ABSTRACT

Title: MUTATIONAL ANALYSIS OF THE DOWNHILL FOLDING PROTEIN GPW: TOWARDS TUNING STABILITY OF A MOLECULAR RHEOSTAT CANDIDATE

Adam Fung, Ph.D., 2008

Directed By: Associate Professor Victor Muñoz
Department of Chemistry & Biochemistry

A popular convention derived from early experimental evidence of single-domain proteins pointed towards a common mechanism of achieving their native three-dimensional structure. Concomitantly, protein function has been linked to the simple two-state approach wherein the folded state is associated with its biologically active conformation. The more recent discovery of downhill protein folding has provided a sharp contrast to this simple treatment. A catalog of qualitative signatures has been developed to distinguish between these different folding mechanisms at the two extremes (downhill and two-state). Additionally, the introduction of physical models to measure the protein folding ensemble allows the direct measurement of both the thermodynamic and kinetic barrier heights. The end result of such a quantitative approach is a more distinct partition separating the two scenarios. This methodology has been applied to the bacteriophage lambda protein gpW (gene product W) where a clear assessment of its folding behavior has been obtained. GpW is involved at the connector region at which assembly of bacteriophage heads and tails occurs. Without gpW, infectious virions are incapable of forming. This protein is also suspected of having a role in DNA binding and

packaging. Contrary to its expected two-state behavior, gpW folds with a marginal barrier (< 3 RT) as determined through an examination of the thermodynamic and kinetic behaviors. Possessing a novel fold and being an independently folding protein, gpW represents the first experimentally characterized downhill folder that is not a domain of a larger complex and is known to perform a specific function. This is a clear diversion from the standpoint of a single macrostate being responsible for function and places a significant emphasis on investigating the functional role of downhill folding. With the function of performing multiple duties, gpW is an excellent candidate for functional studies of downhill folding.

MUTATIONAL ANALYSIS OF THE DOWNHILL FOLDING PROTEIN GPW:

TOWARDS TUNING STABILITY OF A MOLECULAR RHEOSTAT CANDIDATE

By

Adam Fung

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements of the degree of
Doctor of Philosophy
2008

Advisory Committee:

Associate Professor Victor Muñoz
Professor Dorothy Beckett - Chair
Professor George Lorimer
Assistant Professor Barbara Gerratana
Associate Professor Jeffrey DeStefano – Dean's Representative

Dedication

To Rex

Acknowledgements

I would like to extend my gratitude to Dr. Victor Muñoz for providing me the opportunity to work in his lab and allowing me the chance to develop a great interest in the field of protein folding. The exposure I have gained from him, especially beyond the bench, will shape my outlook throughout my scientific career. He has challenged me in new areas of development and driven me to achieve a higher level of scientific merit and integrity.

My Ph.D. advisory committee deserves a great deal of appreciation for their guidance and willingness to provide their valuable resources, both in the form of lab materials and in scientific knowledge. Dr. George Lorimer was gracious in allowing me to use supplies and equipment as well as an advisory soundboard when I had problems. Dr. Dorothy Beckett could always be counted upon for direction, guidance, as well as great candid conversations and a source of academic and athletic inspiration. Her support in allowing me to use her lab's resources during trying times was invaluable. Dr. Barbara Gerratana is largely responsible for the success of gpW mutant characterizations. Without her guidance and donation, I would still be in front of the HPLC. Each of these professors went beyond their expected commitment. I will forever be grateful for their investment in my success. They have treated me like their peer, and am proud to be a colleague and friend.

Of particular note are the contributions of my colleagues and collaborators – Dr. Peng Li for his contributions in the kinetic experiments, Dr. Athi Naganathan for

the fitting and his continuous computational guidance, Dr. Urmi Doshi for her computational contributions, Drs. Jose Manuel Sanchez-Ruiz and Raquel Godoy-Ruiz from the University of Granada, Spain for performing the calorimetry experiments, Dr. Alan Davidson from the University of Toronto for providing the gpW clone, Dr. Chris Lima from the Sloan-Kettering Institute for the SUMO fusion construct and the ULP1 protease, to Dr. Barbara Gerratana and Ms. Melissa Resto for providing the SUMO and ULP1 clones and their patience through my persistent questions, and finally to Dr. Rob Walker, who's concern for my situation is greatly appreciated, and who also served as an inspiration to me both academically and athletically.

The members of the Muñoz group were an invaluable resource in my development and personal enrichment. In no particular order: Raquel Godoy-Ruiz, Mourad Sadqi, Luis Campos, Peng Li, Urmi Doshi, Athi Naganathan, Jianwei Liu, Tanay Desai, Christina Koshiari, Victoria McParland, Fabiana Oliva, and Raul Perez-Jimenez.

I thank the Graduate School of Maryland and the Department of Chemistry and Biochemistry. Special thanks go to Professor Bryan Eichhorn who, as the admitting chairman in the Department of Chemistry and Biochemistry of my graduate class, took a chance on me and gave me the opportunity to follow my ambitions towards a Ph.D. His confidence in me will never be forgotten. Thanks are also due to the support staff, most notably Paulette Frazier, Levi Gayatao, and Bill Griffin.

To my genuine friends, who are few in number, but colossal in heart. To Rex, my beloved best friend and devoted companion, for putting up with the long hours

alone and weekends following me around the lab. Raquel, who has such a genuine soul and was a source of continuous guidance, friendship and tolerance during our time together in the Muñoz group. Sarah Wehri, whose friendship, council, and cooperation is greatly valued, could always make things better with a cookie. Saranga Naganathan who gained my trust by seeing in me more than what meets the eye. Special thanks go to Yoav, Priscilla, Erin, Julie, and Nikki for their unconditional support, encouragement, motivation, and friendship. And finally, I give thanks to my cycling team, DC Velo, along with my friends in the cycling community for nurturing my outlet to maintain sanity.

Table of Contents

Dedication ii

Acknowledgements iii

Table of Contents vi

List of Figures ix

List of Tables x

1. Introduction and Research Aims

1.1 Introduction 1

1.1.1 The Free Energy Landscape Perspective 2

1.2 Protein Folding Trajectories 4

1.2.1 The Two State Approximation 6

1.2.2 Downhill Protein Folding 9

1.3 Mutational Approaches in Protein Folding 14

1.3.1 Alanine Scanning 15

1.3.2 Phi Value Analysis 18

1.3.3 Psi Value Analysis 22

1.4 Metal Binding 25

1.5 Research Aims and Chapter Summary 27

2. Materials and Methods

2.1 Cloning and Protein Expression 30

2.2 Purification 34

2.3 Determining Protein Concentration 36

2.4 Circular Dichroism 37

2.5 Fluorescence 41

2.6 Fourier Transform Infrared Spectroscopy 43
2.7 Differential Scanning Calorimetry 44
2.8 Time Resolved Infrared Temperature Jump Kinetics 47

3. Multiprobe Characterization of gpW

3.1 Introduction to gpW 49
3.2 Coupled Denaturant Unfolding 56
3.3 Backbone Interactions in gpW Folding 63
3.3.1 Far-UV Circular Dichroism 63
3.3.2 Fourier Transform Infrared Spectroscopy 68
3.4 Probing Global Structure 72
3.4.2 Fluorescence 75
3.5 Spectroscopic Characterization of Bi-Histidine Mutants 78
3.6 Conclusions 82

4. Measurement of Folding Barrier Height by Calorimetry

4.1 Introduction 86
4.2 DSC Data Described by a Two-State Approximation 86
4.2.1 Two-State Criterion 88
4.2.3 Importance of Baseline Determination 89
4.3 Variable Barrier Model 90
4.4 Direct Estimate of gpW Free Energy Barriers 95
4.5 Conclusions 105

5. Kinetic Characterization of gpW

5.1 Introduction 109
5.2 One-Dimensional Free Energy Surface (FES) Kinetic Model 111
5.2.1 Free Energy Barrier Heights from Kinetics 114

5.2.2 Robustness of 1-D FES Model 115
5.3 gpW Infrared T-Jump Kinetics 117
5.4 Conclusions 128

6. gpW Functional Perspectives
6.1 Introduction 130
6.2 Mature Virions of Bacteriophage λ 131
6.3 λ Head-to-Tail Assembly 134
6.4 Molecular Rheostats 135
6.5 Tuning Stability of gpW Bi-His Mutants 137
6.6 Conclusions 141

7. Summary and Future Directives 144

8. Bibliography

List of Figures:

Figure	Description	Page Number
1.1	Two-state free energy reaction coordinate	7
1.2	Representative free energy landscape for the downhill folding protein BBL	11
3.1	Structure of gpW	53
3.2	Contact map of gpW	54
3.3	Double perturbation experiment of WT gpW	62
3.4	Unfolding enthalpy as a function of denaturant	63
3.5	f-CD temperature/wavelength experiments of WT gpW	68
3.6	SVD analysis of f-CD	70
3.7	SVD analysis of FTIR	73
3.8	n-CD temperature/wavelength experiments of WT gpW	76
3.9	SVD analysis of n-CD	77
3.10	Temperature dependent fluorescence spectra	79
3.11	SVD analysis of fluorescence	80
3.12	First derivative and CD unfolding curves	85
3.13	Summary of multi-probe unfolding curves	87
4.1	DSC thermogram of WT gpW	100
4.2	One-dimension free energy surface extracted from variable barrier model of DSC thermogram	102
4.3	Probability distribution of WT gpW from DSC thermogram fit to variable barrier model	104
5.1	IR T-jump relaxation decays near thermal melting temperature	117
5.2	Relaxation rates as a function of temperature fit to one dimensional free energy surface model	117
5.3	Kinetic amplitude of WT gpW fit to 1-D FES model	118
5.4	Simulated kinetic data from 1-D FES model fit	118
5.5	Kinetic amplitudes, relaxation rates as a function of temperature, and one-dimensional free energy landscapes for all helical bi-his mutants	121
5.6	Kinetic amplitudes, relaxation rates as a function of temperature, and one-dimensional free energy landscapes for all β-strand bi-his mutants	123
6.1	gpW bi-his mutant constructs	133
6.2	Thermal stability of bi-histidine mutants upon metal binding	141

List of Tables:

Table	Description	Page Number
4.1	Fitting parameters obtained from WT gpW DSC data fit to variable barrier model	103
5.1	Comparison of melting temperatures, relaxation times, and barrier heights of WT gpW and all bi-his mutants	120

1. Introduction and Research Aims

1.1 Introduction

The modern advent of the protein folding field began with a simple thought experiment, aptly referred to as Levinthal's paradox, questioning the driving forces governing the energetic minimization of a polypeptide chain. Proteins, being large macromolecular structures, are composed many numbers of atoms spanning a range of several orders of magnitude. Each of these atoms is ultimately restricted to a unique three dimensional conformation. With the seemingly infinite number of degrees of freedom, the arrival of a newly synthesized polypeptide to its native conformation was estimated to take longer than the age of the universe (Levinthal, C. 1968). From a functional perspective of a simple two-state system, the native state is the functionally active state. For a protein to exert its function there must be a common driving force which compels the energetic bias towards an energy minimum. In the case of a simple two state system, a global minimum is achieved.

The evolution of Levinthal's paradox from theory to experiment took a leap forward in the course of Anfinsen's work with ribonuclease A (RNase A) (Anfinsen, C. B. 1973). Upon unfolding and disrupting catalytic activity via chemical denaturation, RNase A was discovered to fully recover its three-dimensional structure and function upon removal of the denaturing agents. It was then postulated that the diffusional search across the immense three-dimensional space available was not

randomly sampled, but instead, all of the information necessary to prejudice the polypeptide towards its native configuration was contained in the amino acid sequence. This is a rather intuitive conclusion given that protein folding proceeds over several orders of magnitude commonly on the order of microseconds to seconds. Therefore, local interactions are likely to be the major influence directing the unfolding / folding equilibrium towards the ordered state in native conditions. This feature is especially striking given the consideration that amino acids vary in their physical and chemical properties, yet primary sequence alterations can still lead to similar secondary and tertiary structures as well as sharing functional roles among evolutionary descendents. This suggests that Natural Selection favors a particular overall configuration despite differences in chemical composition. The end result is that through this selective process, the energetic and kinetic restraints are respected and conserved.

1.1.1 The Free Energy Landscape Perspective

The general theme of evolutionary selection has been further refined and recapitulated by means of the "principal of minimal frustration" by Wolynes and co-workers (Bryngelson J. D. 1995, Onuchic J. N. 1997). Beyond the view of pathways proposed by Levinthal and typical of chemical reactions, Wolynes proposed a statistical view of the protein folding landscape borrowed from the statistical mechanics of random heteropolymers. Unlike the rugged landscapes of these

polymers made up of a random assortment of amino acids, natural proteins have a smoother landscape bearing resemblance to a funnel with chain entropy being the major deterrent to folding to the native ensemble of structures.

In the funnel view of energy landscape theory, the vertical axis represents the stabilization energy achieved by partially folded microstates. Much like beads on a string, there are a number of ways to arrange the chain so that each bead is oriented in its final position. With respect to proteins, successive levels of stabilization free energy are sampled to achieve the native state. Each level represents an ensemble of structures with the same degree of stabilizing free energy, or a balance in the loss of chain entropy with the gain in bond energy. Thus, a protein's course down the funnel is combinatorially dictated by the number of native contacts formed (those that are present in the native state). Folding is slowed by the formation of non-native interactions which can be thought of as traps in the folding process. These traps invariably slow folding, but the magnitude is dependent on the energy associated with breaking these associations and proceeding downhill on the energy landscape. The slope, or gradient, of the funnel in addition to the energy landscape "roughness" determines the speed limit for protein folding. Each protein funnel is unique, with a steep slope increasing the rate of reaching the free energy minimum while a broad slope would impose more lengthy folding kinetics. This is a great example of the statistical nature of protein folding and is a partial resolution to Levinthal's random search.

The ‘roughness’ of the energy landscape is influenced by the degree of nativeness as the protein proceeds down the funnel-like landscape. A native interaction can be thought of as a smoothing of the folding trajectory thus favoring more stability in the free energy hierarchy. As mentioned above, non-native interactions result in a more rugged landscape with traps and valleys separating the downhill trajectory toward the native ensemble by energetic barriers which slow folding. The frictional forces exerted upon the protein chain in breaking non-native contacts in these transient traps determines the relative temperature dependent diffusion constant of the folding reaction coordinate. Of the many potential folding routes, those that pay low entropic costs while gaining stabilization energy are favored, highlighting the minimal frustration principal. The rate limiting step of folding is synonymous with the transition state ensemble, which is a common potential among pathways in the free energy landscape and determines the folding rate. Experimentally, these features of the folding funnel aid in calculating folding kinetics as a function of stabilization energy and minimizing degrees of freedom of the peptide chain. Functionally, free energy landscape theory suggests that protein topology is the ultimate factor in governing a protein’s path to folding. Support for this conclusion is gained by the observation that most ultrafast folding proteins are composed primarily of α-helical structure which relies specifically on local interactions as opposed to long range contacts (Naganathan, A. N. 2007).

1.2 Protein Folding Trajectories

Following the work of Anfinson, many small single domain proteins (~110 amino acids in length) have been the subject of unfolding and refolding studies in which the folding rates were found to occur in less than one second (Jackson, S. E. 1998, Burton, R. E. 1996, Crane, J. C. 2000). The earliest protein folding studies focused on model systems exhibiting a three-state mechanism which proceed to fold by populating stable intermediate species (Kim, P. S. 1982, Kim, P. S. 1990). The assumption was that protein folding proceeded much like metabolic pathways, namely by distinct intermediate structures on a well defined pathway. By trapping these intermediates, their structural features could be characterized. Typical model proteins which could be manipulated to trap these populated intermediates as a function of proline *cis / trans* isomerism, disulfide bond shuffling, or heme dislocation include Ribonuclease A (Garel J. R. 1973), bovine pancreatic trypsin inhibitor (Weissman, J. S. 1991, Weissman, J. S. 1992a, Weissman, J. S. 1992b, Weissman, J. S. 1993, Weissman, J. S. 1995) and cytochrome c (Englander, S. W. 1998) respectively. Paradoxically, these folding intermediates exhibiting biphasic kinetics are a product of misfolding or kinetic traps as the least resistive pathway to folding is monophasic. In other words, intermediates are a product of ignoring the actual rate limiting step as a protein gains secondary and tertiary structure. The view that this step-wise process solved Levinthal's paradox lost in popularity upon Jackson and Fersht's work on barley chimotrypsin inhibitor 2 (CI2) which demonstrated the

two-state behavior of this protein with agreement between energetic parameters obtained through equilibrium and kinetics (Jackson S. E. 1991a).

1.2.1 Two State Approximation

After validating the incidence of apparent two-state behavior in CI2, the focus in protein folding shifted away from schemes to trap intermediates to characterizing the high energy transition state. Interpreting two state behavior using chemical models, the unfolded state is separated from the native state by a sufficiently high (> 3 RT) energy barrier. At the saddle point of this free energy barrier sits the transition state, more aptly referred to as the transition state ensemble, which acts as the common rate limiting species. Quantitatively, this model represents overcoming the entropic cost of fixing the peptide chain with the gain in stabilization energy (Figure 1.1). A simple equilibrium expression can be written:

$$F \underset{k_u}{\overset{k_f}{\rightleftharpoons}} U \tag{1.1}$$

where F and U represent the native and unfolded states kinetically controlled by rate constants k_f and k_u, respectively. This simple model is in clear conflict with energy landscape theory.

Statistically speaking, a two-state model can be represented by the probability distribution of macrostates. Since a two-state system is an “all or nothing” transition,

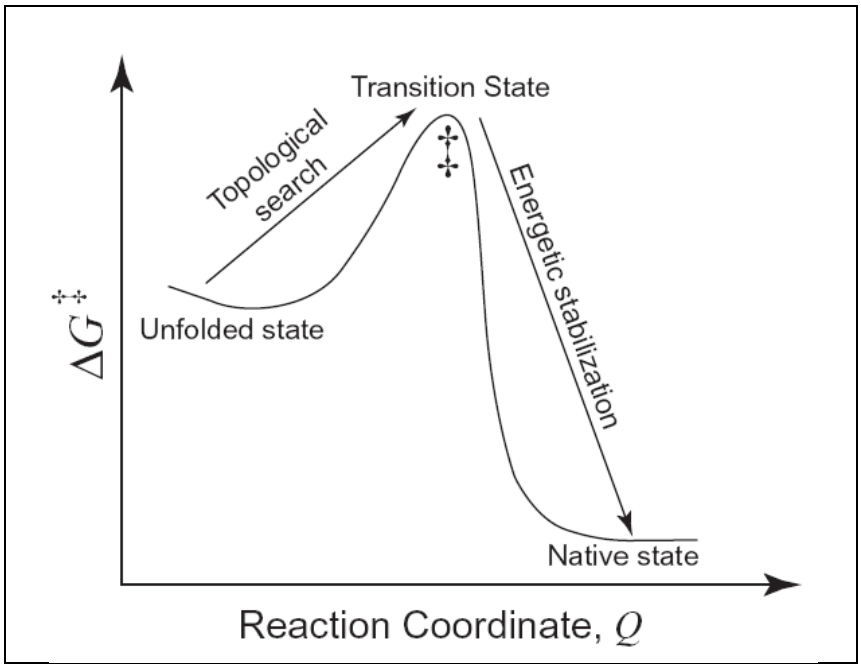

Figure 1.1 A simple two-dimensional reaction coordinate with free energy as the order parameter illustrating the energetic barrier separating the two ground states. The metastable transition state is at the free energy maxima where additional native interactions push the folding in the downhill direction leading to the lowest energy native state.

only two species are populated. In natively biased conditions, the balance of the population density would entirely favor the native state and *vice versa* for unfolding conditions (variables including: pH, temperature, pressure, presence of chemical denaturants). At the chemical equilibrium, or denaturation midpoint (T_m, C_m) where thermodynamic terms are zero, fifty percent of the population would occupy both the native and unfolded states.

In recent years, folding studies have been fit with prejudice to a simple chemical two-state model by enforcing rate equilibrium free energy relationships (REFER) through mutational analysis (Fersht, A. R. 1992) and ignoring the physical significance of experimental baselines. In equilibrium, denaturant dependent (temperature or chemical) unfolding transitions monitored by spectroscopic probes that are sigmoidal in shape are taken as evidence for a simple two-state transition and thought to express 'cooperativity' in the sense of an all-or-nothing transition. Kinetic relaxation experiments that can be fit to a single exponential are thought to represent crossing a free energy barrier, and thus would indicate two-state folding. However, this equality has been recently questioned by evidence that single exponential kinetics don't necessarily guarantee the presence of an energetic barrier (Muñoz, V. 2002, Sadqi, M. 2003).

A wealth of experimental evidence has allowed the creation of a "laundry list" of criteria to classify a true two-state system. As mentioned previously, spectroscopic techniques reveal sigmoidal unfolding as a function of denaturant with multiple

probes displaying the same denaturant dependent behavior. Differential scanning calorimetry experiments of two state proteins are represented by a single sharp transition with the van't Hoff enthalpy equaling the calorimetric enthalpy. Likewise, kinetic experiments exhibiting single exponential relaxation rates and observable unfolding / refolding rates with a linear dependence on denaturant concentration are characteristic of two-state proteins. Finally, and perhaps most importantly, experimental parameters (ΔH, ΔS, ΔCp, denaturant *m*-values) should agree when compared between equilibrium and kinetics.

A great number of proteins have been classified with a blanket two-state model. Experimental evidence suggests that true examples of two-state folding are rare and that rather than a rigid sequential chemical two-state description, folding mechanisms are dynamic and an ensemble view of the energetic sampling is a much more accurate view. This is buttressed by the experimental detection of downhill folding (Garcia-Mira, M. M. 2002, Naganathan, A. N. 2005a, Sadqi, M. 2006) proteins and those that fold with marginal barriers (Fung, A. 2008, Munoz, V. 2007, Naganathan, A. N. 2007).

1.2.2 Downhill Protein Folding

An expected scenario proposed by energy landscape theory in contrast to the two-state approximation is downhill, or one-state, protein folding (Bryngelson, J. D.

