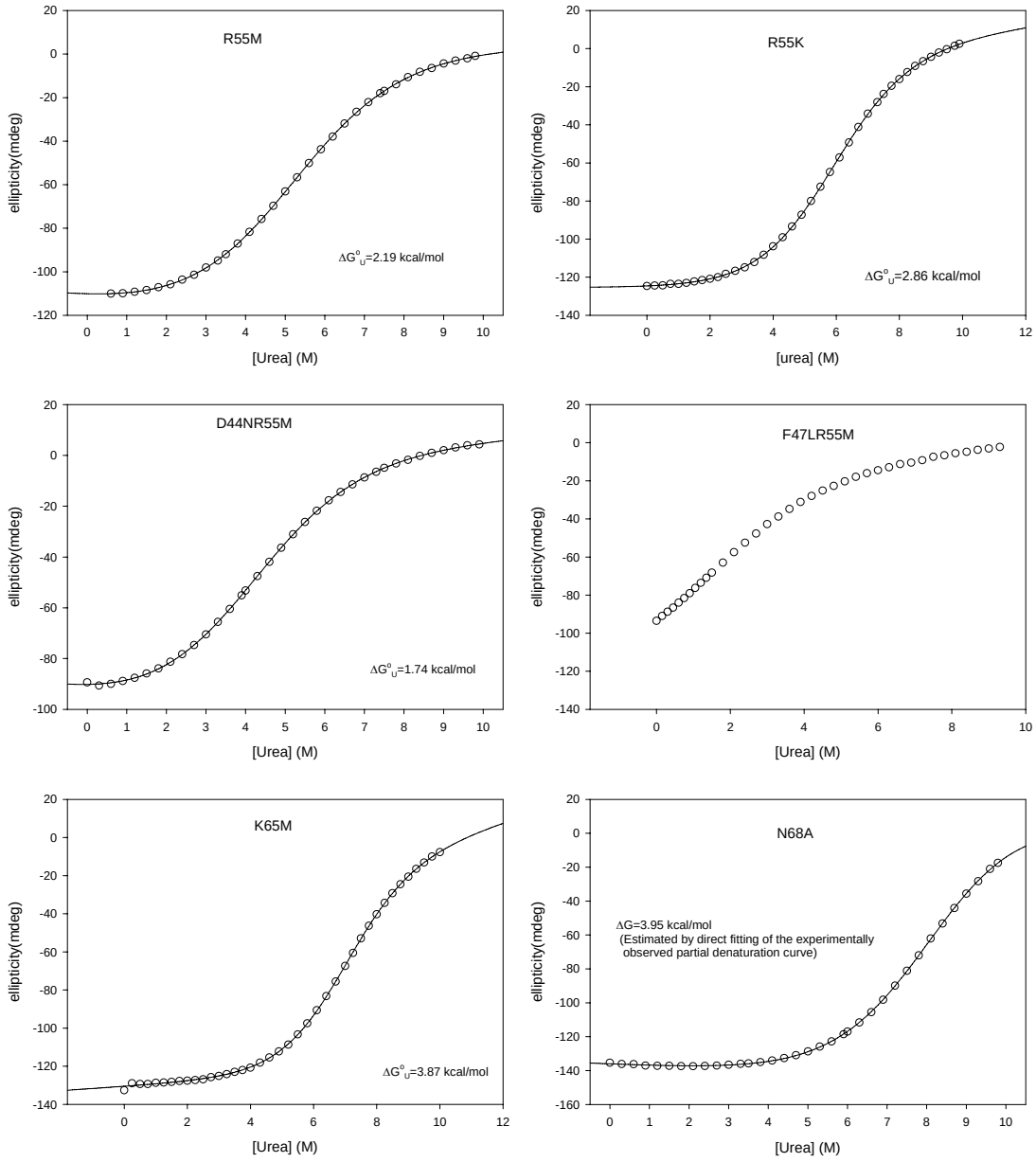
Appendix 3. *continued.*



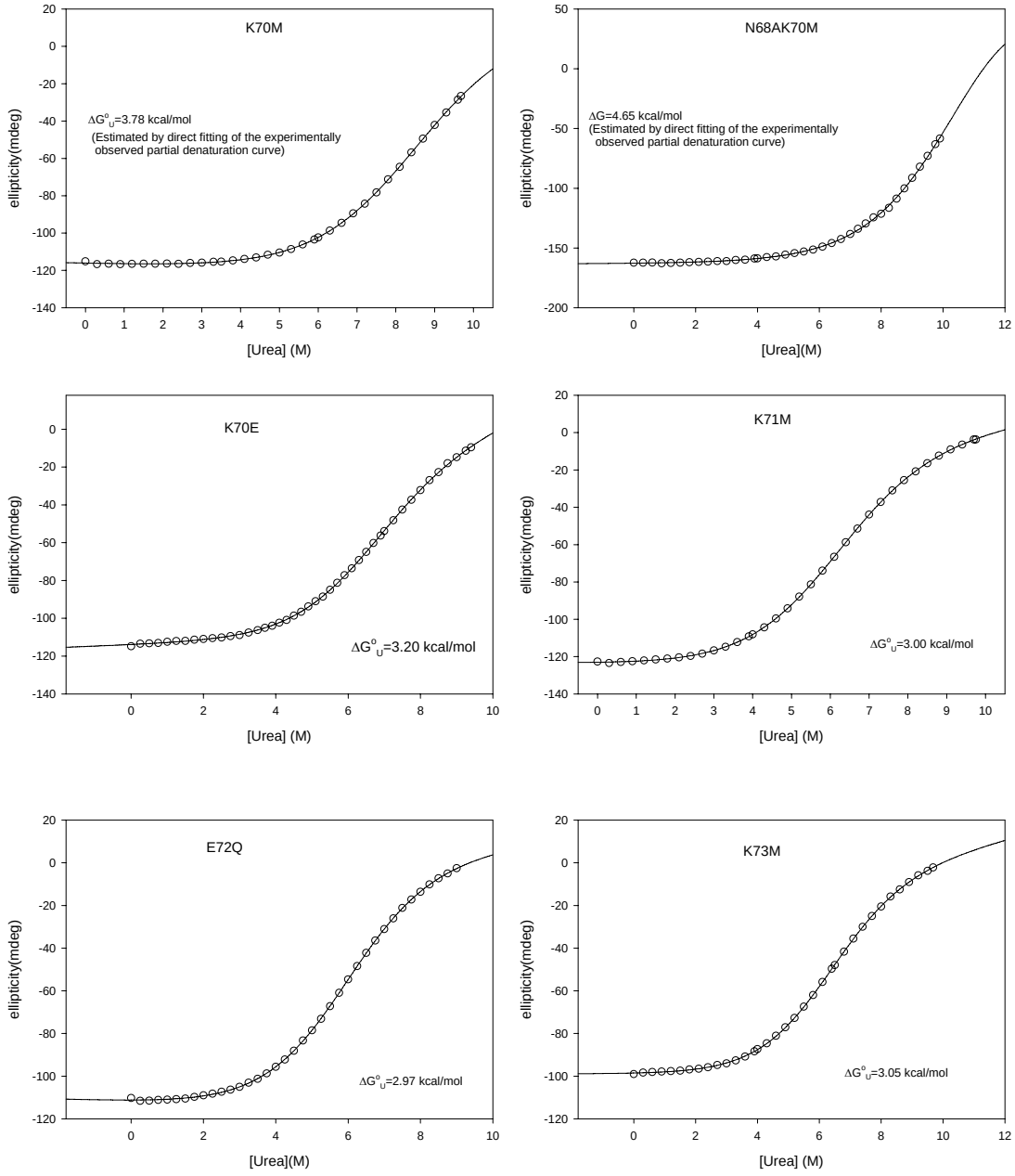
Appendix 4. Urea unfolding curves for WT HP36 and its mutants in 10 mM sodium acetate and 150 mM sodium chloride at pH 5.0 and 25°C.