1995). Folding in the downhill regime proceeds without the opposition of a free energy barrier separating the ground states. A free energy gradient is still present, but the absence of a barrier precludes the existence of a transition state. Therefore, movement along a simple reaction coordinate is an environmentally sensitive product of the continuous minimization of free energy with a nominal cost in configurational entropy. In other words, a shift from native to destabilizing conditions results in a unimodal distribution of structurally distinct microstates favoring structural organization in harmony with environmental conditions (Figure 1.2).

Like a two-state scenario, the folding rate can be dependent on a diffusion controlled kinetic bottleneck. Unlike in the two-state scenario in which the free energy barrier dictates the rate limiting step, the effective diffusion coefficient (D_{eff}), or kinetic prefactor (k_o), is the sole director of the flux along the reaction coordinate, and thus the roughness of the free energy landscape. This process is synonymous with the downhill phase of the two-state activated complex in funnel-like energy landscapes. If the diffusion constant permits rapid movement along the reaction coordinate, the folding rate is a single averaging of this downhill progression. However, if D_{eff} becomes sufficiently small, intrachain dynamics slow folding and the induced roughness of the energy landscape gives rise to a slow phase leading to biphasic kinetics.

The first experimental evidence of downhill protein folding was reported in 2002 by Muñoz and co-workers on the peripheral subunit binding domain (PSBD)

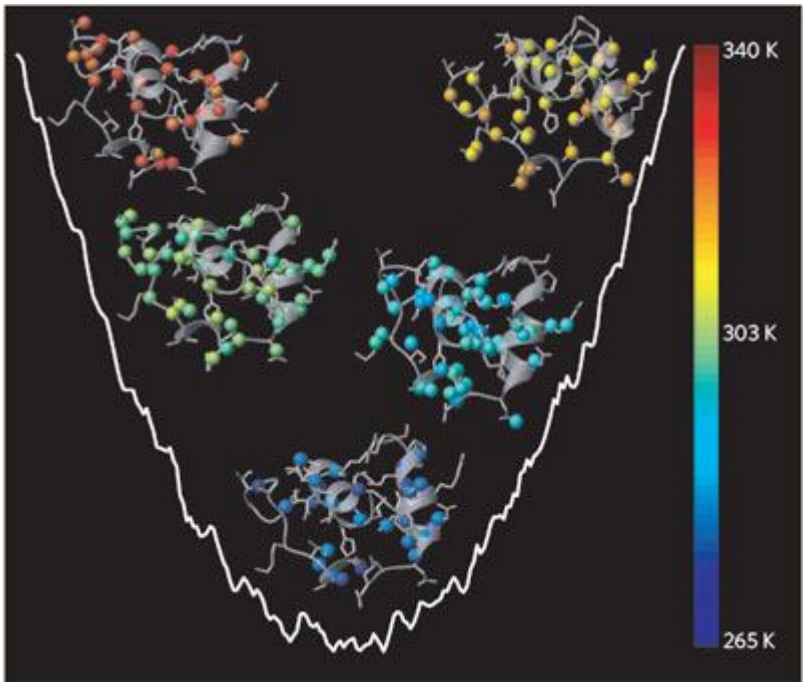

Figure 1.2 Two-dimensional slice of a representative free energy landscape for the downhill folding protein BBL as a function of temperature (indicated by color). Free energy (y-axis) is indicated by the degree of structural formation along the reaction coordinate (x-axis) with high temperatures corresponding to high-energy structures. Formation of structure is continuous with negligible free energy barriers arising from the diffusion along the reaction coordinate (i.e. roughness). (Figure from Kelly, J. W. 2006)

protein BBL from *E. coli* (Garcia-Mira, M. M. 2002). The folding behavior of this predominantly helical, independently folding, 40 residue protein was investigated with a systematic equilibrium characterization. Experiments included following the global conformational transition using differential scanning calorimetry (DSC), tertiary environment using fluorescence and Förster resonance energy transfer (FRET), and far-UV circular dichroism (f-CD) to monitor fluctuations in secondary structure.

The initial findings by Muñoz indicated a broad range (295 – 335 K) of melting temperatures (T_ms) dependent on experimental probe. Further, secondary structure composition was found to be wavelength dependent in which different features are monitored at distinctive wavelengths. This lack of agreement is in stark contrast to the expected outcome with a two-state system. Further, probability density maps constructed from a statistical mechanical model clearly illustrated free energy profiles with an absence of a free energy barrier, satisfying the fundamental downhill folding criterion.

Oliva and Muñoz generated an additional tool in diagnosing downhill folders using f-CD to monitor the behavior of local interactions influenced by temperature and urea (Oliva, F. Y. 2004). These double perturbation experiments probing the α-helix content in BBL reproduced theoretical simulations of the phenomenological model that showed downhill folders to display crossing baselines and variable temperatures of maximal signal (T_{max}) induced by cold denaturation at increasing urea

concentrations. This method also highlighted the need to perform a global two-state fit of the temperature / urea data in addition to single two-state fits as opposed to only performing single fits. This method was expanded by comparing thermodynamic parameters extracted from the individual fits compared to the global two-state fit (Nananathan, A. N. 2005a). ΔH and ΔS obtained from single fits for BBL displayed a distinct denaturant dependent curvature indicative of downhill folding while the global fit showed a linear dependence on denaturant concentration signifying a two-state folder. The combination of these two exhaustive approaches affirmed the experimental characterization of BBL and several length and chromophore dependent variants as downhill folders after being challenged in the literature (Ferguson, N. 2004, Ferguson, N. 2005). The development of the Variable Barrier model to analyze barrier heights in calorimetric profiles lead to quantitatively assigning the absence of a free energy barrier (Munoz, V. 2004) and the combination of these methods have lead to the characterization of another protein and the subject of the current work, gpW, which folds essentially downhill with a marginal barrier estimated at 0.6 kJ/mol (Fung, A. 2008). More recent experiments were performed at the atomic level using nuclear magnetic resonance (NMR) to follow the unfolding of 158 backbone and sidechain protons in BBL. Temperature dependent unfolding was monitored by proton chemical shifts. As in previous experiments, a broad range (~60 K) of thermal unfolding was observed with many protons exhibiting sigmoidal transitions with others displaying two identifiable transitions. This is evidence for a complex unfolding behavior consistent with a continuous process with the free energy

landscape possessing a single energetic minimum, which shifts upon thermal perturbation. The average unfolding profile obtained by NMR also overlaid nicely with the f-CD unfolding transition. This indicates the global unfolding of the average ensemble observed by NMR is coincident with the comparatively low resolution technique monitoring local interactions.

1.3 Mutational Approaches in Protein Folding

The ability to manipulate the primary sequence of proteins has become much simpler to accomplish with the advent of site-directed mutagenesis. Protein engineering has lead to tremendous success in elucidating the functional roles of amino acids in proteins at the level of individual residues. Mutational studies have been applied to determine functional residues involved in enzymatic reactions, protein-protein interactions, protein-DNA interactions, and protein folding among others. With such a wealth of structural information becoming more readily available, protein engineering using rational design has developed into a common process utilized in protein recognition studies. Application of these investigations has largely been applied towards drug development. Advances have even become sophisticated enough to develop de novo protein structures using computational algorithms which possess novel folds and a single native structure (Kuhlman, B. 2003).

The native state of a folded protein is stabilized by mostly non-covalent interactions. These interactions include hydrogen bonding, hydrophobic forces, electrostatics, van der Waals interactions, and dipole-dipole interactions. The native state is additionally stabilized by expelling solvent molecules from the interior of the protein hydrophobic core. By examining the functional residues involved in establishing and maintaining these interactions within a protein's three-dimensional structure, the forces directing the process of folding may be partially resolved. Taken one step further, mutational analyses provide a platform for investigating the thermodynamic and kinetic contributions of individual residues in the process of protein folding. Three methods have served as a launching point towards understanding the forces governing protein folding. Alanine scanning has proven particularly suited in determining functional "hot spots." Development of the phi (ϕ) value analysis permits quantifying the energetic destabilization of the transition state as a function of a particular point mutation referenced to the native state of the wild type protein. An extension of the ϕ-value analysis, the psi (ψ) value analysis, was developed to utilize metal coordination of an engineered bi-histidine site to explore folding pathways of the transition state ensemble. Each of these methods is deserving of merit by their ability to monitor structure, function, and energetics, though individually, their relatively narrow application is a function of each method's limitations.

1.3.1 Alanine Scanning

Proteins, the workhorses of the cell, are largely responsible for performing a multitude of duties. Individual proteins generally have a single role where functional diversity is minimal, thus necessitating the existence of equal numbers of proteins to unique functions within the cell. As each protein has a specific role to play, it becomes a central enquiry to discover which residues are critical in executing these functions.

Alanine scanning mutagenesis is one such method to achieve a residue by residue account of amino acid functionality. One such area of concentration where this method is readily applied is protein-ligand interactions. Non-covalent interactions at these interfaces are a product of spatial organization between the receptor and ligand. Often, several functional residues are clustered together in three dimensional space creating a sort of epitope. High resolution biophysical techniques such as X-ray crystallography and multidimensional NMR reveal these contact points, but fail to expose the energetic contributions of each residue.

Rapid characterization of sidechain function can be achieved by systematic alanine substitution. By substituting a wild type amino acid with alanine, the sidechain is truncated to the β-carbon thus deleting all possible hydrophobic or electrostatic interactions contributed by sidechain functional groups. This is however considered a conservative mutation since no conformational entropy is introduced into the peptide backbone upon substitution. This strategy proved useful in

determining the sidechains involved in human growth hormone receptor-ligand interactions (Cunningham, B. C.), structure-function studies of human tissue-type plasminogen activator (Bennett, W. F.), and lethal phenotypes as a function of surface charge distribution in the yeast ACT1 gene (Wertman, K. F. 1992).

Traditional alanine scanning mutagenesis involved *in vitro* activity assays subsequent to expression and purification of each mutant. This labor intensive process has evolved with the advent of combinatorial protein libraries in which mutants are readily generated according to the ability to alter several amino acids by changing just one nucleotide of the DNA codon. Phenotypes can then be rapidly screened by phage display (Smith, G. P. 1997, Sidhu, S. S. 2000, Weiss, G. A. 2000).

In addition to protein function, sidechain functional groups are a critical component in the free energy minimization of a protein upon folding. Bond formation and hydrophobic packing stabilizes local and long range interactions as a protein folds. The increase in conformational enthalpy as a function of structural order along with the concomitant decrease in backbone conformational entropy is a manifestation of a protein achieving its native state; the macrostate with the lowest free energy.

Through the course of evolution, amino acid conservation has been a product of Natural Selection. This is evident in multiple sequence alignments across families of functionally related proteins. Properties inherent to amino acid composition such as size and charge contribute in an additive fashion to minimizing the free energy of

the native state. If such properties provide stability as a function of position, mutations at these positions reveal their essential roles in protein stability for the wild type residue. This was apparent in experiments using *lac* repressor as a canvas for single amino acid mutations (Miller, J. H. 1979). Nearly 1500 alterations at 142 positions were created, and the results showed that over half had no effect on protein structure or stability. However, amino acids essential to structure or function significantly destabilized or even eliminated vital contacts contributing to the native stability.

A study on T4 lysozyme indicated that surface residues were most amendable to alanine substitution. Thermostability was increased upon mutation of solvent exposed residues while buried or partially buried amino acids decreased the melting temperature. Another interesting finding from this mutational study was that a salt bridge between two neighboring residues was not a factor in overall protein stability (Heinz, D. W. 1992). A similar alanine mutational study of the Arc repressor revealed that structure and function was equal to that of wild type construct with stability matching or even exceeding that of the wild type. Additionally, folding rates of some mutants were up to 30 times faster and unfolding rates up to 50 times slower than wild type (Bowie, J. U. 1989). These studies imply that some residues may form unfavorable interactions and imply that all non-covalent interactions are not created equal.

These results suggest that proteins are largely amendable to changes at non-critical positions, but key residues comprising a subset of the overall primary sequence are responsible for a greater part of protein structure, stability and function.

1.3.2 Phi Value Analysis

In the late 1980's, protein engineering studies on tyrosyl-tRNA synthetase revealed the contributions to the binding energy of different sidechains to enzyme catalysis (Fersht, A.R. 1988). At the same time but independently, mutations on bovine pancreatic trypsin inhibitor by Goldenberg *et al* helped distinguish a difference between the destabilization free energy of the native state relative to the unfolded state and the stabilization energy of the intermediates (Goldenberg, D. P. 1989). Similarly, Matouschek *et al* found that the hydrophobic core of barnase was compromised upon mutation and that the N-terminus was weakly structured (Matouschek, A. 1989).

Conceived from the classical Brønstead and Leffler analyses in physical-organic chemistry, the phi-value analysis was conceptualized to relate structure of the transition state or intermediates to a reference state in protein folding. Formalized by Fersht to interpret energetic perturbations to protein folding, the ϕ-value analysis sought to bridge the structural gap between native and denatured states by characterizing the transition state using kinetics and structure-activity relationships.

Briefly, by making conservative mutations mostly at solvent exposed amino acids within a secondary structural element, the equilibrium destabilization free energy relative to the native state could be quantified. As kinetics are the only way of analyzing the metastable transition state, the unfolding and folding rates would then be measured.

A mutation may perturb the free energy of the native state relative to the denatured state, quantified $\Delta\Delta G_{N\text{-}D}$. That same mutation may destabilize the free energy of the transition state relative to the denatured state $\Delta\Delta G_{\ddagger\text{-}D}$. The kinetic destabilization free energy relative to the equilibrium destabilization free energy produces a ϕ-value between 0 and 1. A formal expression is therefore represented by $\phi_F = \Delta\Delta G_{\ddagger\text{-}D}/\Delta\Delta G_{N\text{-}D}$.

It is therefore assumed that a value of 0 indicates that the same degree of structural formation is present between the transition state and unfolded state. That is to say that upon mutation, a particular site is destabilized in the transition state the same amount as the denatured state. A ϕ-value of 1 signifies that the transition state has the same structure as the native state. Analogous to the value of zero, the stabilization free energy of the transition state is perturbed equally to that of the native state when a value of one is achieved.

Experimentally, ϕ-value analyses on the protein barnase indicated that many residues reveal ϕ-values near one and were clustered together (Serrano, L. 1992b, Serrano, L. 1992a). This was interpreted to mean that the transition state possessed a

high degree of nativeness both in terms of structure formation and environment and that a small subset of residues was responsible in forming a nucleation site for the protein to collapse upon. Further experiments on core residues of the two state protein chymotrypsin inhibitor 2 revealed fractional and ϕ-values near zero (Jackson, S. E. 1993a, Jackson, S. E. 1993b). The authors concluded that none of these residues significantly impacted protein folding and the percentage of structural organization was proportional to the fractional ϕ-values. Since, ϕ-value analysis has been applied to an array of proteins to investigate the role of individual amino acids on protein folding (Goldenberg, D. P. 1999). From a structural standpoint, much work has been done to determine if a general trend in transition state formation can be collated into diffuse or polarized (Itzhaki, L. S. 1995, Grantcharova, V. P. 1998, Gruebele, M. 1998, Riddle, D. S. 1999). This has lead to much debate about the suitability of drawing structural conclusions from ϕ-values (Sosnick, T. R. 2004, Feng, H. 2004, Sanchez, I. E. 2003, Bulaj, G. 2001, Ozkan, S. B. 2001, Fersht, A. R. 2004, Krantz, B.A. 2004, Raleigh, D. P. 2005).

The ϕ-value analysis makes several assumptions that are key in assigning the degree of structure formation to the value obtained from the free energy ratio: (a) mutations are conservative, (b) mutations do not perturb the structure of the transition state, (c) the structure of the folded and unfolded states are not perturbed upon mutation, (d) folded regions of the transition state have native-like structure, (e) the folding pathway is not altered (ie the transition state of the mutant and wild type is the same), (f) new interactions are not formed as a result of mutation, (g) the free energy

landscape is smooth with clear free energy maxima and minima and proceeds over a well defined trajectory, and (h) slope of equilibrium and kinetic denaturant dependent folding and unfolding free energy (ie m_f and m_u values) is not significantly altered.

However, implicit in these being assumptions, little assurance is offered to validate drawing structural conclusions from energetic changes. Additionally, an empirical catalog of ϕ-values compiled from the vast array of proteins subjected to this mutational analysis reveals that ϕ-values are mostly distributed around 0.3. Fractional values may arise as a result of partial structure formation of a single transition state. Alternatively, if the transition state is an ensemble of multiple transition state conformations, perturbing an interaction important in one of these structures will reduce flux through that trajectory giving an artificially low ϕ-value. An independent calculation by Raleigh and Plaxco on seven well characterized two-state proteins concludes that only 1.6% of all ϕ-values for these proteins were greater than 0.8 using the criterion that $\Delta\Delta G_{eq} > 7$ kJ/mol as proposed by Sanchez and Kiefhaber (Raleigh, D. P. 2005, Sanchez, I. E. 2003). The presence of negative ϕ-values or ϕ-values greater than one appears common (Ozkan, S. B. 2001), and are translated to mean that the residue being probed has a kinetic role or non-native interactions are formed. This conclusion disobeys one of the assumptions of the method, and the amino acid in question is routinely ignored. In fact, these values, along with a great number of "reasonable" ϕ-values, suggest the free energy change upon mutation is too minor to register an accurate determination. This underlines the

error associated with a two point mutational analysis and highlights the fact that beyond quantifying energetic differences between the transition state and a reference state, inferring structure formation from ϕ-values is critically flawed.

1.3.3 Psi Value Analysis

The difficulty in inferring transition state structure from energetic differences quantified by ϕ-values ignited the development of an analogous method, namely the ψ-value analysis. Utilizing the metal binding capacity of histidine, bi-histidine sites were engineered into known positions throughout GCN4 and ubiquitin with the aim of being able to continuously tune the stability of a bi-histidine site with increasing concentrations of metal ions (Krantz, B. A. 2001, Krantz, B. A. 2004). Sites were engineered in close proximity making use of the i, i+4 orientation in alpha helices and the calculated placement in three dimensional space across beta strands. Structural formation is directly probed as a function of metal binding due to the known placement and proximity of the two histidine residues.

In the native state, the bi-histidine site is positioned such that the metal ions are free to bind, only dependent on the affinity, or native dissociation constant (K_N) of the reaction. Conversely, in the unfolded state, the two histidine residues are not in an appropriate orientation conducive to metal binding. However, there is potential for diffusive arrangement of the bi-histidine site in the unfolded ensemble, and is

accounted for with an equilibrium constant, γ, and a separate dissociation constant, K_U. A linked equilibrium relates the $\Delta\Delta G_{eq}$ with increasing stability upon metal binding. Reaction kinetics are similarly linked dependent on the metal binding site being present or absent in the transition state. The flux, ρ, through a particular pathway as a function of binding site formation (transition state structure) measures the population change via reaction rates, $k_{present}$ and k_{absent}. A Leffler plot of the kinetic data illustrates a continuous shift in stability as a function of metal ion binding. The redistribution of free energy as more stability is achieved with increasing metal concentrations can give rise to linearity or curvature in the Leffler plot depending on the binding efficiency. From the instantaneous slope as a function of binding stability, the ψ-value can be defined $\psi = \partial\Delta\Delta G^{\ddagger}_{f} / \partial\Delta\Delta G_{eq}$. A rigorous discussion on the calculation of ψ-values can be found by Sosnick in Chemical Reviews (Sosnick, T. R. 2006).

Two possibilities exist at the extremes where the slope is linear. When the slope is zero across the range of increasing metal ion concentrations, metal binding sites in the transition state ensemble have the same population as the unfolded state and $\psi = 0$. If the slope is one, the bi-histidine site in the transition state ensemble is assembled to the same degree as the native state, thus $\psi = 1$. Unlike in the ϕ-value analysis, fractional ψ-values represent the fraction of the transition state ensemble that have the bi-histidine site formed, and thus are a direct measure of structure

formation. These ψ-values are extracted by the derivative of the Leffler plot slope and are dependent on metal ion concentration.

ψ-value analysis on a N-terminal bi-histidine mutation of a cross linked version of GCN4 with a homogeneous transition state concluded that metal binding did not supersede the influence of disulfide mediated N- or C-terminal connectivity of the coiled coil (Sosnick, T. R. 2001). Another aspect of ψ-value analysis, determining pathway heterogeneity, was explored with a dimeric version of GCN4. By increasing metal ion concentration, the bi-histidine nucleation site was populated by the transition state ensemble with greater frequency demonstrated by a shift towards unity of the ψ-value. Expanding the utility of this analysis, the α/β protein ubiquitin was subjected to experimental characterization by ψ-value analysis. A composite transition state structure was assembled by engineering bi-his sites both to helical as well as β-strand sites. Unsurprisingly, a very different view of the transition state structure was observed by interpreting ψ-values compared to that obtained from ϕ-values.

Sosnick *et. al.* assert that ψ-value analysis gains significant ground on more traditional mutational analyses with respect to protein folding. ψ-values report on known binding partners and a single interaction is probed as a reflection of structural topology. Recently, Bodenreider and Kiefhaber showed by simulation that curvature in Leffler plots was an effect of parallel folding/unfolding pathways and that the ligand binding equilibrium was dependent on transition state decay (Bodenreider, C.

2005). In addition, it is intuitive to question the effect of the binding constants on folding. The ψ-value analysis assumes that metal binding is only dependent on bi-his site formation. In one scenario, the dissociation affinity could be sufficiently high that equilibrium between bound and unbound is significantly favored by the former. This would unilaterally indicate a transition state ensemble that was considerably native-like. It is scenarios like these which are cause for caution when considering ψ-values between zero and unity.

1.4 Metal Binding

Regulated by sequence, size, topology, and energetics, proteins have been found to fold over a diverse array of times and by distinct mechanisms. In addition to the behavior of model single domain proteins, an important consideration should be directed at cofactor mediated folding reactions. Proteins such as cytochrome b_{562} and myoglobin have been found to retain their non-covalent metal binding interactions even after unfolding (Robinson, C. R. 1997, Bertini, I. 1997). The implication here is that metal binding domains may at least be partially responsible for protein folding. The consequence of metal binding can also be described in the ubiquitous zinc-finger superfamily of proteins which constitute a large number of transcription factors, DNA binding proteins, and proteins involved in regulation. Here, a zinc ion coordinates the folding of a consensus region composed of a 2-stranded β-sheet and an α-helix through non-covalent interactions with a combination

of cysteine and histidine residues. In the absence of the zinc ion, folding of this domain is deficient significantly impacting the activity of these DNA binding domains.

Protein biochemists have long exploited the metal binding properties of these particular residues (Cys and His) or sequence homologs of known metal binding domains to construct a folding nucleation site and induce structure formation and stability. As examples: the heat capacity change due to helix formation at various temperatures has been measured by ITC of an EF-hand calcium binding motif (Lopez, M. M. 2002), the free energy contributions of helix initiation and the influence of C-capping residues (Goch, G. 2003), residue specific enthalpy of the helix-coil transition (Richardson, J. M. 2005), increases in helical conformations of up to 90% in short peptides as a result of bi-dentate i, i+4 coordination of cadmium, zinc, or copper (Ghadiri, M. R. 1990), and most recently by Krantz and Sosnick to induce helix nucleation and explore pathway heterogeneity (ψ-analysis Section 1.3.3) (Krantz, B. A. 2001, Sosnick, T. R. 2006).

It is obvious that metalloproteins have provided many details of simple features in the protein folding process. What is clear is that the behavior of these metal binding proteins and structural homologs can act as interesting scaffolds to investigate more complex behaviors in the protein folding problem. Indeed, the metal-binding properties of histidine will be taken advantage of to investigate the functional corollary of downhill protein folding.

1.5 Research Aims and Chapter Summary

The delicate progression of a newly translated polypeptide chain from a random conformation of amino acids to its well-defined three-dimensional structure represents the terminal step in the Central Dogma of biology. Early predictions suggested a timescale of achieving the native state to be of infinite length. Experimental data provided evidence that proteins intrinsically possess all the information necessary to direct protein folding as a product of their amino acid sequence. Free energy landscape theory highlighted the ensemble view of folding and revolutionized the way of thinking about protein folding processes from classical chemical models to a statistical approach. This theory also predicted the existence of barrierless folding in contrast to the conclusion that fast-folding single domain proteins fold in a two-state manner. Downhill folding was experimentally confirmed through exhaustive characterization of the PSBD protein BBL. The use of several different techniques to monitor and compare the behavior of probes reporting on local and global environments confirmed the existence of the energetically and structurally continuous process. The efforts by Muñoz in developing experimental approaches and physical models have been unparalleled in their ability to differentiate between the most common mechanisms of folding for small proteins, most notably downhill folding. In fact, independent analysis of many of these single-domain proteins suggest downhill folding is more common than originally thought and may belie the assumption that native structures represent the only conformation with functional significance. Early folding studies took advantage of protein engineering to explore a

residue-by-residue map of functionally significant amino acids. These mutational approaches have been a significant source in unraveling the energetic forces responsible for protein stability and have more recently been steered with a 'directed evolution' approach to identify structural assembly in the folding process. In addition to reassigning previously classified two-state folders, finding additional proteins exhibiting downhill behavior becomes a critical quest in defining a functional relevance of proteins folding in the absence of an energetic barrier. The work presented forward is an effort in adding a new member to growing list of proteins folding in the downhill regime and attempting to add a functional perspective to barrierless folding via mutational modifications which take advantage of the observed continuous folding / unfolding.

Chapter 2 serves as a catalog of the materials and procedures used to prepare and perform the exhaustive characterization of the model protein, gpW. It also functions as an introduction to the experimental techniques used to characterize protein folding / unfolding transitions. Each method discussed serves as a unique way to monitor secondary or tertiary structural variations in conjunction with protein folding / unfolding both by equilibrium and kinetic methods. In chapter 3, an introduction to gpW will be provided along with a description of the bi-histidine mutants. This will lead into a rigorous characterization of the equilibrium behavior of gpW and associated bi-histidine mutants by the established multiprobe analysis. Chapter 4 provides an in-depth look at differential scanning calorimetry and its use in directly extracting free energy barrier heights using the variable-barrier model.

Results from applying this model to DSC data from WT gpW are presented. A description of the accompanying kinetic behavior of gpW and bi-histidine mutants is discussed in chapter 5. Chapter 6 will provide a more in-depth discussion of bacteriophage λ assembly and the functional role of gpW. Discussion of the biological implications associated with the experimental evidence presented in chapters 3 – 5 will also be included along with tentative results from metal binding experiments attempting to tune local structure with consequences reflected in global stability.

2. Materials and Methods:

2.1 Cloning and Protein Expression

The gpW gene was supplied by Alan Davidson in the pET-21d vector (Novagen) with 5' NcoI and 3' SalI restriction sites (Maxwell, K. L. 2001). The original construct contained a 16 residue C-terminal tail, which included a hexa-histidine tag and the sequence encoding a FLAG epitope. Due to the unstructured nature of the six C-terminal residues, in addition to the his-tag and FLAG epitope encoding residues, an oligonucleotide primer (MWG Biotech) was designed to insert a stop codon following the sixty second residue to truncate the sequence from sixty eight residues, thus eliminating the unstructured, yet hydrophobic C-terminal residues which contributed to low solubility in the original construct. In addition, the same mutagenic oligonucleotide contained the sequence to replace the SalI restriction site with the HindIII restriction site in order to subclone the gene into the pBAT vector. The gene was amplified with the mutagenic oligo using polymerase chain reaction (PCR) using 200 mM dNTPs, 400 nM oligos, pfu Turbo DNA polymerase (Stratagene), and a template DNA concentration of 150 ng/µl. Amplification was achieved with initial heating at 95^{o}C for 1 minute followed by 25 cycles of a 30 second denaturation step at 95^{o}C, a 30 second annealing step at 58^{o}C, and a 1 minute extension step at 72^{o}C. These cycles were followed by a 5 minute extension at 72^{o}C before cooling to 4^{o}C. The PCR product was purified using the QIAquick PCR

purification kit (Qiagen) and subsequently doubly digested with NcoI and HindIII restriction enzymes (Invitrogen). The digestion product was again purified with the QIAquick kit and ligated into the pBAT vector using T4 DNA ligase (Stratagene). Success of the modified construct was judged by DNA sequencing (University of Maryland, UMBI DNA sequencing facility).

Protein expression was accomplished by transforming the ligation product into BL21 DE3-Gold ultra competent cells (Stratagene). Positive transformants were selected by isolating the plasmid DNA using a QIAprep spin miniprep kit (Qiagen) from single colonies expressing antibiotic resistance to 100 mg/mL ampicillin followed by DNA sequencing (UMD CBR DNA sequencing facility). Successful clones were screened for protein overexpression levels by inoculating 5 mL Luria Broth (Fisher) cultures and inducing with isopropyl β-D-1-thiogalactopyranoside (IPTG) (Invitrogen) to a final concentration of 1mM when the OD_{600} of the culture reached 0.8. Time points were collected at 0, 1, 2, 3, 6, and 12-16 hours post induction. Colonies producing the greatest quantity of gpW were judged from expression levels visualized via 10-20% Tricine gels (Invitrogen) sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Gels were run at 80 mA and 125 volts. The colony generating the highest yield was selected, inoculated into an overnight culture of 5 mL LB broth including 100 mg/mL ampicillin, and induced for expression with 1 mM IPTG at 4 L scale for 3 hours once the OD_{600} reached 0.8. Cells were harvested by centrifugation at 3000 rpm for 45 minutes in a Beckman benchtop centrifuge. The pellet was resuspended in a lysis buffer containing 20 mM

sodium phosphate 200 mM sodium chloride pH 6.0. Resuspended cells were lysed on ice via sonication. The suspension was then centrifuged at 25,000 rpm for 35 minutes at 4^{o}C in a Beckman ultracentrifuge using a Ti-45 rotor. The supernatant was filtered using Millex GV 0.22 μm PVDF syringe filters (Millipore) in preparation for purification by chromotography.

gpW bi-histidine (bi-his) mutants were selected and designed based on secondary (α-helical mutants E5H E6G A9H, R11H A12G, K45H A48G E49H) or tertiary structural interactions (β-strand mutants E34H V23H, E35H Q27H) (Mutants will henceforth be referred to as E5H, R11H, K45H, V23H, and Q27H). Mutants were generated using a similar PCR mutagenic strategy as with the wild type truncation. The wild type gene was amplified with the respective mutagenic oligo using PCR. Each mutant was optimized to obtain the maximum PCR product. For E5H and R11H, 10 ng of template DNA was combined with 125 ng of the respective oligo, and 2.5 units of pfu Turbo DNA polymerase with a final concentration of 400 μM dNTPs in the reaction mixture of 50 μL. For K45H, 5 ng of template DNA was combined with 125 ng of the respective oligo, and 2.5 units of pfu Turbo DNA polymerase with a final concentration of 400 μM dNTPs in the reaction mixture of 50 μL. For the β-strand double mutants, sequential cycles of PCR mutagenesis were necessary since the mutation sites are located far from each other in sequence space. E34H was generated using 60 ng of template DNA combined with 125 ng of the respective oligo, 2.5 units of pfu Turbo DNA polymerase and a final concentration of

400 μM dNTPs in the reaction mixture of 50 μL. V23H and Q27H required 10 ng of template DNA was combined with 125 ng of the respective oligo, 2.5 units of pfu Turbo DNA polymerase and a final concentration of 400 μM dNTPs in the reaction mixture of 50 μL. Each double or triple mutant required a separate PCR run due to the varying thermal melting temperatures of the respective oligos. The thermal cycle for each mutant included an initial denaturation step at 95°C for 30 seconds followed by 24 cycles of a 30 second denaturation step at 95°C, a 1 minute annealing step at X°C, and a 10 minute extension step at 72°C. These cycles were followed by a final 10 minute extension at 72°C before cooling to 4°C. E5H, R11H, K45H, E34H, V23H, and Q27H required an annealing temperature of 69°C, 68°C, 64°C, 70°C, 65°C, and 66°C respectively (represented by 'X' in the PCR cycle description above). Each time, the PCR product was purified using the QIAquick PCR purification kit (Qiagen).

Following mutagenic production and DNA sequencing to confirm mutations, each mutant was transformed separately into BL21 DE3-Gold ultra competent cells. Transformant screening and protein overexpression was performed in the same manner as wild type gpW. However, a low yield was revealed for each of the α-helical bi-his mutants making protein production and purification unfeasible. The β-strand mutants produced acceptable yields thereby allowing scale up production of V23H and Q27H. This was accomplished using the same protocol as wild type.

To attain a higher yield of bi-his mutants, the SUMO fusion system was utilized. The pSMT3 plasmid was donated from Dr. Christopher Lima (Mossessova, E. 2000). To subclone gpW bi-his mutants into the pET-28d vector containing the SUMO fusion protein the 5' NcoI restriction site was replaced with a BamHI site in the same manner the SalI site was replaced with HindIII in the wild type construct. The PCR program was identical to that used to generate the bi-his mutants, although an annealing temperature of 67^{o}C was used. The PCR product was sequentially digested using BamHI and HindIII restriction enzymes (Invitrogen) and purified after each single digest using the QIAquick kit. T4 DNA ligase was used to ligate the bi-his mutant gene at 4^{o}C in a thermal cycler overnight. The ligation product was transformed into BL21 DE3-Gold competent cells and positive transformants were selectively identified as having antibiotic resistance to 50 mg/mL kanamycin in addition to DNA sequencing. To screen for overexpression, 5 mL cultures were induced with 1 mM IPTG or 0.2% lactose once the OD_{600} reached 0.8 and incubated at 37, 30, and 20^{o}C. Time points were collected pre-induction, 4, 24, and 48 hours post induction. Expression levels were judged by visualization of SDS-PAGE as described before. 4 liter cultures were centrifuged in a Beckman centrifuge at 7000 rpm for 30 minutes to pellet the cells. The pellet was resuspended in 20 mM sodium phosphate, 500 mM sodium chloride, 20 mM imidazole pH 7.5. Cells were lysed by sonication and clarified by centrifugation in the Beckman ultracentrifuge at 25,000 rpm for 35 minutes, and filtered as before.

The ULP1 protease, which cleaves the SUMO fusion protein from the target protein (gpW), was donated with the gene subcloned in a pET-21d vector (Noavagen) with kanamycin resitance by Dr. Christopher Lima (Mossessova, E. 2000). The pULP1 plasmid was transformed into BL21 DE3 Gold *E. coli* cells and successful transformants were selected as described for pSMT3.

2.2 Purification

gpW wild type was purified with a Beckman dual channel HPLC with 126 pumps by injecting 5 mL aliquots of the filtered supernatant onto a XK26 size exclusion column packed with HiLoad 26/60 Superdex 75 prep grade resin (Amersham Biosciences). The mobile phase was 20 mM sodium phosphate 200 mM sodium chloride pH 6.0. Fractions were manually collected and analyzed for purity using SDS-PAGE. Fractions containing gpW were pooled and dialyzed exhaustively against nanopure water using dialysis cassettes with a 3000 molecular weight cutoff (MWCO) (Pierce) before lyophilization and storage at -20^{o}C.

gpW V23H and Q27H were purified after supernatant precipitation with 50% ammonium sulfate to remove large molecular weight proteins. The ammonium sulfate precipitate was resuspended in 50 mM Tris pH 8.0 and purified using SP Sepharose Fast Flow cation exchange resin (GE Healthcare). gpW was eluted using a linear gradient to 500 mM sodium chloride. Cation exchange was followed by

reversed phase chromatography using a Clipeus C18 10mm 250 x 20 mm semi/prep column (Higgins Analytical). The reversed phase running buffer consisted of 5% acetonitrile 0.1% trifluoro acetic acid. gpW mutant proteins were eluted by increasing to 95% acetonitrile using a linear gradient. Samples were analyzed for content and purity as described above, pooled, and directly lyophilized after freezing at -80 ^{o}C.

ULP1 protease was purified using a HisTrap FF affinity column (GE Healthcare) packed with 5 mL Ni-Sepharose resin per column. The resin was charged by injecting 2.5 mL 100 mM nickel sulfate for each column. Three columns were linked together in series and the supernatant was loaded onto the columns via a superloop with 50 mM Tris-HCl, 500 mM sodium chloride, 1 mM BME pH 8.0 as the running buffer. ULP1 was eluted by increasing the imidazole concentration from 20 to 500 mM with a linear gradient.

Mutant gpW-SUMO fusion constructs were purified with the same HisTrap columns. Supernatant was loaded as with ULP1 with 20 mM sodium phosphate, 500 mM sodium chloride, 20 mM imidazole as the running buffer. His-tagged gpW-SUMO fusion proteins were eluted via a linear gradient of from 20 to 500 mM imidazole.

The cleavage of the SUMO protein from gpW mutants was accomplished by combining pooled fractions of SUMO-gpW mutants respectively and adding ULP1 protease in a 1:100 (v:v) ratio and dialyzing overnight at room temperature in 2000

MWCO cassettes in 50 mM Tris-HCl, 500 mM sodium chloride, 1 mM BME pH 8.0. After dialysis, the cleavage solution containing SUMO, ULP1, and gpW was loaded onto two HisTrap columns linked in series and charged with Ni^{2+} ions. Due to the metal binding affinity of the bi-histidine mutation, gpW mutants were eluted using a linear gradient from 20 to 175 mM imidazole. SUMO and ULP1 were then eluted with a step to 500 mM imidazole. Pure gpW mutants were then exhaustively dialyzed in nanopure water before lyophilization and storage at -20°C.

2.3 Determining Protein Concentration

Wild type protein concentration was measured using a native molar extinction coefficient (ε_{native}) of 1888 M^{-1} cm^{-1} according to Beer's Law (equation 2.1) in a Cary 100 diode array spectrophotometer. Absorbance was measured at 275 nm, ε is the extinction coefficient, b is the cuvette pathlength, and c is concentration in molarity.

$$Abs_{275} = \varepsilon \times b \times c \tag{2.1}$$

The native extinction coefficient was determined by preparing in triplicate protein samples in water and 6 M guanidinium hydrochloride (GdnHCl). ε_{native} was calculated based on the following equation (2.2):

$$\varepsilon_{native} = \frac{Abs_{native} \times \varepsilon_{denat}}{Abs_{denat}} \tag{2.2}$$

where Abs is equal to the absorbance at 275nm corresponding to tyrosine absorbance in the UV absorbance spectrum. ε_{denat} is assumed to be 1280 M^{-1} cm^{-1} which is the extinction coefficient of free tyrosine.

Bi-histidine mutants exhibited more complex behavior and produced a variable measurement of absorbance by UV at 275 nm in native conditions despite multiple sample preparations and trials. Thus, the concentrations of these mutants were determined by denaturing a mutant sample from a concentrated stock solution. Samples were diluted 1:4 (protein:GdnHCl) and the concentration was calculated using Beer's Law. Since there is only one tyrosine in gpW, the extinction coefficient used to calculate concentration was 1280 M^{-1} cm^{-1}. This is the extinction coefficient corresponding to free tyrosine. This value is applied based on the assumption that the single tyrosine present in gpW is fully solvated in the unfolded state under the influence of chemical denaturant as confirmed by double perturbation studies on WT.

2.4 Circular Dichroism (CD)

Circular dichroism is a common technique used to reveal the conformation of biological molecules such as nucleic acids and proteins, which are composed of chiral, or optically active, components. The signal measured using CD is a function of the wavelength dependence of an optically active solution's ability to absorb circularly polarized light. Circular dichroism is expressed as the difference between the molar absorption of left- and right-handed circularly polarized light (Equation 2.3).

$$\Delta\varepsilon = \varepsilon_L - \varepsilon_R \tag{2.3}$$

where ε_L and ε_R are the molar absorptions of the left and right-handed circularly polarized light respectively.

The common unit of measure is ellipticity with units of millidegrees. The relationship between the differential absorption of circularly polarized light can be seen in equation 2.4

$$\theta_{obs}(\lambda) = 32.98(\Delta\varepsilon)\cdot l \cdot c \tag{2.4}$$

where λ is wavelength, l is cuvette pathlenth, and c is protein concentration in molar units.

As a means of normalizing the ellipticity signal in order to compare proteins varying in the number of amino acid residues and to correct for concentration differences, the mean residue ellipticity is calculated and plotted according to equation 2.5

$$[\theta] = \frac{\theta_{obs}(\lambda)}{n_{pb}\cdot 10 \cdot l \cdot c} \tag{2.5}$$

where n_{pb} is the number of peptide bonds in the protein (N-1 amino acids).

The far ultraviolet (F-UV) CD spectrum of a protein monitors excitations of the amide backbone over the wavelength range 250 – 190 nm. In this range, the

signal is dependent on secondary structure content. Elements of secondary structure can be followed by monitoring characteristic wavelengths indicative of the degree of 2^{o} structural formation. α-helices exhibit three characteristic bands, and often dominate a protein CD spectrum if present: a negative band at 222 nm which is a function of the n-π* excitation of the electrons on the carbonyl oxygen, a negative and positive band at 208 nm and 193 nm, respectively, corresponding to exciton splitting of the π-π* transition. β-strands can be characterized by a negative and positive band of similar magnitude at 217 nm and 195 nm, respectively. The band at 217 nm arises from the peptide n-π* transition while exciton splitting is responsible for the π-π* transitions at 195 nm. The absorbencies of β-strands have smaller amplitudes than α-helices, and are often dwarfed as a function of the proportion of helical residues and/or helix length. β-turns and loops are not discussed in detail here, but also exhibit wavelength dependent electronic transitions. It is also worth noting that aromatic residues can make weak, but detectable contributions to the F-UV CD spectrum.

The near ultraviolet (N-UV) CD spectrum takes advantage of the presence of aromatic residues (Trp, Tyr, Phe), which act as the chromophores giving rise to CD signal. These aromatic groups have characteristic π-π* absorbance bands found in the wavelength range of 320 – 250 nm. Positive signals from 250 – 270 can be attributed to phenylalanine residues, signals from 270 – 290 arise from tyrosine residues, and signals between 280 – 300 are attributable to tryptophan residues. Disulfide bonds

also contribute weak signals throughout the N-UV CD region, but are ignored here since the current model system lacks disulfide linkages. N-UV CD provides a measurement of the tertiary structure as a function of the environment surrounding the aromatic side chains, and thus act as good probes to conformational shifts or ligand binding.

Current Work: Far UV CD experiments were performed at a target protein concentration of 20 uM in 20 mM sodium phosphate buffer pH 6.0 with varying concentrations of guanidine hydrochloride using a pathlength of 1 mm. GdnHCl solutions were prepared in 20 mM sodium phosphate buffer pH 6.0 yielding final GdnHCl concentrations ranging from 0.0 to 3.0M. Final GdnHCl concentration was determined by refractometry using a Fisher Sci. refractometer.

Near UV CD experiments were performed with a target protein concentration of 80 μM in 20 mM sodium phosphate buffer pH 6.0 using a 1 cm pathlength.

Circular dichroism spectroscopy experiments were carried out in a Jasco J-810 spectropolarimeter thermostated with a Jasco Peltier-Type system. Wavelength scans were performed between 210 and 250 nm in continuous scanning mode at 10nm/min with a 2nm bandwidth and a response time of 16 seconds. Far-UV CD measurements in the absence of GdnHCl were monitored over a wavelength range of 250-190. Wavelength scans in the near UV were performed between 320 – 250 nm using otherwise identical scanning parameters to far UV measurements. All temperature wavelength scans were performed over a temperature range of 268-362 K with 3 K

increments. Temperature ramps had a slope of 3 K/min with a 2 minute equilibration prior to data acquisition at each temperature.