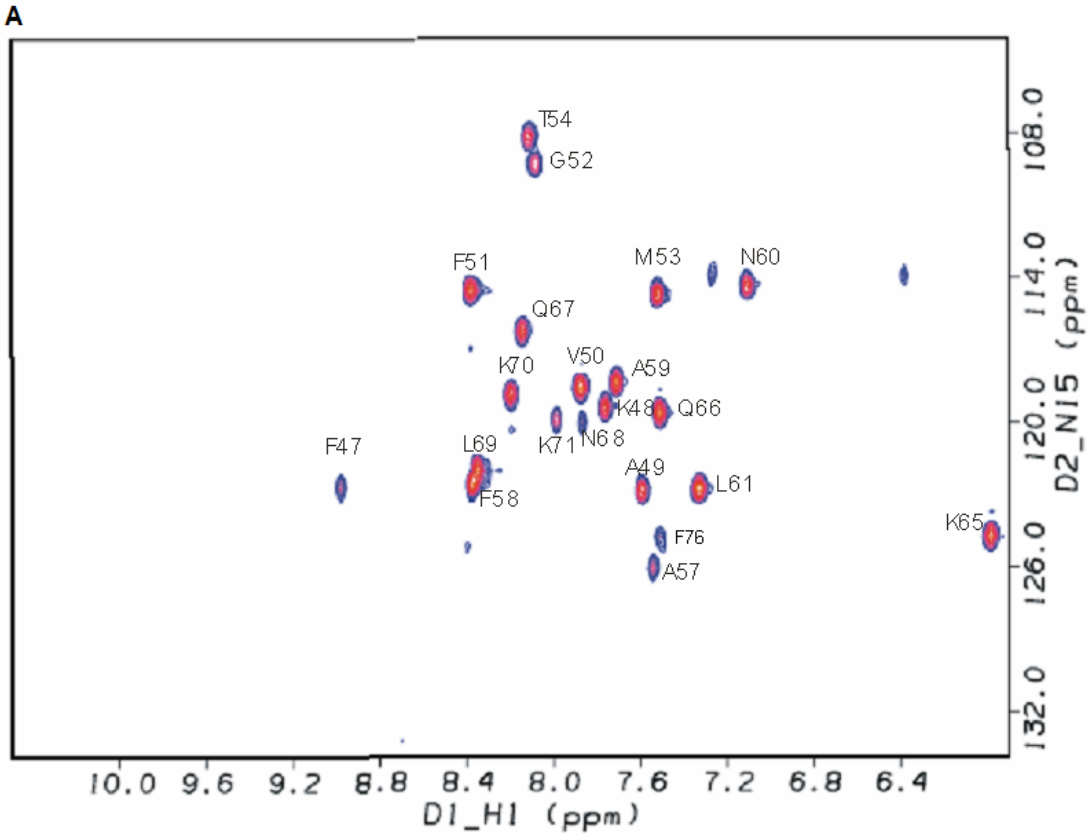
Appendix 4. *continued.*



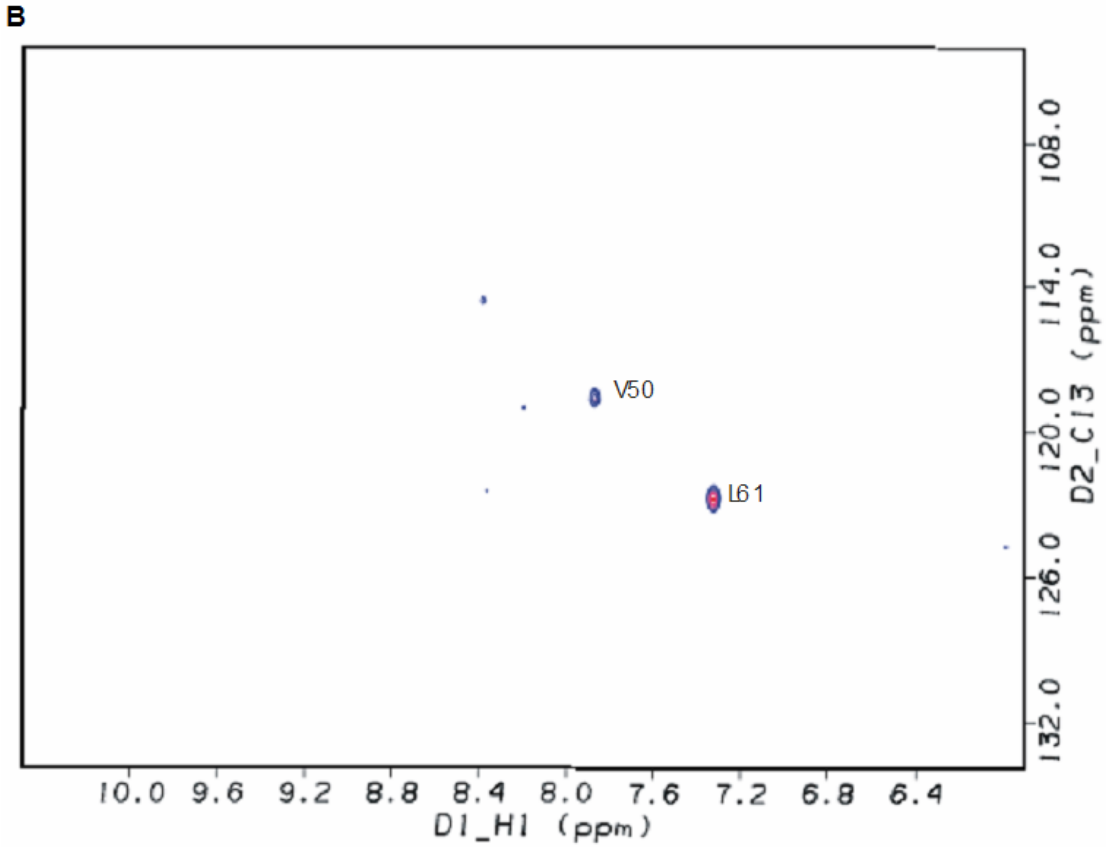
Appendix 4. *continued.*



Appendix 5. Amide H/^{2}H exchange experiments of ^{15}N-labeled HP36 (A) the first HSQC spectrum taken 27 min afer dry protein was mix with $^{2}H_2O$; (B) the HSQC spectrum taken after 6 hours.