2.5 Fluorescence

As with absorption spectroscopy like CD, fluorescence takes advantage of chromophores that absorb light. The signal output is a product of the physical and chemical environment. The common chromophores for proteins are aromatic residues which take advantage of the network on non-bonding electrons in the aromatic ring structure. The fluorescence of a chromophore is dependent on a molecule absorbing light and transitioning from the ground state to the excited state. If the energy absorbed exceeds the quantity needed to reach the first electronic excited state, the excess energy is dissipated as vibrational energy through heat or interaction with solvent. In order to return to the ground state, light is emitted or transferred non-radiatively. Fluorescence is a measure of the lifetime of a photon emitted as a byproduct of the decay of the excited state. This is expressed as quantum yield, Q, which is the ratio of the number of photons emitted to the number of photons absorbed (equation 2.6) where a photon is defined by $E = h \cdot v$ where *h* is Plank's constant and *v* is the frequency of the light.

$$Q = \frac{photons_{emitted}}{photons_{absorbed}} \tag{2.6}$$

Since fluorescence is the result of a degenerative process whereby the byproduct, emitted light, is lower in energy than the incident photon, *Q* is under the control of internal and external factors such as the number of vibrational levels, rigidity of the chromophore, and environment. Internal factors are of less concern in biological molecules because the absorbing fluorophores are typically rigid aromatic rings with delocalized π electrons and characterized by π-π* transitions resulting in strong absorbance and fluorescence. The environmental factors most important in influencing the reduction of the quantum yield is broadly referred to as quenching, which is defined as collisional processes affected by solvent, temperature, pH, neighboring residues, and fluorophore concentration. Thus, fluorescence is a powerful tool to monitor the tertiary structure of proteins containing either intrinsic or unnatural fluorophores.

Current Work: Fluorescence emission experiments of gpW were dependent on the excitation of the intrinsic tyrosine residue. Measurements were collected using a Flurolog-3 Spectrofluorimeter (Jobin Yovin, Inc.) with temperature control regulated by a Peltier system. All measurements were performed using a 1 cm pathlength quartz cuvette at a protein concentration of 6 μM. The excitation wavelength was 280 nm. Excitation and emission slit widths were 2 nm with an integration time of 0.25 seconds. The temperature dependent emission scans were performed every 2 K between the range of 273 – 361 K and were equilibrated for 2 minutes prior to data collection.

2.6 Fourier Transform Infrared Spectroscopy (FTIR)

A member of absorption spectroscopy, FTIR is another method to investigate protein structure. FTIR spectra are generated by transitions between ground state vibrational levels as a result of light absorption. The modes of vibration of functional groups such as amides, carbonyls, and methyl groups are sensitive to changes in structure, conformation, and environment. The characteristic stretching and bending vibrations give rise to infrared absorption as a function of the atomic dipole moment. The six types of vibrations are as follows: symmetrical stretching, asymmetrical stretching, in-plane scissoring and rocking, and out-of-plane wagging and twisting. Hydrogen bonding and the presence of secondary structure are the major factors which determine band intensity.

Bands in the amide I region are primarily the result of C=O stretching. Small contributions to this region are also made by C-N stretching and N-H bending motions. The amide I region is defined by the wavelength range 1600 – 1700 cm^{-1}. Secondary structural elements can be monitored in this region; namely the α-helical conformation (1632 and 1650 cm^{-1}), β-sheet conformation (1685 cm^{-1}), and random coil conformations (1658 cm^{-1}). The major drawback of monitoring the amide I region is the overlapping of secondary structure peaks. However, the protein signal obtained in this region has much higher amplitude compared to the amide II and amide III regions. Another drawback is that the –OH stretching vibrations from aqueous solvent (water) and water vapor significantly absorb in this region, and must

be subtracted from the protein solution spectrum. One way to get around this is to use deuterated solvents to reduce the absorbance of water, degass samples, and exhaustively purge the sample compartment with dry nitrogen gas.

The amide II (1480 – 1575) and the amide III regions (1220 – 1320) also provide some information regarding secondary structure, but more as a means of support to the amide I absorptions due to a protein's relatively weak signal strength at lower wavenumber.

Current Work: Equilibrium FTIR experiments were performed on a thermostated FTS-300 IR Spectrometer (BioRad) at 2 cm^{-1} resolution using $CaCl_2$ windows and a 50 μm Teflon spacer. The protein sample was prepared at 4 mg/mL concentration in 99.9% D_2O, 20mM Sodium Phosphate buffer at pH 7. Temperature dependent measurements were performed between the ranges of 277 – 365 K regulated with a circulating water bath. The sample cell temperature was monitored with a Fluke 50s thermocouple.

2.7 Differential Scanning Calorimetry

DSC provides a unique opportunity to observe the geography between the unfolded and folded states. Using this technique, the absolute heat capacities of the native and denatured states can be defined and the thermodynamic effect of this transition between macrostates can be evaluated.

The instrument consists of two identical cells, one as a reference for the buffer, and another for the protein sample. The basis for this design is to measure the difference in the electrical power required to maintain both cells at equal temperature. The differential power ($\mu J\ s^{-1}$) is normalized by the scanning rate to reveal the difference in heat capacity between both cells ($\mu J/deg$). The apparent heat capacity of the protein solution can then be directly extracted according to equation 2.7 where C_p^{sol} is the absolute heat capacity of the protein solution and C_p^{solv} is the absolute heat capacity of the solvent (buffer only).

$$\Delta C_p^{app} = C_p^{sol} - C_p^{solv} \tag{2.7}$$

As part of the routine calibration of the instrument, a buffer-buffer baseline, where buffer is loaded in both cells, is evaluated both before and after the scan of the protein solution to monitor for baseline drift and correct for instrumental effects. It is also important to use the dialysis buffer as a reference to replicate the solvent as closely as possible in the protein solution.

Calculating an accurate protein concentration is paramount. Small errors in concentration are propagated in the analysis of the partial molar heat capacity as indicated in equation 2.8 where c is concentration in mM, V_o is the calorimeter cell volume in mL, and V_{prot} and V_{solv} are the protein and solvent molar volumes, respectively. This calculation of the absolute heat capacity of the protein is a necessary product to directly calculate the barrier height from DSC thermograms.

$$C_p^{prot} = \frac{\Delta C_p^{app}}{c \cdot V_o \cdot 10^{-6}} + \frac{V^{prot}}{V^{solv}} C_p^{solv} \tag{2.8}$$

Another dimension of the concentration issue is the formation of aggregates upon thermal unfolding due to high concentrations. To circumvent this latter problem, multiple protein concentrations are scanned over the temperature range to more accurately calculate the protein's absolute heat capacity (Kholodenko, V. 1999, Georgescu, R. E. 2001). A more thorough explanation of the absolute heat capacity calculation, the physical and chemical information obtained by DSC, and the direct extraction of the barrier height will presented with the equilibrium characterization of gpW found in chapter 3.

Current Work: DSC experiments were performed in collaboration with Dr. Jose M. Sanchez-Ruiz at the University of Granada, Spain with a VP-DSC calorimeter from MicroCal (Northampton, MA) at a scan rate of 1.5 K/min. Protein samples were prepared by exhaustive dialysis against the 20 mM sodium phosphate pH 6.0. Samples were degassed and kept under 30 psi excess pressure in the 0.5mL calorimetric cells to prevent degassing. Proper equilibration of the instrument was determined by performing several buffer-buffer baselines before each experiment. Several reheating runs were performed to determine the reversibility of denaturation. An additional buffer baseline was obtained following the experiment to monitor possible drift in the instrument baseline. Baseline reproducibility was very high and consistent with previous experiments performed with the same instrument. Three

independent DSC experiments were performed under identical buffer concentrations, but with protein concentrations ranging from 0.6 – 1.0 mg/mL. Absolute heat capacity values and associated errors were calculated from the protein concentration dependence of the apparent heat capacities, as described above.

2.8 Time-resolved Infrared Temperature Jump Kinetics

Ultrafast folding proteins cannot be studied with the conventional stopped flow technique, which has a dead-time of ~1 – 2 milliseconds. Continuous flow devices reduce the dead-time to the tens of microseconds, but this limits the temperature range from which to monitor the relaxation rate. A laser-induced temperature jump apparatus is the favored method by which to study microsecond folding proteins. The measurement relies on a pump/probe setup whereby a pump beam is shifted by a Raman cell to the wavelength that allows vibrational excitation of water. Once this beam is focused on the sample, a temperature jump is achieved causing the solvent to spontaneously heat up in the nanosecond time scale. The relaxation from the elevated temperature to the original equilibrium temperature gives rise to the folding kinetic amplitudes that are monitored by a probe beam.

Current Work: The fundamental of a Continuum Surelite I-10 Nd-YAG laser (1064 nm) with a 7 ns pulse-width and run at 2 Hz was shifted to ~1.9 μm with a 1 m path length Raman cell (Lightage) filled with a mixture of Ar and H_2 at 1000 psi to heat water by vibrational excitation. Heating pulses of ~20 mJ was used to generate

~10 – 12 K jumps. A continuous wave (CW) lead salt diode laser (Laser Components) and a MCT (Mercury-Cadmium-Telluride) detector (Kolmar Technologies) with 50MHz bandwidth was used to monitor time-dependent changes in IR absorption at 1632 cm^{-1}. D_2O buffer was used for background subtraction and also as an internal thermometer to determine the T-jump magnitude. The IR cell and sample preparation were identical to the equilibrium FTIR experiments.

3. Multiprobe Characterization of gpW

3.1 Introduction to gpW

Translation of the bacteriophage lambda (λ) gene W results in the production of gene product W (gpW), a small single domain α/β protein with a novel fold. A search of structural databases failed to produce any structural homologs, thus gpW is the smallest monomeric protein containing both alpha and beta structure without disulfide bonds or cofactors. gpW is organized by two anti-parallel α-helices pinned against a single β-sheet composed of two anti-parallel strands. The full length construct is composed of 68 amino acids and the three-dimensional solution structure was resolved by NMR (Maxwell K. L. 2002) (pdb ID 1HYW). Residues 4-17 (α1) and 40-52 (α2) comprise the first and second α-helices respectively while residues 23-27 (β1) and 31-35 (β2) form the two β-strands. gpW also contains an intrinsic chromophore, in the form of the amino acid tyrosine, that serves as a probe of tertiary structure and is important in determining protein concentration (Figure 3.1). According to the assignment of structured residues in the solution structure, gpW possesses a long unstructured C-terminal tail. This tail contains mostly polar residues with multiple glycine residues and a proline known to destabilize secondary structure. A well defined hydrophobic core is assembled via contribution from each of the

secondary structures (Leu7, Ala8, Ala10, Leu14, Leu 17, Ala24, Val33, Val40, Leu43, Ile47, and Leu50).

At first glance, initial predictions would suggest gpW to exhibit two-state behavior with a folding rate in the millisecond timescale. The helical structure prediction algorithm AGADIR predicts a very high intrinsic helical propensity for both $\alpha 1$ and $\alpha 2$ suggesting a large structural contribution from local interactions. A contact map of gpW shows a large number of local interactions as expected for a protein with α-helical features (Figure 3.2). Long range contacts between the N- and C-termini of the two helices are an additional feature illustrated by the contact map indicating their close proximity due to packing against the β-hairpin and desolvating the hydrophobic core. The relative contact order (RCO) for gpW was calculated from the protein databank (pdb) file (pdb ID 1HYW) and is dependent on both the distance between all atoms and overall chain topology. RCO is smaller for proteins that are stabilized primarily by local interactions and is higher when long-range contacts are formed. To calculate RCO, the distances between hydrogen and any other atoms are screened out in addition to distances beyond a cutoff of 6 angstroms. In other words, only distances between non-hydrogen (heavy atoms) are taken. Any two residues are considered to be in contact if at least one of their atoms is involved in contact. The residue separation for each pair of contacting residues is computed and the resulting contact order is simply calculated as the sum of residue separation divided by the protein length multiplied by the number of contacts considered. Formally, RCO is calculated by:

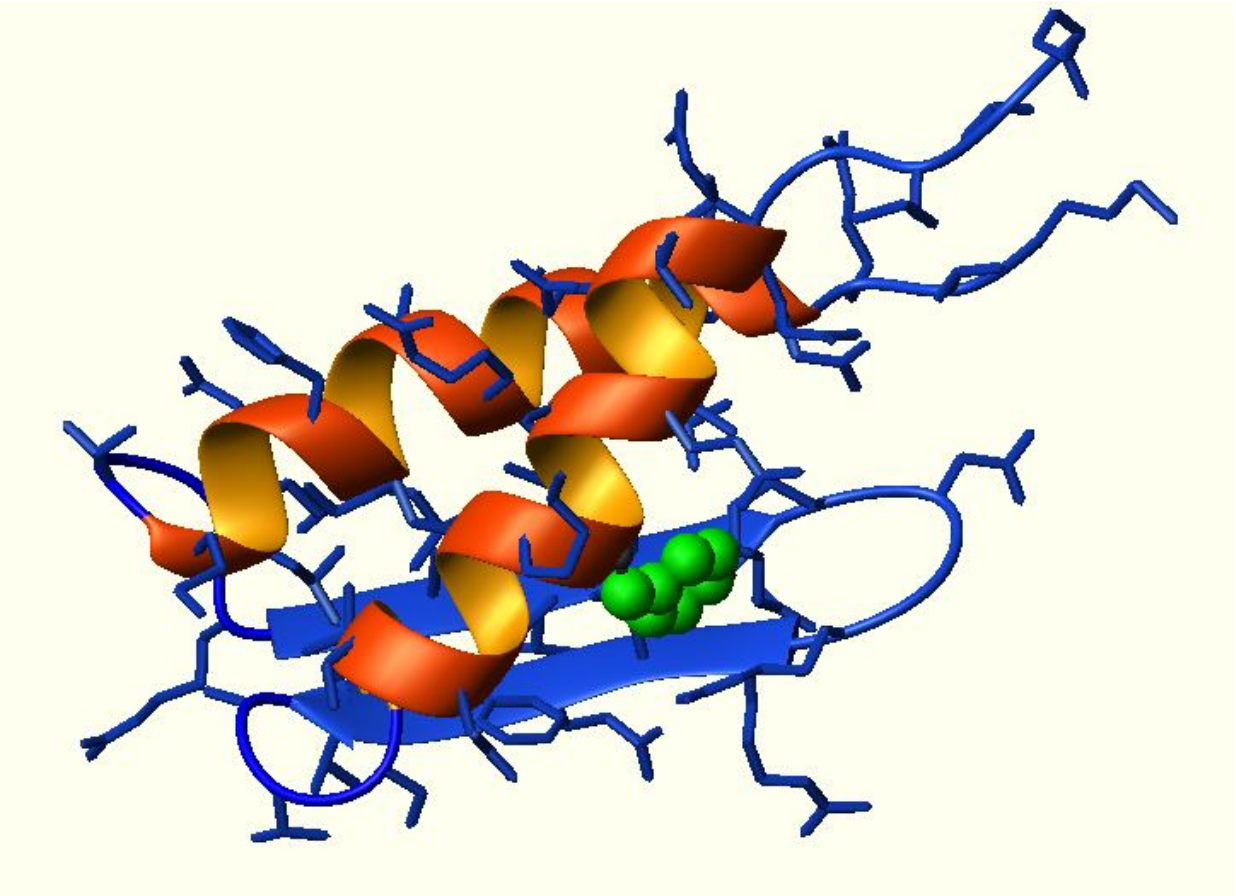

Figure 3.1 Three-dimensional representation of gpW secondary structure and overall topology. Side-chains are depicted as sticks and are colored blue. The single native tyrosine on α-helix 2 is shown in ball-and-stick format colored green. The long C-terminal tail is not pictured due to its unstructured conformation, and thus its absence from the NMR resolved pdb entry (pdb ID 1HYW) (Maxwell, K. L. 2001) (Graphic generated by MolMol Koradi, R. 1996)

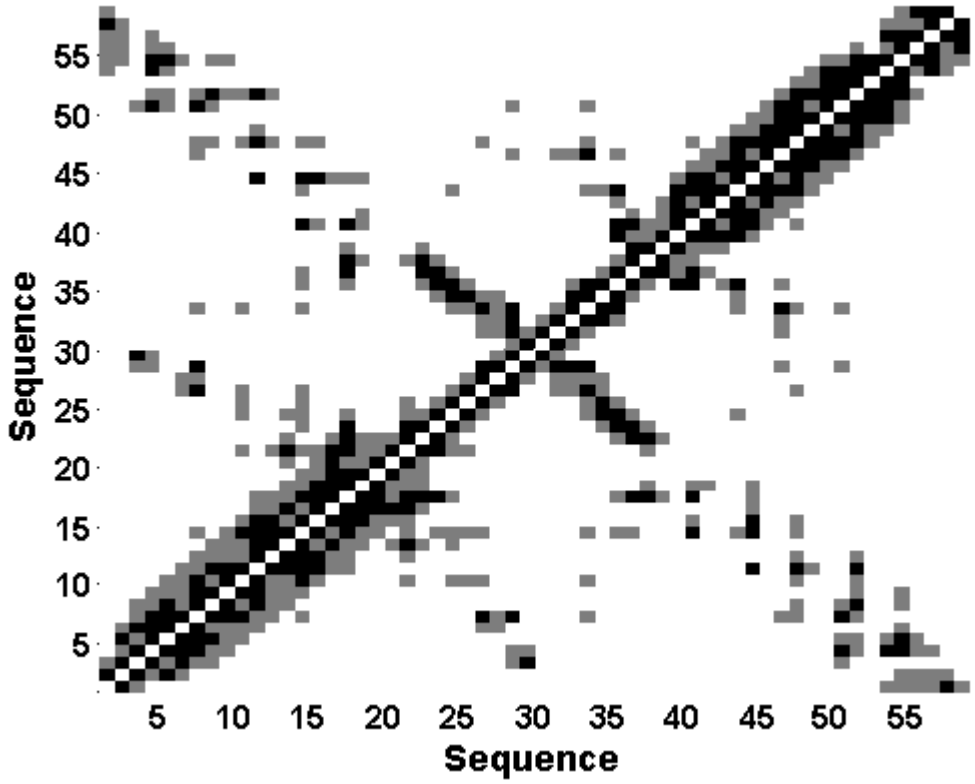

Figure 3.2 Contact map of gpW of all non-hydrogen atoms with 6Å cutoff. Strong interactions (atomic contacts greater than 20) are shown in black while weaker contact order (atomic contacts greater than 0 but less than 20) is shown in grey. Backbone interactions dominate the diagonal in the α-helical regions as expected. Long-range contacts are formed due to the hydrogen bonding between the two β-strands, between α1 and α2, and from the two helices packing against the β-hairpin loop.

$$RCO = \frac{1}{L \cdot N} \sum^{N} \Delta L_{ij} \tag{3.1}$$

where N is the number of contacts (within the 6Å cutoff) between heavy atoms, L is the number of amino acids, and ΔL_{ij} is the number of residues separating the interacting pair (Plaxco K. W. 1998). The RCO for gpW is calculated to be 15.323, which is in accordance with a protein folding in the millisecond range. The absolute contact order (ACO) is reported to be 8.8875 and is equal to RCO multiplied by protein length. This parameter reports on the contact order as a function of protein size, but has been found to be less correlated than RCO with folding rates of two-state folding proteins.

A feature based on an experimental correlation between folding rate and size suggests gpW should follow two-state behavior. Being composed of 68 amino acids, the length of gpW is similar to that of other two-state-like proteins and much longer than the ultra-fast folding proteins which fold on the order if microseconds. A recent development by Naganathan and Muñoz building on the work of Thirumalai (Thirumalai, D. J. 1995) found a strong correlation between folding rates and protein size of 69 proteins and peptides (including proteins exhibiting downhill, two-state, and three-state behavior) ranging in length from 16-396 residues (Naganathan, A. N. 2005b). For a protein of 68 residues, folding rate is expected to be in the range of milliseconds, corresponding well to predictions from contact order. Additionally, estimating the barrier height of this protein as a function of this correlation implies a

significant free energy barrier of approximately 8-30 kJ mol^{-1} estimated from the pool of proteins for which thermodynamic data is available.

The WT gpW construct used in this work was truncated to 62 residues by removing 6 unstructured, destabilizing, and hydrophobic C-terminal residues to resolve solubility issues discovered in NMR studies. The truncation was also made to eliminate N- and C- terminal tags inserted by Maxwell for ease of purification (Maxwell K. L. 2001). This does not affect the contact order calculations and resulting folding time prediction as the truncated residues are not included in the pdb sequence. The length dependent correlation holds true for this shorter 'wild type' sequence as well.

Taking advantage of the findings from the metal-binding domains discussed in section 1.4 and building on the concept of tunable stability presented by Sosnick while assuming two-state folding, gpW mutants were originally to be engineered with the intention of exploring the formation of structure in the transition state ensemble. Since both the ϕ- and ψ-value analyses are unable to elucidate structural formation due to their one-dimensional origin, expanding these analyses into a second dimension should provide the level of detail necessary to characterize the high energy species. By introducing a bi-histidine site at various points in the protein, the local stability of this site could be tuned with increasing metal concentration. Coupling this local perturbation with a point mutation at another location within the protein should enable monitoring the effect of local perturbations on global structure and stability.

This coupling can be conceptualized by the increase in the level of detailed gained by performing two-, or three-dimensional nuclear magnetic resonance as compared to one-dimensional experiments.

Five bi-histidine mutant constructs were engineered at various positions throughout the protein. Three mutants were designed to probe the global stability as a function of tunable stability in the α-helical sequences, while two mutants were designed to probe stability induced by pinning the two β-strands together. The α-helical mutants were engineered taking advantage of the i, i+4 orientations of helix side-chain moieties and are as follows: E5H E6G A9H (N-terminal half of α1), R11H A12G (C-terminal half of α1 and utilizing the single native histidine at position 15), K45H A48G E49H (central region of α2). The two β-sheet mutants were engineered such that the solvent exposed face of the β-sheet could incorporate the bi-histidine site with one histidine on each β-strand so metal binding would facilitate the proximity of the two strands to initiate and stabilize the hydrogen-bonded network. For the mutations E34H V23H and E34H Q27H, both shared the mutation E34H in the middle of β2, but differentially served as nucleation sites near the C-terminal end of β1 (V23H) or nearer to the β-hairpin (Q27H). The helical mutants employ i, i+4 bi-histidine sites with a glycine residue substituted in between. The purpose of this is to ensure destabilization of the helix with the hopes of rescuing stability upon metal binding. Glycine is known to be highly destabilizing with low helical propensity due to conformational flexibility imparted by its lack of sidechain. Of particular interest,

after imputting each of the mutant sequences in AGADIR, the predicted helical propensity for helical mutants at experimental conditions is approximately 50% of that for wild-type suggesting a potentially significant impact on helix stability upon mutation. For the β mutants, the predicted helical propensity is expectedly the same as wild-type.

Contact order and sequence length suggest gpW is a two-state folder. The rest of this chapter is dedicated to the experimental equilibrium characterization of wild type and bi-histidine mutant gpW variants and a description of the models and fitting methods. Section 3.2 discusses a double perturbation experiment of wild type gpW comparing the unfolding behavior as a function of coupled denaturants. Section 3.3 is a presentation of the experimental characterization of backbone secondary structure indicative of the local environment and its unfolding pattern. Section 3.4 focuses on the global environment using the lone native tyrosine residue as a reporter of tertiary structure and section 3.5 offers conclusions developed from the experimental data.

As was demonstrated in comparing the unfolding events of BBL at atomic resolution and comparatively low resolution techniques, circular dichroism proved to recapitulate the ensemble average of the atom-by-atom behavior. This is certainly expected, but a further product of the analyses developed by Muñoz compels experimentalists to use multiple techniques probing local and tertiary structural features to achieve a comprehensive view of protein folding / unfolding mechanism by exploring the agreement between methods.

3.2 Coupled Denaturant Unfolding

To attain a complete picture to determine the folding mechanism of gpW, first a temperature-GdnHCl double perturbation experiment following the unfolding transition at 222 nm was employed to monitor the complex coupling between denaturants. Extending beyond a single perturbation experiment which could be described as chemical denaturation held at constant temperature, double perturbation experiments take advantage of the different mechanisms influencing unfolding between denaturing agents. In two-state proteins, the addition of chemical denaturant is reflected by altering the barrier height between the unfolded and folded states. Thus, adding GdnHCl shifts the equilibrium toward the unfolded state, but does not significantly alter the structure of the ground states. In contrast, the free energy profile of a downhill folder will become broader as a response to gains in conformational entropy in the unimodal distribution. For a two-state system, coupling between denaturing agents may give rise to varying responses, but the experimental output remains a linear combination of the two ground states and should follow a Maxwell relationship (Oliva, F. Y. 2004). For a downhill folding protein, the continuous shift in structural sampling prohibits a linear description of the structural ensemble and provides individual snapshots in the unfolding transition.

Figure 3.3 is a plot of the double perturbation experiment of gpW showing the CD mean residue ellipticity (MRE) at 222 nm. The unfolding curves are expectedly sigmoidal at varying GdnHCl concentrations with distinct curvature at low

temperature becoming apparent at increasing GdnHCl concentrations displaying the onset of cold denaturation; a sign of a shift in the conformational ensemble. A global two-state fit is also pictured and is shown as red curves overlaying each dataset. The fitting model assumes a linear dependence of both ΔH and ΔS on denaturant concentration as expected for a two-state protein (Equations 3.2, 3.3, and 3.4)

$$\Delta H([D]) = \Delta H_m^o + v_1[D] \tag{3.2}$$

$$\Delta S([D]) = \Delta S_m^o + v_2[D] \tag{3.3}$$

$$\Delta S_m^o = \Delta H_m^o / T_m \tag{3.4}$$

where ΔH_m^o and ΔS_m^o are the denaturant free enthalpy and entropy at the melting temperature (T_m) (where p_f = p_u and ΔG=0). To complete the thermodynamic description of the system, the temperature dependent equilibrium constant (K(T)) (Equation 3.5), the partition function (Q) and folding (p_f) and unfolding probabilities (p_u) (Equation 3.6 and 3.7), and Gibbs free energy are included (Equation 3.8).

$$K(T) = \frac{[U]}{[F]} = \frac{k_u}{k_f} = e^{\left(\frac{-\Delta G(T)}{RT}\right)} \tag{3.5}$$

$$Q = 1 + K(T) \tag{3.6}$$

$$p_f = \frac{1}{1+K(T)} \text{ and } p_u = \frac{K(T)}{1+K(T)} \tag{3.7}$$

$$\Delta G(T) = \Delta H_m + \Delta C_p(T - T_m) - T\left[\Delta S_m + \Delta C_p \ln\left(\frac{T}{T_m}\right)\right] \quad (3.8)$$

ΔH_m and ΔS_m are the enthalpy and entropy of the reference state and ΔC_p is the change in heat capacity between the folded and unfolded states.

The denaturant dependent folded baseline (S_f) is linearly dependent on temperature (T) (Equation 3.9). Similarly, the unfolded baseline (S_u) is both linearly dependent on temperature (T) and denaturant [D] where $v_3 - v_7$ are floating fitting parameters which define the shape, intercept, temperature and chemical denaturant dependence of the baselines (Equation 3.10).

$$S_f = v_3 + v_4(T - T_{ref}) \quad (3.9)$$

$$S_u = v_5 + v_6(T - T_{ref}) + v_7[D] \quad (3.10)$$

A total of 10 parameters are necessary to develop the global two-state fit (ΔH, ΔS, T_m, v_1-v_7) with an equal number of parameters describing the baselines as thermodynamic parameters. This highlights a particular importance in defining experimental baselines.

At a first approximation, the global two-state fit appears to mirror the raw data very well. Upon closer inspection however, the fits at lower GdnHCl concentrations are poorly replicated. For instance, at 0 M GdnHCl, the stability is underestimated, while overestimating stability at 1 M. This same overestimation appears systematic

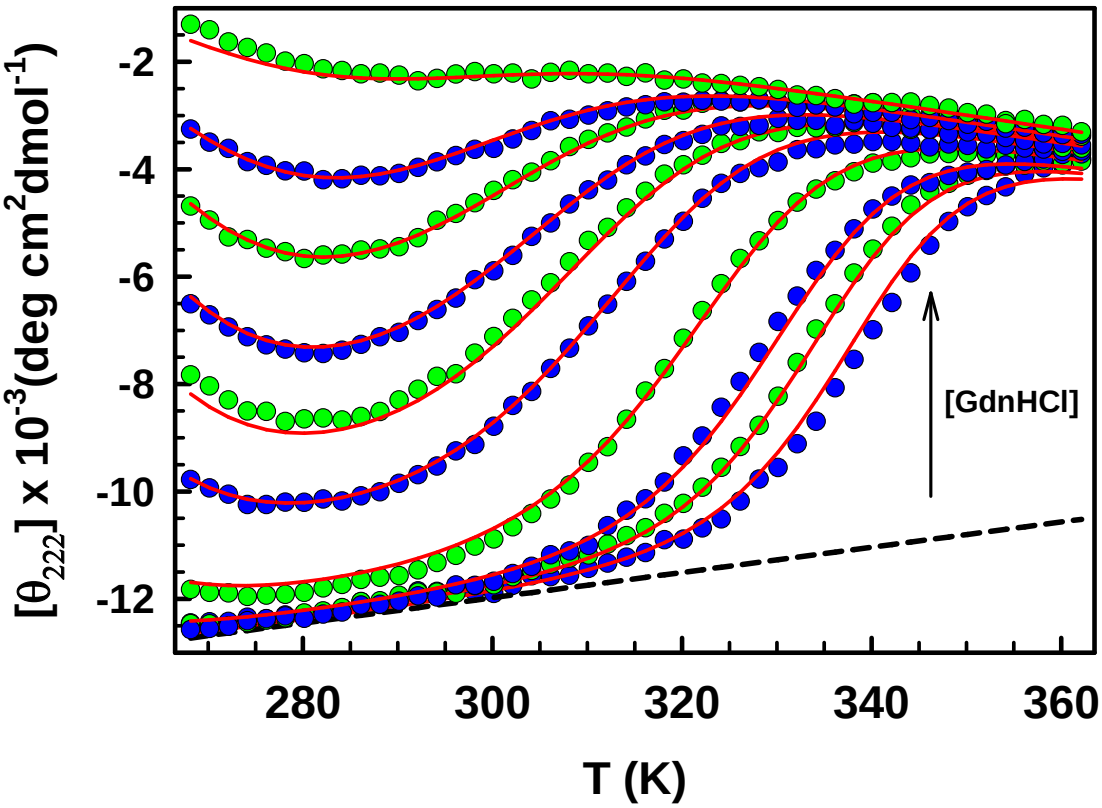

Figure 3.3 Temperature-GdnHCl double perturbation experiment of gpW. GdnHCl concentrations are 0.0, 0.25, 0.5, 1.0, 1.5, 1.75, 2.0, 2.25, 2.5, and 3.0 M with blue and green symbols distinguishing each step in denaturant concentration. Red lines correspond to the best fit of a global two-state model. The black dashed line represents the native baseline.

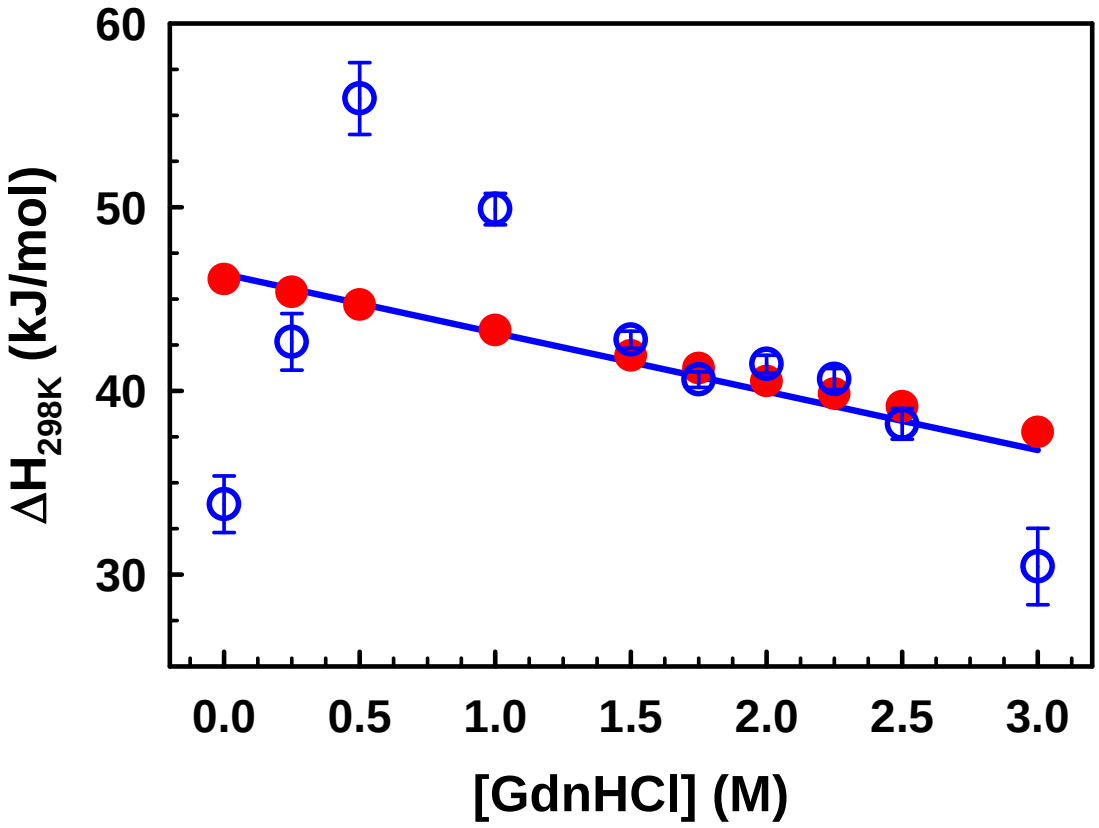

Figure 3.4 A plot of the apparent unfolding enthalpy (ΔH_{298K}) at 298 K on GdnHCl concentration (open blue circles). Also plotted are the values obtained from the global fit (filled red circles) and a linear regression of the values from individual fits (solid blue line).

as it can also be seen at 0.25, 0.5, 1.0, and 1.75 M GdnHCl. This suggests that, although the error is small, the number of fitting parameters, especially in defining S_f and S_u, are too many for an accurate representation of the raw data.

To alleviate these fitting errors and provide a reference for comparison of best-fit parameters, two-state fits to each individual GdnHCl dependent thermal unfolding curve can be done. This is achieved by fixing the native baseline and ΔC_p obtained from the global fit as a starting reference for the 0 M curve. For each subsequent unfolding curve, the parameters from the previous single fit were used as reference. Thus, only the T_m, ΔH, and two parameters to describe the unfolding baseline are necessary. As mentioned earlier, ΔH for a two-state system is expected to follow a linear dependence with denaturant concentration as only two states are populated. As can be seen in Figure 3.4, the apparent unfolding enthalpy at 298 K (blue circles) does not follow the expected pattern for a two-state folding protein. Instead, a sharply curving plot is observed with a maximum at 0.50 M GdnHCl. The reason for this result is that for a downhill system, as GdnHCl perturbs the ensemble, the free energy basin becomes flatter as more entropy is introduced reflective of the structural continuum. A plot of ΔH from each unfolding curve obtained from the global fit imitates the expected linear dependence for a two-state system with a linear regression of the individual fits overlaying nearly identically. This indicates that the complex coupling between denaturants cannot be represented by the global fit. An

identical result is observed in a plot of ΔS versus GdnHCl, where again, a linear dependence would be expected for a two-state system (data not shown).

A second signature of downhill folding that can be obtained is illustrated once again in the global two-state fit of the double perturbation plot, notably T_{max}. This is the temperature at which the signal at 222 nm exhibits its maximum (most negative mean residue ellipticity) in each GdnHCl denaturation curve. In a two-state example, T_{max} varies minimally and is expected to produce a nearly horizontal line as a function of chemical denaturant due to a well defined population of the ensemble. In a downhill scenario, T_{max} progressively shifts to the right with increasing chemical denaturant as a result of a continual succession of structure formation. As can be seen in Figure 3.3, T_{max} for gpW shifts from 268 at 0 M GdnHCl to ~290 at 3 M GdnHCl; a shift of ~22 degrees which is evidence for a downhill folder.

3.3 Backbone Interactions in gpW Folding

The implementation of multiple techniques is an essential practice in order to diagnose the folding mechanism for a protein (i.e. downhill or two-state). A two-state system is expected to exhibit identical behavior independent of the method used to monitor protein unfolding. The double perturbation experiments discussed above have offered initial proof that gpW is not a two-state system. To further explore this

result, the thermal stability of the peptide backbone was investigated using far-uv circular dichroism (f-CD) and Fourier Transform infrared (FTIR) spectroscopy.

3.3.1 Far-UV Circular Dichroism

F-CD is a robust method to monitor the ensemble average of the peptide backbone, thus providing an indirect measure of secondary structure. The presence of α-helical structure gives rise to dominant expression of the corresponding characteristic wavelengths in the f-CD spectrum (i.e. 193, 208, and 222 nm) and is dependent on helix length which can be examined by spectral decomposition. Since the native structure of gpW contains a significant amount of helical identity, f-CD is an excellent method to track backbone unfolding. Figure 3.5A is a plot of the temperature dependent matrix of spectra collected at pH 6.0 in 20 mM sodium phosphate buffer. At low temperature, characteristic bands of α-helical structure are readily apparent. As temperature is increased, the signal at those wavelengths diminishes while the appearance of bands associated with random coil structure dominates at high temperature. By following the temperature dependent ensemble transition over the entire far-UV spectrum, a simple test analogous to probe dependence can be completed by comparing the wavelength dependence of unfolding. Strictly for a two-state system, coincidence of unfolding behavior is expected independent of the wavelength monitored. Comparatively, as the structural ensemble of a downhill system shifts with temperature, varying populations of helix

and coil are expected. Figure 3.5b illustrates the wavelength dependence monitored at wavelengths indicative of helical (193, 208, 222) and coil (200) structure. The T_ms obtained by taking the first derivative of the unfolding curves show minor deviations spanning a range of 5 K (data not shown). Another difference seen is in the slopes of the pre- and post-transition baselines. The observed change is less dramatic than is seen for the global downhill folder BBL, suggesting the existence of a free energy barrier. The spread of melting temperatures and dissimilar slopes are narrowly different. Nonetheless, this result offers further evidence that gpW is non-two-state.

The temperature-wavelength spectra permits snapshots to be taken of the structural ensemble along the free energy reaction coordinate using singular value decomposition (SVD). This analysis separates the signals contributing to the overall spectra. By ranking these contributions, the signals that are a result of instrument noise can be filtered out leaving only the major sources of the signal intact. The temperature-wavelength spectra pictured in figure 3.5A are recomposed from the first three components of the SVD calculation. Figure 3.6 is a collective illustration of these three most important components. Panel A shows the relative f-CD signal which represents the average spectrum over the temperature range. This basis spectrum contributes ~81 % to the total and is indicative of α-helical structure with negative bands at 208 and 222 and a positive band at 193. The normalized amplitude of this first component in panel D shows a transition demonstrating the high degree of helical character at low temperature melting away as temperature is increased. A

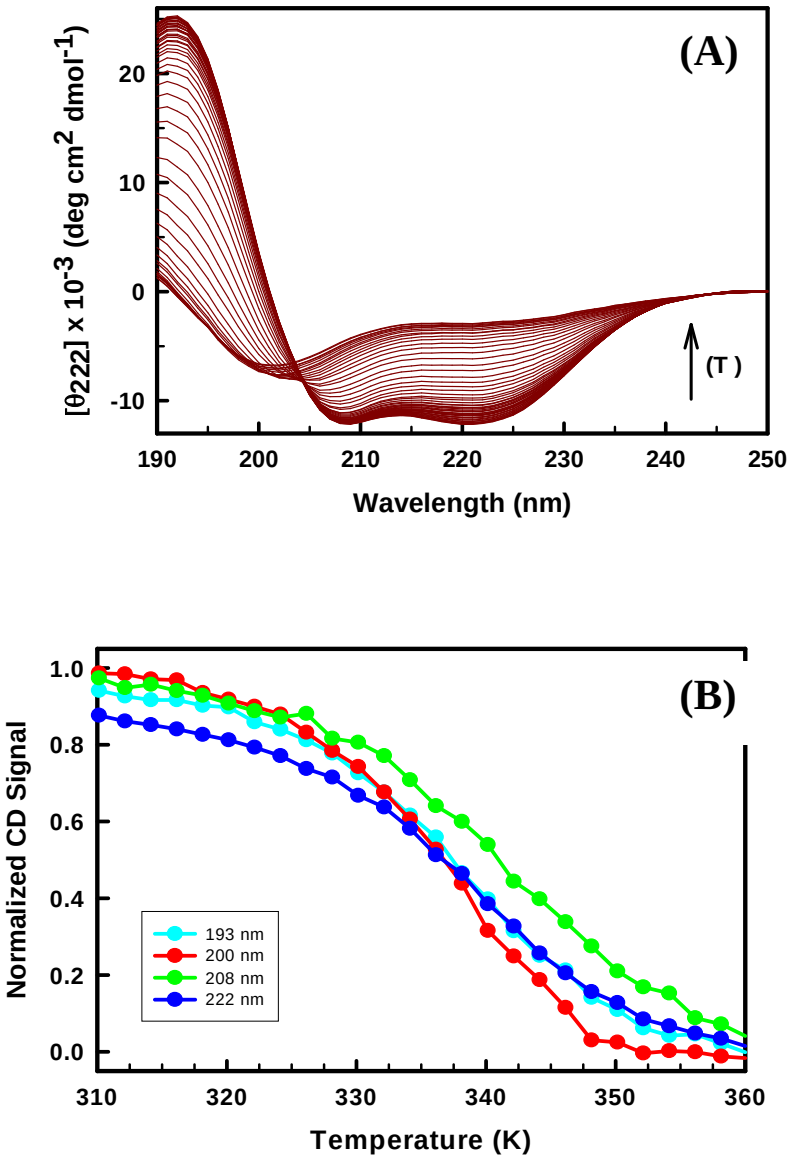

Figure 3.5 (A) WT gpW far-ultraviolet circular dichroism spectra over a temperature range from 268 to 372 K progressing from a high degree of secondary to a random coil like spectrum at high temperature. (B) Wavelength dependence of wild type gpW unfolding as monitored by far-ultraviolet circular dichroism.

picture of contrast to the helical content is shown in panel B. This second component (~18 % of the total) is indicative of a random coil spectrum with the amplitude (Panel E) a mirror image of that for the helix. As observed in the panel A, the relative signal at 208 and 222 nm are nearly equal with a slightly more negative signal at 208 nm. This suggests medium-length helices are involved in the transition as longer helices are characterized by a very large negative signal at 222 nm while shorter helices have a greater negative band at 208 nm (Garcia-Mira, M., personal experiments with myoglobin and cytochrome C). The third component, contributing less than 1% of the total, from the SVD analysis (Panel C) is a picture of this transition showing the pattern of change in signal strength. The amplitude shown in panel F suggests that as temperature rises, helical content decreases (shorter helices) before denaturing to a predominantly coil-like structure. The appearance of shorter helices points to fraying at the ends, as a minor loss in enthalpy accompanies a large gain in conformational entropy. Thus, the appearance of a third component in f-CD measurements of helical proteins allows a quantitative measure of helix-length. Even though this component is such a minor contributor to the sum, its significance cannot be discounted. This just means the total differential of signal change is small, and therefore emphasizes the power of the SVD analysis.