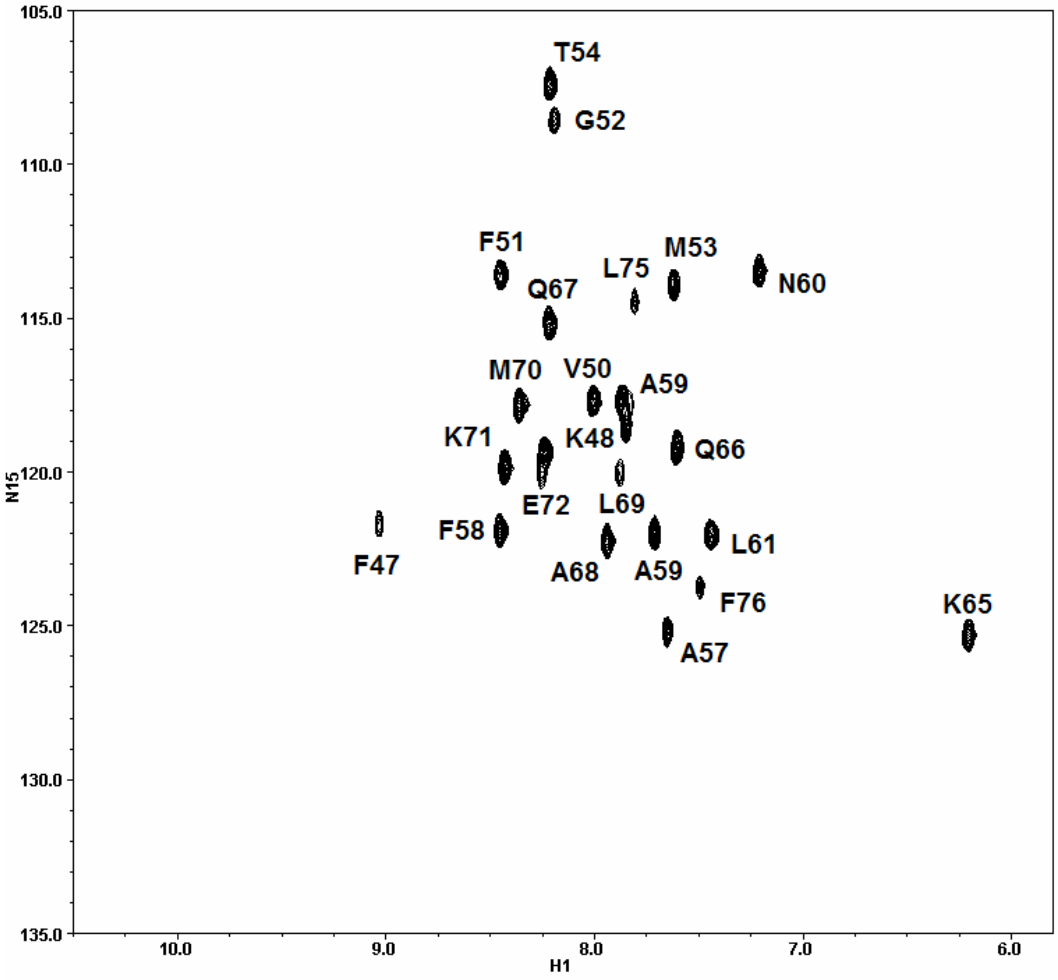


Appendix 5. *continued.*



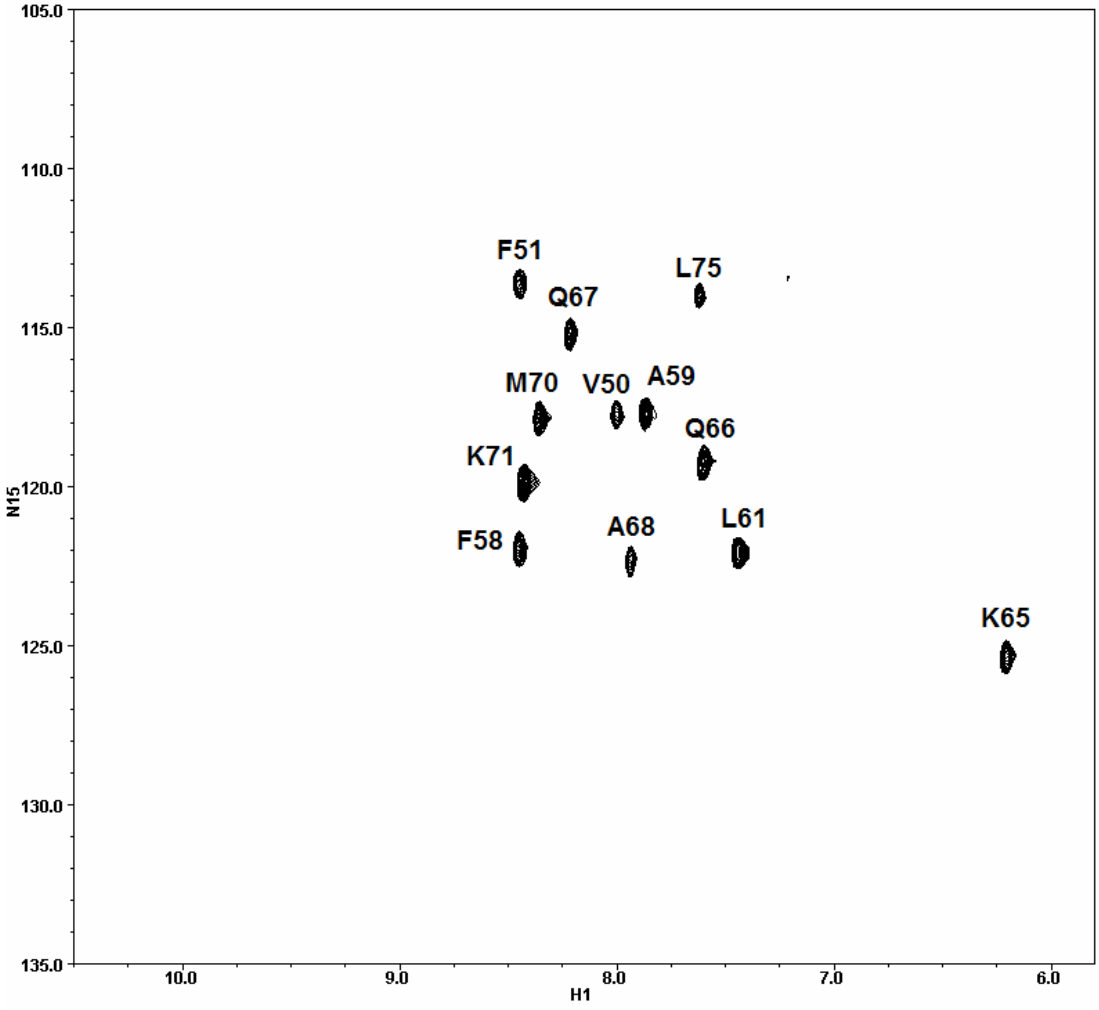
Appendix 6. Amide H/^{2}H exchange experiments of ^{15}N-labeled N68AK70M (A) the first HSQC spectrum taken 24 min afer dry protein was mixed with $^{2}H_2O$; (B) the HSQC spectrum taken after 6 hours; (C) the HSQC spectrum taken after 17 hours.

A



Appendix 6. *continued.*

B



Appendix 6. *continued.*

C