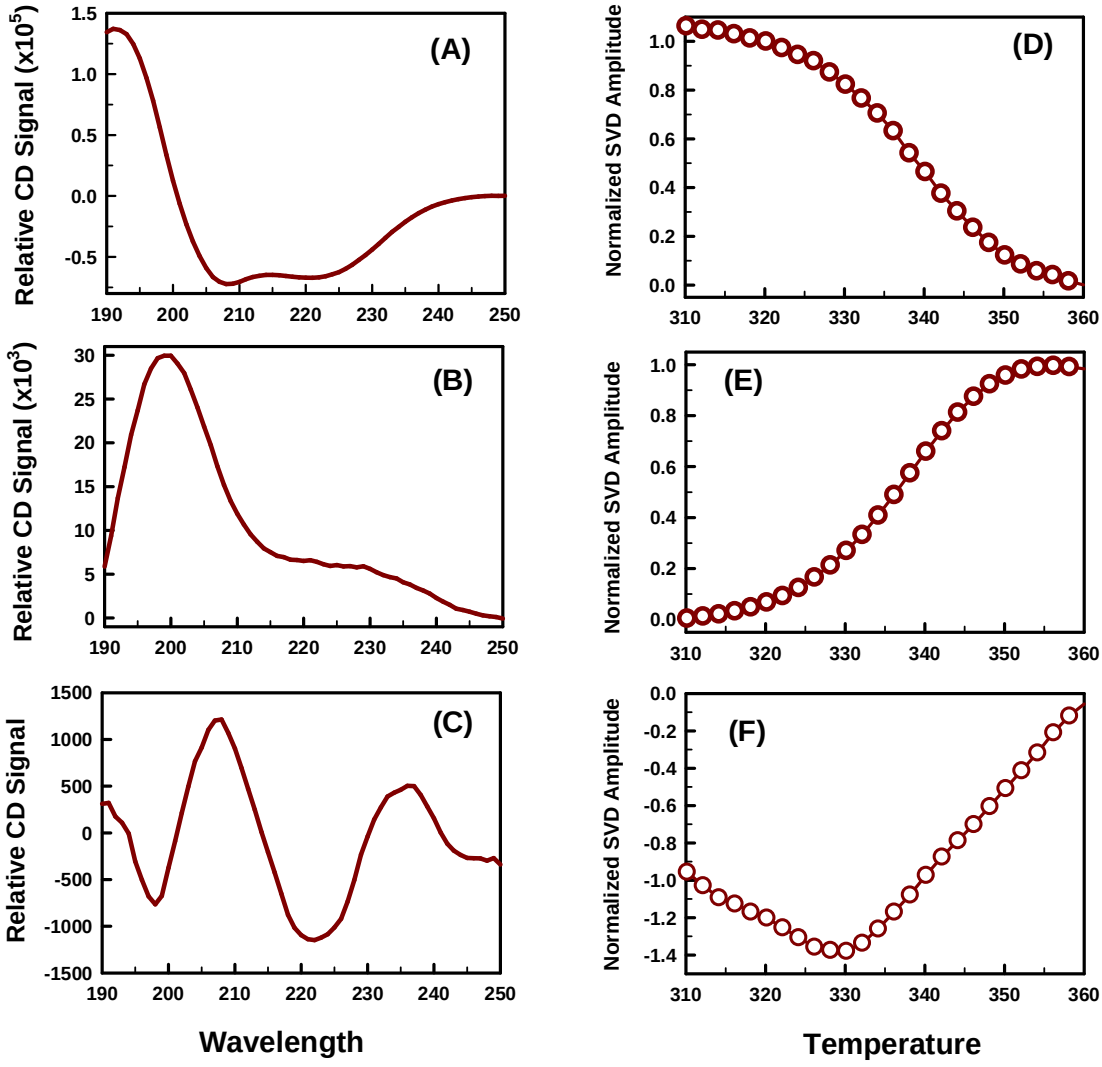

Figure 3.6 First, second, and third components (A, B, and C) from SVD analysis of f-CD temperature-wavelength melts of WT gpW with respective signal amplitude shown below (D, E, and F) for WT gpW.

3.3.2 Fourier Transform Infrared Spectroscopy (FTIR)

Fourier transform infrared spectroscopy is sensitive to bond stretching and bending vibrations inherent in chemical systems. The amide I region (1600 – 1700 cm^{-1}) of the FTIR spectrum is responsive to secondary structural transitions and serves as a compliment to f-CD measurements in the multiprobe analysis. This region particularly measures backbone carbonyl stretching vibrations with minor signals from C-N stretching and N-H bending. Of particular interest in analyzing gpW, are the signals at 1632 and 1652 cm^{-1} for a-helix and ≥1658 cm^{-1} for random coil structures. A view of the FTIR thermal melt absorbance spectra collected in ^{2}H sodium phosphate buffer pD 6.0 shows a large peak spanning 1620-1650 cm^{-1} (Figure 3.7A). With signal overlap common in FTIR, deconvolution of the overlapping absorbance bands would reveal a large positive band ~1650 and another, yet smaller band at 1632 cm^{-1} both corresponding to α-helical absorbance as expected. As temperature increases, the signal strength at ≥1658 cm–1 becomes increasingly apparent signaling the familiar helix-coil transition observed by f-CD. An extra band is observed at 1672 cm-1 and is caused by carboxylate stretching modes of trifluoro acetic acid (TFA) contamination present in the reversed phase mobile phase of which the purified sample was directly lyophilized. This absorbance band partially obscures a more detailed analysis of the unfolding signal. In panel B, the absorbance at 1632 cm^{-1} is plotted as a function of temperature to illustrate the loss of helical signal. The

shape of the curve is reminiscent of a sigmoidal curve; however, data collection at high temperatures is not possible precluding the ability to observe the post-transition baseline. As observed in f-CD, a sloping pre-transition baseline is present, suggestive of a non-two-state transition due to complex behavior in helix unfolding. This, of course, does not prevent a two-state fit to the curve where a T_m of 340 K is obtained in agreement with that calculated for f-CD.

Analysis of the FTIR data by SVD exposes three components of the total spectra. Figure 3.7C is the average spectrum that contributes approximately 91% to the total output. The corresponding amplitude shows an unfinished transition with a slightly shallower pre-transition baseline compared with the absorbance change at 1632 cm^{-1} (Panel B). Taking the first derivative of this plot reveals a T_m of 340, in agreement with f-CD. The second component (contributing ~7%) appears to represent the decomposition of the expected overlapping absorbance bands (panel E). Unlike the raw data (Panel A) where two overlapping bands can be imagined, this second component suggests a single absorption peak associated with α-helix structure and verifies the ability to monitor α-helical structure at 1632 cm^{-1}. Moreover, this single peak can be taken as evidence that the backbone carbonyl groups of the helices are predominantly solvent exposed. The third component, contributing just over 1% to the total, illustrates the opposing signals at

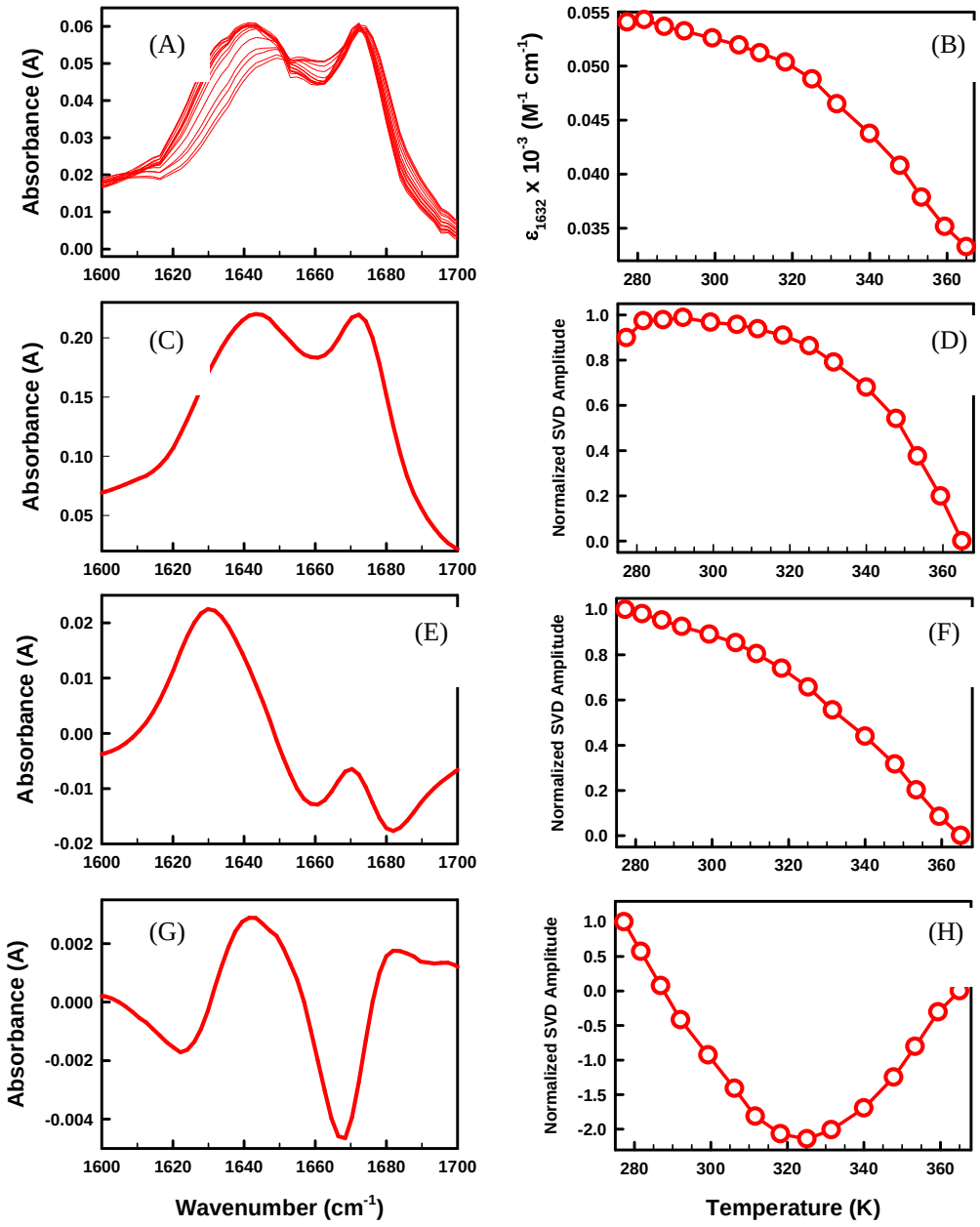

Figure 3.7 (A) Raw FTIR absorbance spectra (B) Temperature dependent unfolding of helical absorbance monitored at 1632 cm^{-1} (C, E, and G) First, second, and third SVD basis spectra and associated amplitudes (D, F, and H) respectively for WT gpW.

the expected wavelengths corresponding to backbone structure signifying the loss in helical composition and concomitant gain in coil-like structure (Panel G). The amplitude of this component is reminiscent of the third component of the f-CD affirming the complimentary agreement between methods. A minority fourth component is also observed which replicates the total average basis spectrum, but the peak at 1672 cm-1 (TFA) overshadows further analysis of the unfolded structure and isn't considered. The amplitude of this fourth component appears to perhaps illustrate the temperature independent behavior of this absorbance, supplying more evidence that this absorbance is primarily due to TFA.

3.4 Probing Global Structure

The presence of amino acids with conjugated electron systems (i.e. aromatic residues tryptophan, tyrosine, and phenylalanine) can be used to probe the solvent exposure surrounding the chromophore. This measurement of tertiary structure is routinely performed using near-UV circular dichroism (n-CD) and/or fluorescence spectroscopy. The single native tyrosine residue located at position 46 (Y46) in the middle of the second α-helix acts as a convenient probe for examining tertiary structure. A comparison to the behavior of the peptide backbone monitored via f-CD and FTIR permits a more detailed inspection of the overall unfolding mechanism of WT gpW.

3.4.1 Near-UV Circular Dichroism (n-CD)

The manifestation of tertiary structure in the folding / unfolding ensemble of gpW can be measured using n-CD by following the electronic absorptions of the lone tyrosine residue. The temperature-wavelength spectra in 20 mM sodium phosphate pH 6.0 are plotted in figure 3.8A as a succession of blue curves with the native conformation giving rise to the most positive signal with a maximum at 280 nm, consistent with tyrosine absorbance (270 – 290 nm). The lack of any positive signal at 280 nm at the high end of the temperature range suggests a complete lack of residual structure in the unfolded state. Figure 3.8B shows the mean residue ellipticity monitored at 280 nm over the entire range of temperatures. A clear sloping pre-transition baseline is observed with the first derivative of the transition indicating a T_m of 335 K, in agreement with that produced from a two state fit.

Applying the same SVD treatment to the n-CD data, there appear to be four components. The first is the average basis spectra (Figure 3.9A) comprising ~70% of the total with the associated amplitude showing a near duplicate curve to that of the unfolding monitored at 280 nm (Panel E). Interestingly, the second component (contributing ~27%) bears striking resemblance to the protein absorption spectrum with a maximum at 275 nm concurring with the extinction coefficient of tyrosine (Panel B). By visual inspection, the amplitude of the second component appears to follow the same pattern as the amplitude of the first (Panel F). However, taking the first derivative of

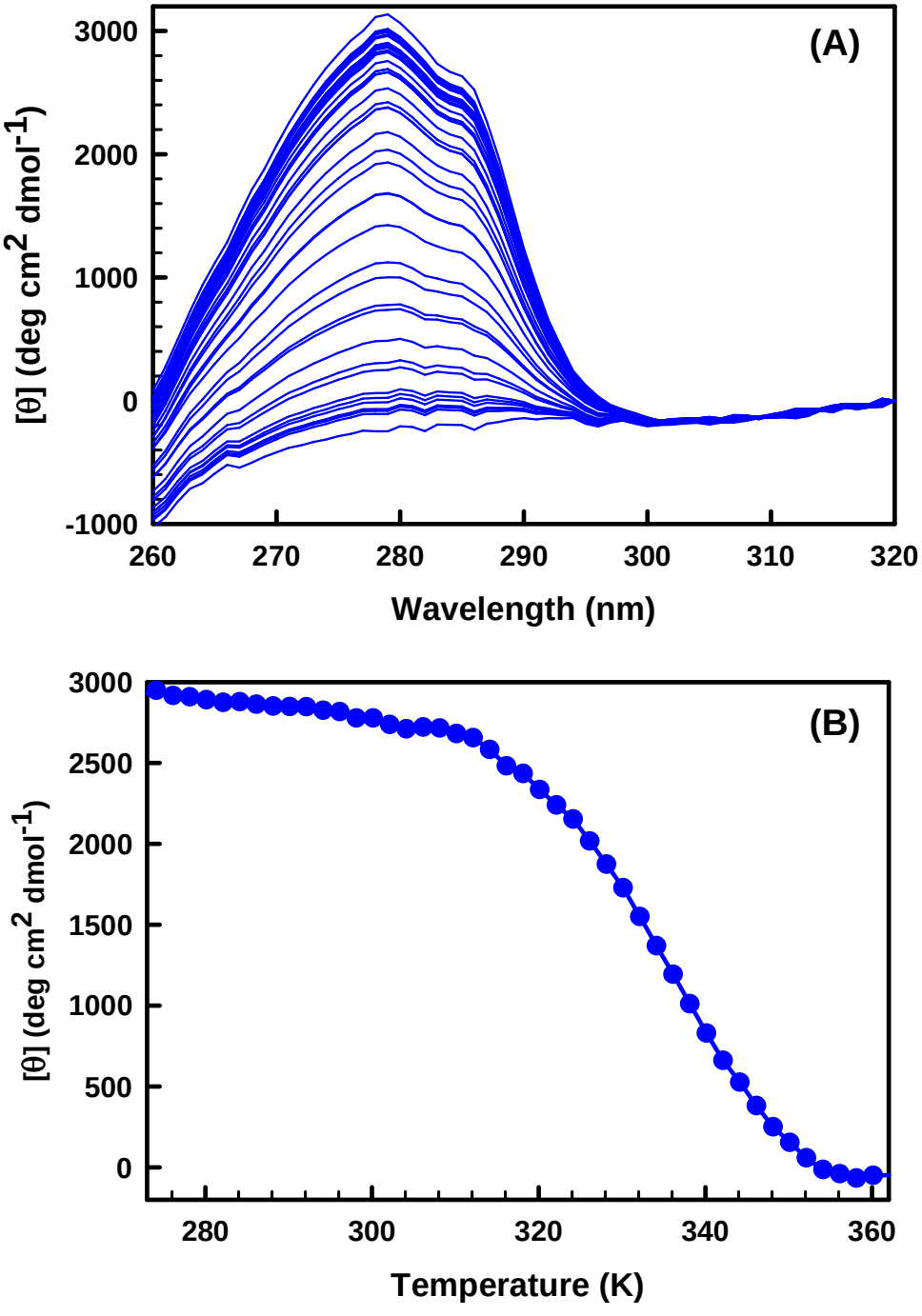

Figure 3.8 (A) WT gpW Temperature-wavelength n-CD spectra (B) Ellipticity monitored at 380 nm over the temperature range (273 – 362 K).

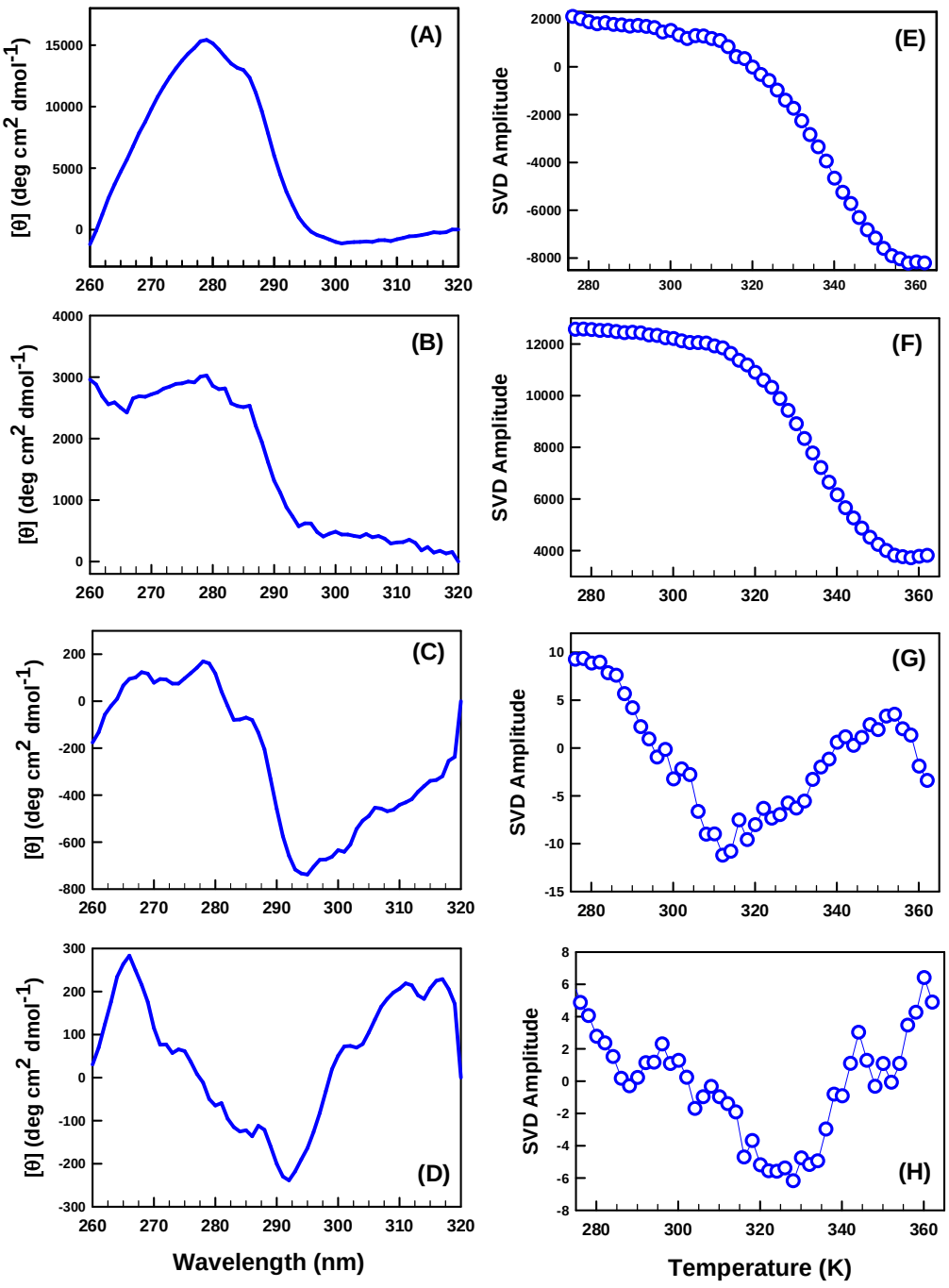

Figure 3.9 SVD components (Panels A, B, C, and D) of n-CD and corresponding amplitudes (Panels E, F, G, and H) for WT gpW.

each yields a T_m of 335 K for the first and 340 K for the second component. The Tm for the second component agrees with that calculated in f-CD and FTIR experiment suggesting that this tyrosine not only reports on the tertiary environment, but also the backbone structure. This is not surprising given the location of Y46 in the second helix, and therefore a third probe of the backbone structure is gained. In panel C and G, the third component is shown which appears to resemble the signal change of tyrosine resulting from thermal unfolding. This component contributes roughly 3% to the total basis. Component 4 looks very similar to that of the third component from f-CD contributing 1.3% to the total. However, the significance of this component is questionable as the first derivative reveals a very broad and noisy minimum. It is noted however that this minimum is very near 335 K as are the first component amplitude and signal at 280 nm. Therefore, the n-CD experiments give rise to an apparent T_m of 335 K which is representative of gpW tertiary structure, but also allows the assessment of the local environment surrounding the native tyrosine. This produces an apparent T_m of 340 K in agreement with f-CD and FTIR.

3.4.2 Fluorescence

The single native tyrosine residue functions as an intrinsic fluorophore allowing the indirect observation of tertiary structure via fluorescence. As a complement to the n-CD experiments, fluorescence experiments were performed at 6

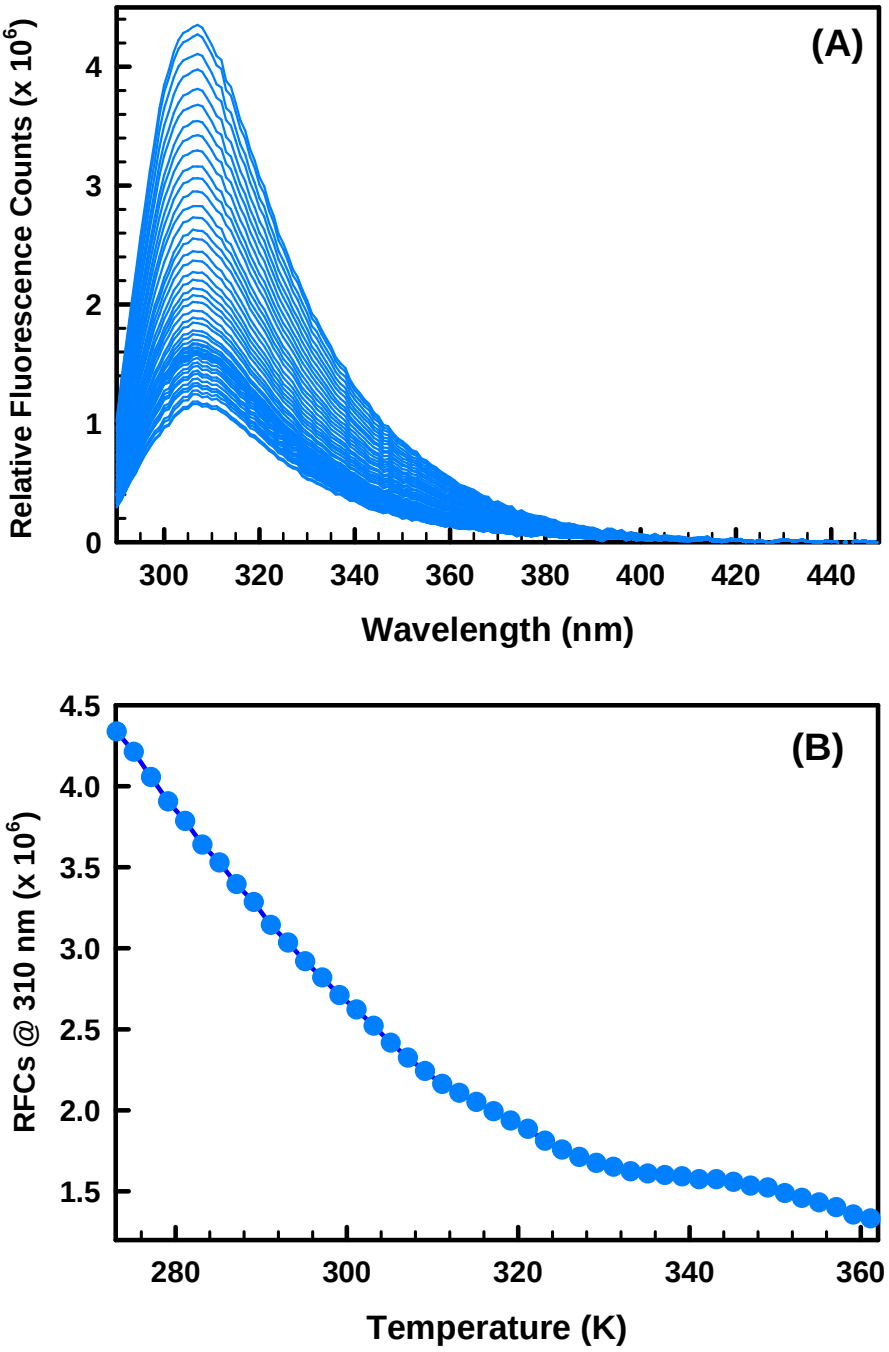

Figure 3.10 (A) Temperature dependent fluorescence spectra of WT gpW (B) Degradation of signal maximum observed at 310 nm

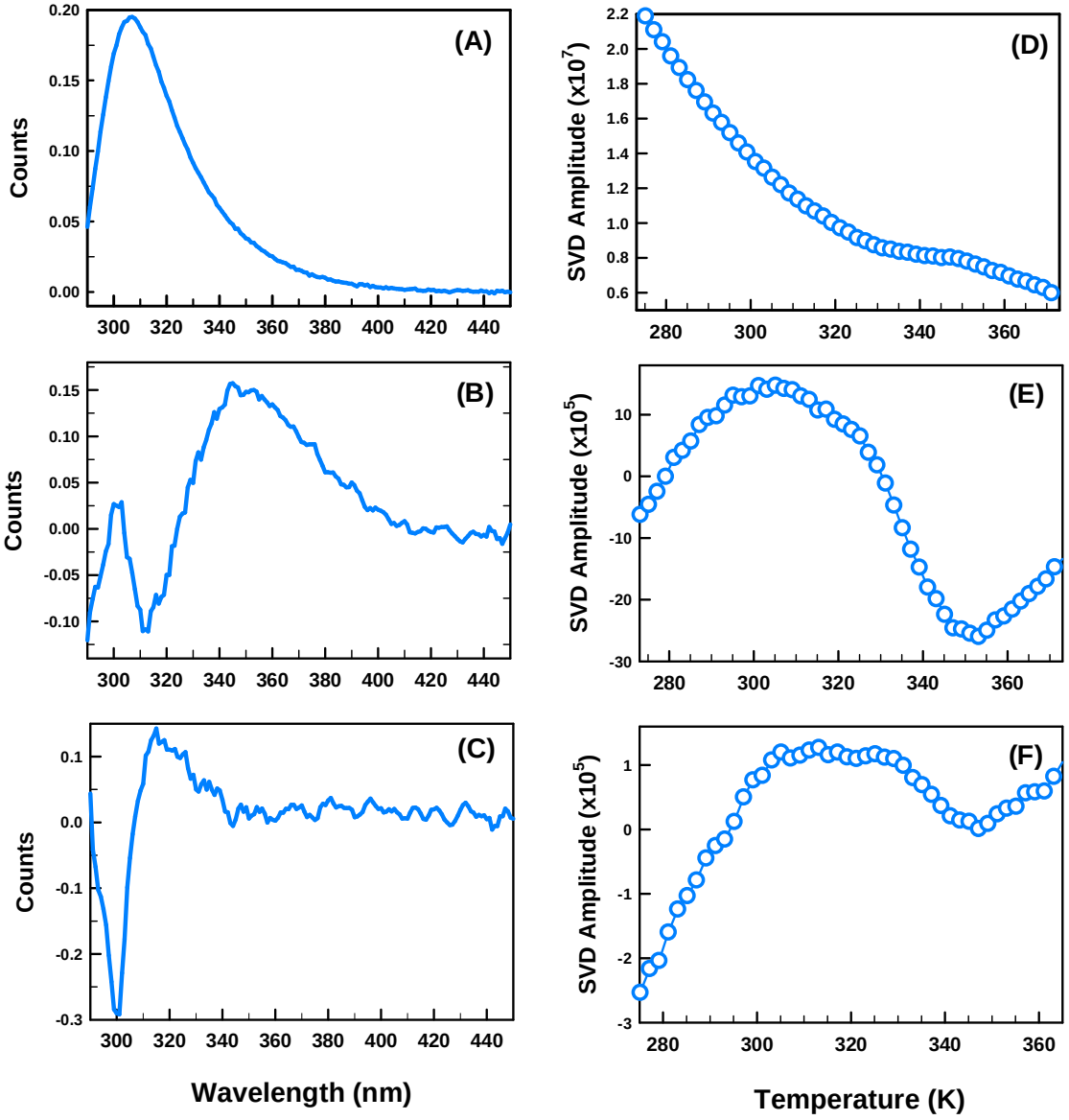

Figure 3.11 SVD components (A, B, and C) and corresponding amplitudes (D, E, and F) for WT gpW.

uM concentration in 20 mM NaP pH 6. The sensitivity of this technique is emphasized by the minimal protein concentration needed, which is largely due to the desolvation of Y46 and its apparent packing within the hydrophobic interior minimizing quenching from solvent. Since experiments were performed in a polar solvent (water), the fluorescence quantum yield of tyrosine will decrease as temperature increases signaling exposure to solvent and gpW unfolding. Figure 3.10 illustrates the gradual decrease in relative fluorescence counts (RFCs). In panel A, the temperature-wavelength spectra are presented with the greatest signal corresponding to the native state. Panel B shows the temperature dependent decrease in relative fluorescence monitored at the peak maxima, 310 nm. A nearly linear decrease is observed which is accounted for in the first component of the SVD analysis (Figure 3.11A and D). The second component reports more directly on the internal environment of the tyrosine residue evidenced by the opposite signs at 310 and 350 nm corresponding to the wavelength redshift as unfolding occurs (Panel B). The sharply positive sloping pre- and post-transition baselines of the second component amplitude are due to this sign change (Panel E). From a two-state fit and first derivative analysis of this amplitude, a T_m of 340 K is obtained in agreement with the first component of the n-CD experiment. The third component may be indicative of a temperature dependence of the emission wavelength.

3.5 Spectroscopic Characterization of Bi-Histidine Mutants

The multiprobe analysis by equilibrium spectroscopic techniques was repeated for each of the gpW bi-histidine mutants as was done for WT gpW to assess stability and equilibrium unfolding behavior. Each mutant was subjected to analysis by following the unfolding over the entire f-CD spectrum (250 – 190). In each case, there were observable wavelength dependencies at the characteristic wavelengths indicative of secondary structure (193, 200, 208, 222) (data not shown). This provides a good first indication that these mutants fold either downhill or over a marginal barrier like wild-type. Of particular note, protein concentration for each of these experiments had to be doubled (40 μM) to achieve more reasonable signal amplitude. The CD spectrum is affected by both secondary structure and protein concentration. Doubling concentration did not improve the signal-to-noise to WT levels, which indicates that the composition of secondary structure in each of the mutants is not equal to that of WT.

Panel A of figure 3.12 is a summary of the first derivative of the equilibrium unfolding curves monitored as measured by f-CD at 222 nm. Stability is significantly diminished in each of the helical mutants (blue, green, and red curves for E5H, R11H, and K45H respectively) with melting temperatures lowered up to 20 degrees as compared to wild-type (black curve). Beta-sheet mutants, V23H and Q27H, show less destabilization, as expected, due to less disruption of the backbone flexibility compared to helical mutants (cyan and magenta curves). Panel B shows the native

probability as a function of temperature from two-state fits of the f-CD unfolding at 222 nm.

Complimentary to f-CD, FTIR experiments were performed for each bi-his mutant. As seen in wild-type, agreement between the two probes of backbone structure was observed. Thermal melting temperatures were coincident for the two methods.

To probe the environment of the tyrosine residue, n-CD and fluorescence experiments were completed. As was seen from f-CD, the n-CD signal amplitude was lower, but protein concentration remained 80 μM for n-CD due to material limitations. The low amplitude is again indicative of diminished structure; however, it appears to be more dramatic for the helical mutants consistent with what was observed by f-CD. In fact, the E5H E6H A9H mutant has a negligible signal at all temperatures suggesting the tyrosine is exposed to the solvent even under native conditions. This is confirmed by fluorescence in which the temperature dependent emission spectra reflected the observed spectra for free tyrosine (data not shown). As with wild-type, a minor difference in T_m is seen for each mutant in support of the non-cooperative unfolding behavior seen in downhill folders.

Panels C and D of figure 3.12 are illustrative representations of the global environment of the lone tyrosine residue at position 46 as measured by n-CD and fluorescence. Not shown in these panels are the unfolding curves for the E5H mutant due to a lack of signal (color code as in panels A and B).

3.6 Conclusions

gpW represents a new member to the class of single domain proteins being subjected to protein folding studies. Composed of 68 residues (50 folded) comprising two α-helices packed against an anti-parallel two-stranded β-sheet, this protein is the smallest known monomeric fold to include both alpha and beta structure. A superficial examination of gpW suggests it follows a two-state mechanism for folding. Its size and contact order correlate with other single domain two-state folders which fold in the millisecond timescale.

A global two-state fit to an array of unfolding curves as a function of denaturant coupling reproduced the first-order perturbation, but failed to account for the higher order effect of temperature-GdnHCl coupling. This effect is highlighted by the comparison between the global fit and individual two-state fits. Plots of DH and DS obtained from the single fits showed a characteristic curvature mimicking the trend observed for downhill folders. This evidence serves as a first sign that gpW folds in a non-two-state manner.

A Multiprobe analysis of the equilibrium behavior allows a closer inspection of its folding activity. For wild-type gpW, f-CD and FTIR are used to probe the backbone conformation as a measure of secondary structure. N-CD and fluorescence are used to measure the tertiary structure with the former also providing an additional probe of

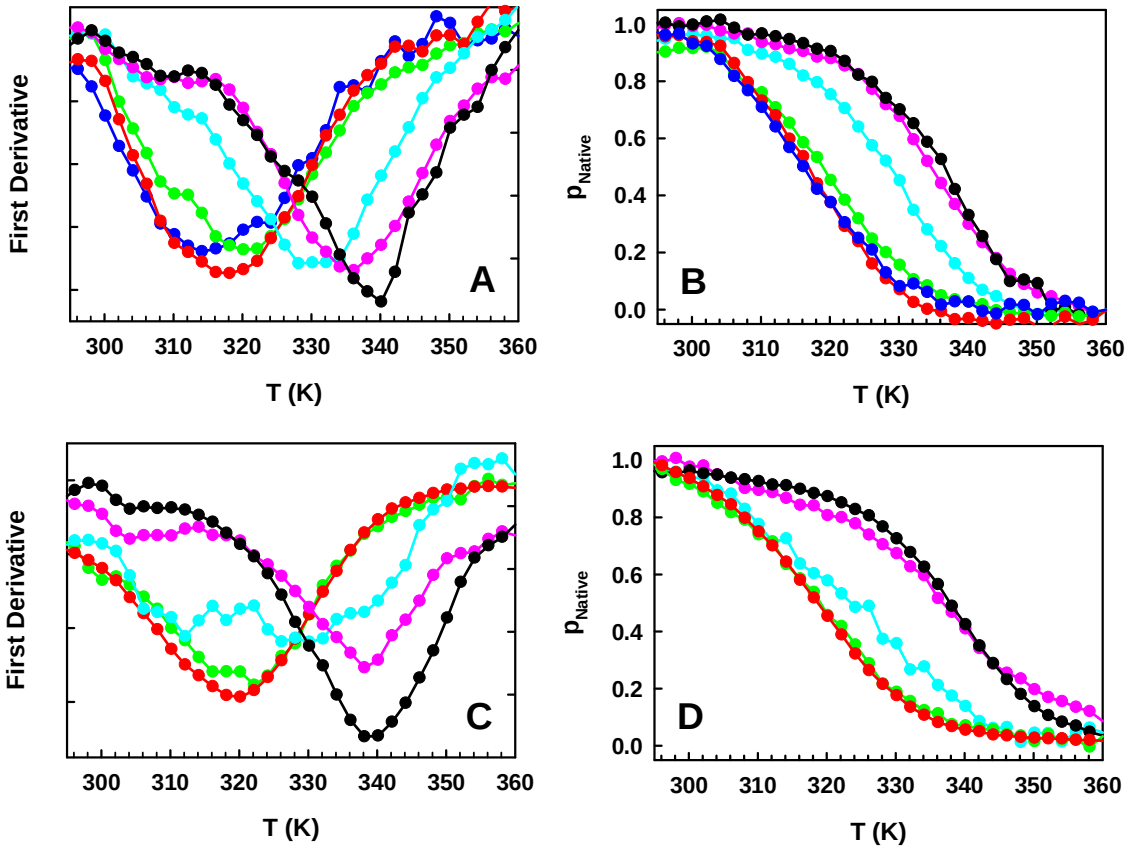

Figure 3.12 (A) Representative first derivatives of backbone unfolding curves and (B) native probabilities for: (blue) E5H; (green) R11H; (red) K45H; (cyan) V23H; (magenta) Q27H; (black) wild-type. Panels C and D are the same as A and B but represent tertiary structure.

the local environment surround the native tyrosine residue positioned in the middle of the second helix. From this multiprobe analysis, a clear trend can be distinguished as

a result of this all-encompassing view of the folding ensemble. Thermal stability data obtained by f-CD and FTIR were complimentary producing a T_m of 340 K from the first derivative analysis and two-state fits of the first component amplitude. An additional probe of the local environment was obtained from the second component of n-CD experiments, which also reveals a Tm at 340 K. Monitoring the global environment by n-CD and fluorescence yields a temperature midpoint of 335 K, a 5 K disparity compared to the secondary structural analysis. This is best illustrated by overlaying the first derivative of each method (Figure 3.13A). This analysis does not rely on constraining the data to a particular model and is a feature inherent in the amplitude independent of baselines. This observation is duplicated by fitting each complimentary experimental pair (f-CD and FTIR, n-CD and fluorescence) of equilibrium unfolding curves to a two-state model (Figure 3.13B). Both figures clearly show the 5 K mismatch in thermal stability between the secondary and tertiary structure. It can therefore be confirmed that the backbone and tertiary structure of WT gpW does not unfold in a cooperative manner. This complex probe-dependent behavior is consistent, though less dramatic, with unfolding of the global downhill folder BBL (Naganathan, A. N. 2005a).

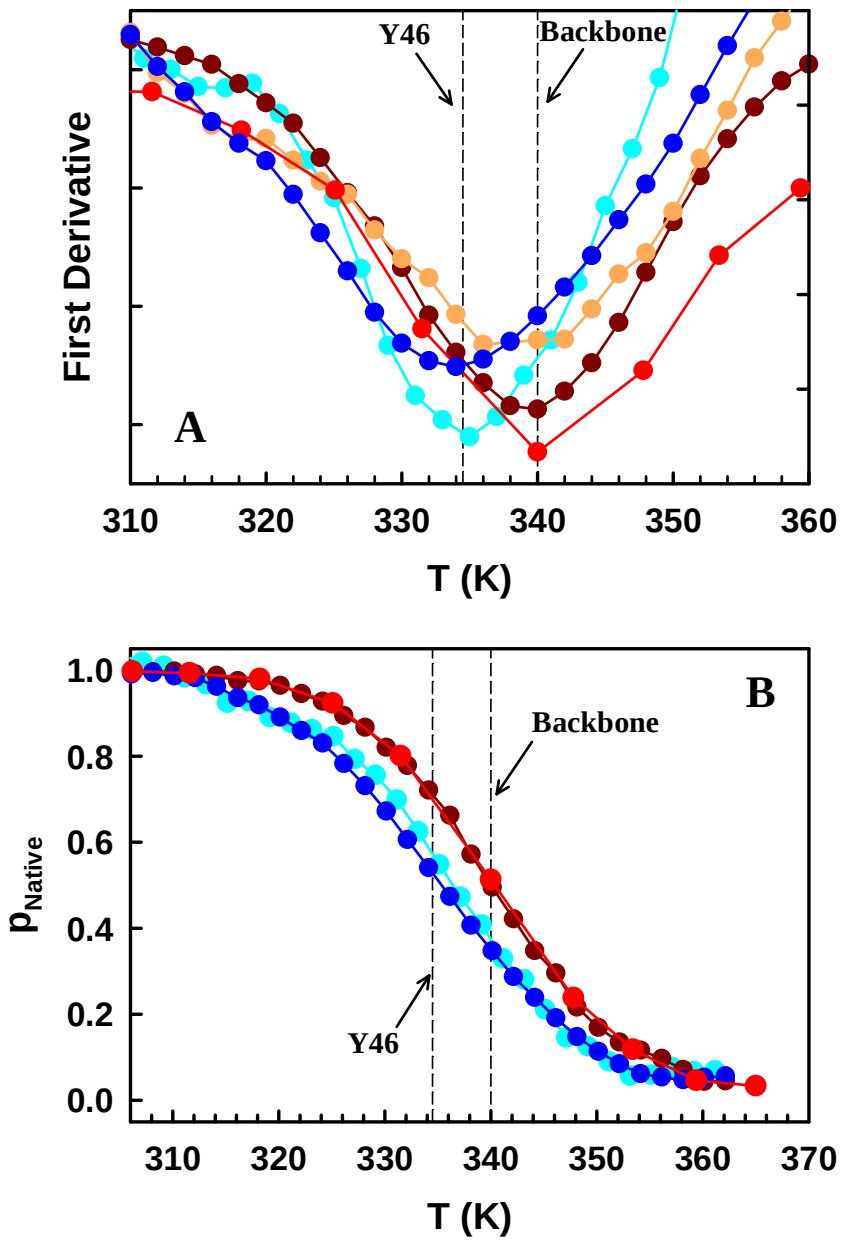

Figure 3.13 (A) First derivative and (B) native probability from two-state fits of equilibrium thermal unfolding curves for WT gpW: f-CD (dk red); FTIR (red); n-CD (dk blue); fluorescence (lt blue); amplitude of second component of n-CD (orange) (only in (A))

Bi-histidine mutants of gpW show similar behavior to the WT construct implying that these mutations do not affect the overall folding behavior. A clear decrease in both secondary and tertiary structure is observed for the helical mutants, and to a lesser extent in the beta-sheet mutants. These results are consistent to some extent due to the introduction of the destabilizing amino acid glycine between the two histidine residues in the helix. In addition to the added conformational flexibility, electrostatics are disrupted with histidine residues taking the place of several polar negatively charged residues. The E5H mutant shows no tertiary structure, which points to the N-terminus of helix one being a prominent feature in global stability. This is confirmed upon inspection of the contact map where long range contacts are present between the N-terminus of $\alpha 1$ and the β-hairpin.

The results from equilibrium thermal unfolding and denaturant coupling suggests gpW and each of the bi-his mutants folds in the downhill regime. The limited span in thermal stability between local and global probes does not provide indisputable evidence of the height of the barrier. The experimental evidence does however support the notion that the barrier height falls within the 3 RT limit for folding over a marginal barrier (Gruebele, M. 2008). A positive barrier height therefore makes observing the population at the top of the barrier possible (Munoz, V. 2007); the experimental signatures of which are the probe-dependent features discussed in this chapter.

The lack of structural detail as a result of ensemble techniques does not allow a quantitative measure of the free energy barrier. However, by employing DSC and kinetics experiments, this free energy barrier to folding can be directly measured and a definitive classification of downhill, marginal, or two-state can be achieved.

4. Measurement of the Folding Barrier Height by Calorimetry

4.1 Introduction

The heat capacity (C_p), or thermal energy storage capacity, of a solution can be measured using differential scanning calorimetry. In the case of a protein solution, the instrument measures the difference in heat capacity between the solute (protein) and solvent (buffer). If the heat capacity of a protein solution is high relative to the bulk solvent, a large amount of energy will be stored as temperature rises. With respect to protein folding, this energy storage process parallels the folding / unfolding transition providing a direct description of the variations in energy at equilibrium over the relevant temperature scale. Therefore, as conformational entropy is gained upon protein unfolding, C_p increases resulting in a positive peak in a DSC thermogram. Thus, DSC provides direct access to the statistical ensemble, or partition function, for the folding / unfolding equilibrium.

4.2 DSC Data Described by a Two-State Approximation

Consider a two-state system consisting of two macrostates, folded (F) and unfolded (U). As described in chapter 1, the equilibrium can be described by:

$$F \underset{k_u}{\overset{k_f}{\rightleftharpoons}} U \tag{4.1}$$

where k_f and k_u are the rate constants for the folded and unfolded states respectively. This equilibrium expression can be more thoughtfully described as a population distribution between macrostates (Lumry, R, 1966). These macrostates are populated by an ensemble of microstates, where in a two-state system there is very low occupancy of states between the two ground states.

The equilibrium constant (equation 4.2), partition function (equation 4.3), folded and unfolded probabilities (equation 4.4), and Gibbs free energy (equation 4.5) (from chapter 3) are described by:

$$K(T) = \frac{[U]}{[F]} = \frac{k_u}{k_f} = e^{\left(\frac{-\Delta G(T)}{RT}\right)} \tag{4.2}$$

$$Q = 1 + K(T) \tag{4.3}$$

$$p_f = \frac{1}{1+K(T)} \text{ and } p_u = \frac{K(T)}{1+K(T)} \tag{4.4}$$

$$\Delta G(T) = \Delta H_m + \Delta C_p (T - T_m) - T\left[\Delta S_m + \Delta C_p \ln\left(\frac{T}{T_m}\right)\right] \tag{4.5}$$

$$\Delta C_p = (C_p^u - C_p^f) \tag{4.6}$$

The calorimetric ΔC_p is defined as the difference in the heat capacity between the folded and unfolded states (equation 4.6) and is framed by the folded and unfolded baselines (area under the curve). A positive ΔC_p originates from an increase in solvent accessible surface area exposed at increasing temperatures. This can be due to an all-or-nothing exposure of hydrophobic residues as in a two-state scenario, or structural plasticity in the case of downhill folders.

4.2.1 Two-State Criterion

The parameter ΔH_m is also known as the van't Hoff enthalpy. Taken at the midpoint temperature, it is obvious that this parameter reports on the probability distribution of the folded and unfolded states. Therefore, a ratio of the empirical calorimetric enthalpy to the model-dependent van't Hoff enthalpy $(\Delta H_{cal} / \Delta H_{vH})$ is a measure of "two-stateness" with a ratio of 1 regarded as a historical marker of a two-state system. Inherent in this definition, the van't Hoff enthalpy is a measure of peak symmetry reflected in the shape of the DSC thermogram. As a result, the temperature dependent enthalpy is reported by the sharpness of the transition. For a high unfolding enthalpy, the shape of the peak will be narrow while a low unfolding enthalpy specifies a weaker temperature dependence resulting in a more broad transition. With the increasing focus on the use small proteins to study the details in protein folding, the assumed two-state enthalpic criterion has come under scrutiny. These small proteins give characteristically broad peaks as an effect of the low

unfolding enthalpy raising an issue related to defining the pre- and post-transition baselines. Since this is a user-defined parameter, a significant amount of error can be introduced in the analysis making the comparison of DSC data between groups difficult (Zhou, Y. 1999, Kaya, H. 2000). As a consequence, greater inspection of the baselines is necessary.

4.2.2 Importance of Baseline Determination

Specifically, the pre- and post-transition baselines correspond to heat capacity variations in the structural ensemble at low and high temperatures. Freire has shown that the baseline at low temperature scales with the number of residues within a protein (Gomez, J. 1995). Therefore, the slope of this "native" baseline leads to a valuable conclusion whereby any slope greater than the predicted baseline suggests a non-two-state scenario. A calorimetric experiment with greater attention paid to pre- and post-transition baselines of several proteins has revealed a hallmark signal indicative of non-two-state behavior (Garcia-Mira, M. M. 2002, Naganathan, A. N. 2005, Naganathan, A. N. 2005b, Irun, M. P. 2001, Dragan, A. I. 2002). Specifically, these proteins were found to have baselines crossing within the transition region of the DSC thermogram. The origin of this effect cannot be rationalized with a physical explanation as this involves a change in sign of the apparent heat capacity at the temperature of intersection. Further, this effect has come to be a defining quality of

global downhill (BBL and PDD (unpublished results)) folders and those that proceed over a marginal barrier (gpW).

4.3 Variable-Barrier Model

With its unique access to the canonical ensemble, DSC provides an impressive tool to inspect free energy barriers in protein folding. Proteins fold by the addition of successive stabilization energies. Compared to chemical reactions which involve the making and breaking of covalent bonds, the sum of the weak interactions formed as a protein folds are 2-3 orders of magnitude smaller in energy. Until recently, the absence of a method to quantify the free energy of folding in equilibrium prevented an absolute measure of the barrier height outside of experimental signatures (Chapter 3). The variable-barrier (VB) model developed by Muñoz and Sanchez-Ruiz (Muñoz, V. 2004) solved this problem and revolutionized the ability to differentiate between two-state, marginal barrier, and global downhill folding.

The VB model gets its origin from the Landau theory of critical transitions (Callen, H. B. 1985). It was developed to alleviate the model specific constraint routinely imposed on experimental data. Instead, the VB model obtains the thermodynamic parameters and the height of the free energy barrier directly from the experimental data without assumptions. Given calorimetric considerations, the heat capacity of the system is directly related to the statistical distribution of enthalpic microstates. Therefore, the partition function can be expressed by equation 4.7 which

allows a continuous treatment of the statistical distribution as opposed to ignoring microstates in a two-state approximation. Therefore, microstate density is dependent on conformational entropy with temperature and heat capacity is the derivative of the temperature dependent average enthalpy.

$$Q = \int \rho(H) e^{\left(-\frac{H}{RT}\right)} dH \tag{4.7}$$

Here, $\rho(H)$ is the microstate density and H is the enthalpy, both of which are temperature independent and fixed over the range of possible populations. The probability, $P(H \mid T)$ of a protein possessing a certain enthalpic quantity, H, is (equation 4.8):

$$P(H \mid T) = \frac{1}{Q} \rho(H) \cdot e^{\left(-\frac{H}{RT}\right)} \tag{4.8}$$

Equation 4.9 can be used to calculate the probability density at a given experimental temperature (T) referenced to temperature T_o:

$$P(H \mid T) = C \cdot P(H \mid T_o) \cdot e^{(-\lambda H)} \tag{4.9}$$

where C is a normalization constant and λ is (equation 4.10):

$$\lambda = \frac{1}{R}\left\{\frac{1}{T} - \frac{1}{T_o}\right\} \tag{4.10}$$

First, and higher-order enthalpy moments ($\langle H^n \rangle$ with n = 1, 2, . . .) can be expressed as:

$$\langle H^n \rangle = \int H^n P(H \mid T) \cdot dH = C \cdot \int H^n P(H \mid T_o) \cdot e^{(-\lambda H)} dH \tag{4.11}$$

And the excess heat capacity referenced to the native state as:

$$C_p^{EX} = \frac{d\langle H \rangle}{dT} = \frac{\langle H^2 \rangle - \langle H \rangle^2}{RT^2} \tag{4.12}$$

If the probability density at T_o is known, the temperature dependent excess heat capacity can be immediately computed from equations 4.9 – 4.12. An inverse Laplace transform of the partition function could also be used to calculate the probability density, but associated assumptions in this approach result in ambiguous results. To alleviate this problem, the probability density can be defined as a free energy function with physical significance and minimal parameters (equation 4.13).

$$P(H \mid T_o) = C' \cdot e\left(-\frac{G_o(H)}{RT_o} \right) \tag{4.13}$$

where $G_o(H)$ is the free energy function at the reference temperature and C' is a normalization constant. As in the Landau theory of phase transitions, $G_o(H)$ is expressed as a four-term power series and that the odd powers are zero at T_o (equation 4.14).

$$G_o(H) = -2\beta \cdot \left(\frac{H}{\alpha}\right)^2 + |\beta| \cdot \left(\frac{H}{\alpha}\right)^4 \tag{4.14}$$

where the meaning of β and α described below. The parameter β can be positive or negative, but to ensure stability of the model due to potentially large enthalpy changes, the coefficient of the quartic term is positive, hence the absolute value of β. By setting the derivative of the free energy functional with respect to enthalpy to zero ($dG_o(H)/dH)=0$) and evaluating d^2Go/dT^2 for β > 0, *Go(H)* shows two minima at +α and –α and a maximum at *H=0*. Thus, for β > 0, there are two macrostates corresponding to the folded (+α) and unfolded (-α) states separated by a free energy barrier β. For β < 0, only one minima is observed at *H=0* signaling a single macrostate. A drawback of the model description thus far is that the free energy functional assumes equal heat capacity for both minima (folded and unfolded states). However, the resultant symmetry can be redefined by introducing an additional α parameter for each sign of H (α_N for H > 0 and α_P for H < 0) and express these parameters in equations 4.15 – 4.17 for ease of fitting.

$$\alpha_N + \alpha_P = \Sigma\alpha \tag{4.15}$$

$$\alpha_N = \Sigma\alpha \cdot \frac{f}{2} \tag{4.16}$$

$$\alpha_P = \Sigma\alpha \cdot \frac{(2-f)}{2} \tag{4.17}$$

$\Sigma\alpha$ approximates the enthalpy difference between the two macrostates in a two-state scenario. f is regarded as an "asymmetry factor" where for a two-state system, $f = 1$ defines an equal population density with identical peak shape for folded and unfolded states, while $f < 1$ leads to a biased distribution with a sharper (narrower range of H) peak for the folded state. Therefore, only 4 parameters are needed to fit DSC data to the VB model (β (barrier height), T_o (reference temperature), $\Sigma\alpha$ (enthalpy at T_o), and f (asymmetry factor). The fitting procedure utilizes an exhaustive routine of fixing values for β and setting a constant $f = 0.1$ to account for the ruggedness of the free energy landscape and ensuring that the free energy minima are global, not local.

The final component to the VB model before fitting can proceed is subtraction of the native baseline. This baseline accounts for the intrinsic heat capacity of proteins focusing the fit exclusively on the transition and enabling the accurate prediction of the barrier height. For proteins with a large free energy barrier and T_m clearly within the temperature range of the instrument, the calorimetric peak is sharp with well-defined baselines. Therefore, the native baseline can be defined be taking the heat capacity values from the pre-transition baseline and directly imposed as the native baseline. For many of the smaller proteins however, the calorimetric peak is broad with poorly resolved baselines. Extracting the native baseline from the pre-transition heat capacity values therefore results in trimming of the data and is dependent on the experimentalist's judgment in extrapolation. This is often the tactic employed to prejudice the data to fit a two-state model (Ferguson, N. 2004). However, this can be overcome by utilizing Freire's empirical baseline from which

the temperature dependent heat capacity is a direct function of the molecular weight of the protein (equation 4.18) (Gomez, J. 1995, Freire, E. 1995).

$$C_p^F = [1.323 + 6.7x10^3 \cdot (T - 273.15)] \cdot M_r \quad J \cdot K^{-1} \cdot mol^{-1} \qquad (4.18)$$

where T is the temperature in K and M_r is the molecular weight of the protein in $g \cdot mol^{-1}$. Essential to this analysis is the use of absolute heat capacity. The fit then only relies on 3 parameters, f, $\Sigma\alpha$, and T_o with the native baseline obtained from well-defined pre-transition C_p values or Freire's empirical baseline fixed. Errors in the fit are generated from the grid-search of fixed β values and compared to the standard deviation (σ). Plotting the barrier height versus the standard deviation allows a visual assessment of the events associated with the unfolding transition (i.e. two-state or barrierless) and an unbiased classification of the folding behavior (Muñoz, V. 2004).

4.4 Direct Estimate of gpW Free Energy Barriers

DSC has been carried out on WT gpW and the barrier heights have been extracted by applying the variable-barrier model to the calorimetric profile. The DSC thermogram for WT gpW is shown in figure 4.1 (blue open circles) with units of absolute heat capacity as is required by the VB model. The thermogram displays a well-resolved peak with a T_m of ~340 K by visual inspection. However, the transition peak is uncharacteristically broad for a protein of this size and high thermal stability.

This feature is neatly illustrated by overlaying the DSC profile of the two-state α-spectrin SH3 domain (Viguera, A. R, 1994) which is composed of 62 residues and characterized by a similar T_m (green curve in figure 4.1). As discussed earlier, the slope of the "native" baseline for WT gpW has a characteristic steepness which exceeds that of Freire's empirical baseline (black open circles in figure 4.1). In fact, a simple two-state fit of the DSC thermogram for WT gpW produces "native" and

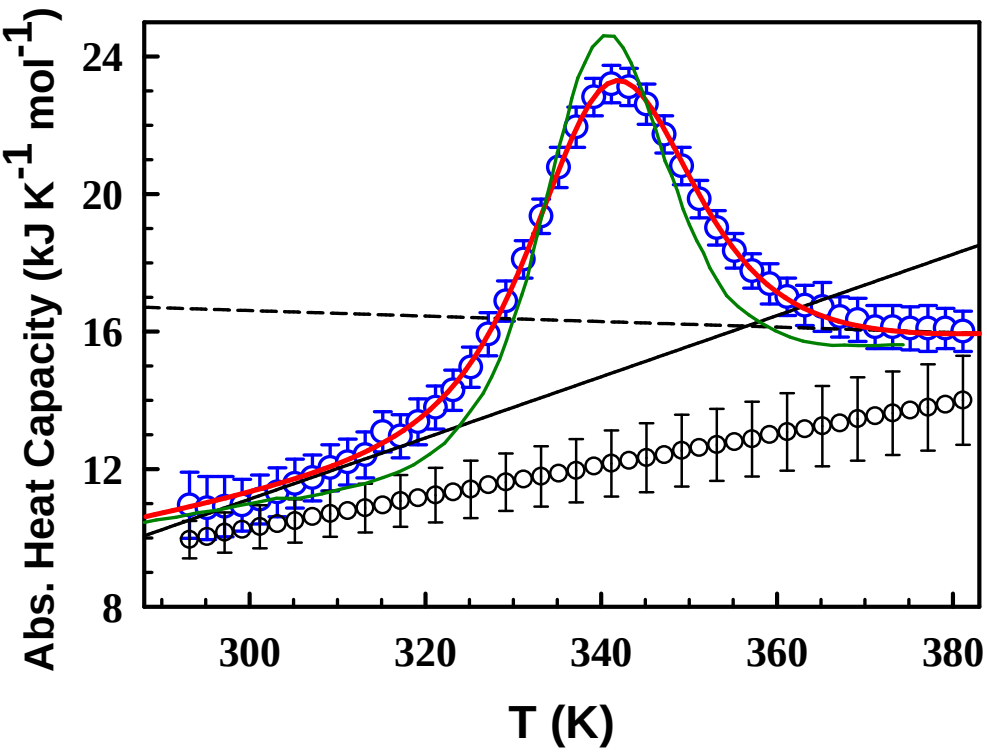

Figure 4.1 DSC thermogram of WT gpW (blue open circles) with error bars showing the standard deviation of 6 experiments performed at different protein concentrations;

native baseline from Freire's empirical correlation (black open circles); best fit to the variable-barrier model (red curve); native (solid black line) and unfolded (dashed black line) baselines from two-state fit; DSC thermogram of a-spectrin SH3 (green curve) from Viguera, A.R. 1994).

"unfolded" baselines that cross in the transition region (black lines in figure 4.1). The combination of these observations is the first calorimetric evidence that gpW possesses only a marginal (≤3RT) thermodynamic barrier to folding.

Applying the VB model, conclusive evidence supports these qualitative observations. The model reproduces the experimental data with outstanding precision using only 4 parameters as discussed above (red curve in figure 4.1). The resultant one-dimensional free energy profile using enthalpy as order parameter illustrates the remarkably low free energy barrier, where the full scale of the free energy functional ($G_o(H)$) on the y-axis is thermal energy (1 RT) (Figure 4.2). The barrier height (β) is calculated to be 0.51 kJ/mol, which is well within the 3RT limit for a marginal barrier. To establish a confidence limit of the fit, the model was re-run at one standard deviation above or below Freire's baseline (error bars of black open circles). The result of down- and up-shifting the baseline was a poorer fit and barrier heights of 0.0, and 2.7 kJ/mol respectively. These results are shown for comparison in Table 4.1. Since the barrier is so small, the probability distribution shows an expected distribution that does not agree with a bimodal dispersion (Figure 4.3). The

distribution moves from the folded ensemble with increasing temperature developing a high enthalpy tail which turns into a hump at the characteristic temperature (T_o), and finally moves toward the unfolded basin with progressively decreasing structure.

4.5 Conclusions

The variable-barrier model of experimental DSC data acts as a stand-alone method in distinguishing the energetic continuity of the protein folding process. It does so by directly measuring the free energy barrier height from raw experimental data without assumptions or constraints of the thermodynamic parameters to a chemical model. In other words, DSC monitors the global unfolding events independent of probe-dependence. The multiprobe analysis of marginal barrier folding proteins has the potential to show coincident unfolding behavior even when monitoring different structural features due to the possible presence of an as-yet-detected probe.

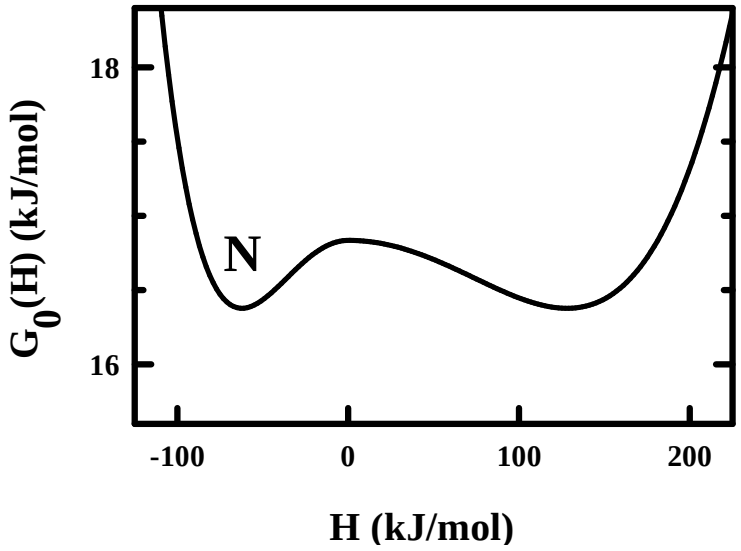

Figure 4.2 One-dimensional free energy surface with enthalpy as order parameter for WT gpW. Scaling of the free energy functional is 3RT (thermal energy).

gpW-WT	T_o	f	$\Sigma\alpha$	β	σ
C_p	68.11	0.5607	192.67	0.51	0.101
$C_{p\text{-(BL-SD)}}$	66.77	0.5171	135.16	0.07	0.386
$C_{p\text{-(BL+SD)}}$	70.25	0.5524	221.87	2.69	0.217

Table 4.1 Summary of parameters from fits to the variable-barrier model from WT gpW DSC data (6 experiments at varying concentration) (T_o (characteristic temperature) in K, f, (asymmetry factor), $\Sigma\alpha$ (enthalpy at T_o) in kJ/mol, β (barrier height in kJ/mol), and σ (standard deviation). Confidence of the fit was performed by subtracting Freire's empirical baseline by subtracting ±1 standard deviation from the

native baseline. Two-state fit yielded T_m=341.23 K (68.08 C) and ΔH=172.61 kJ/mol.

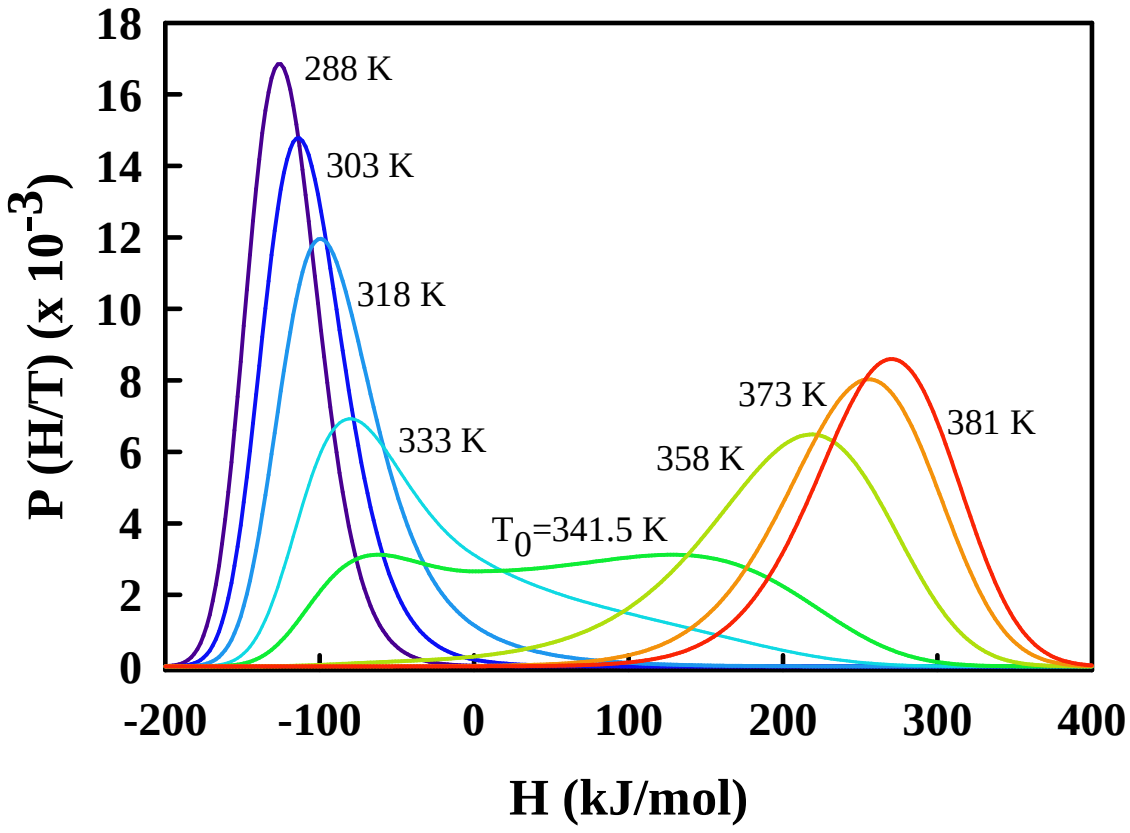

Figure 4.3 Probability distribution of WT gpW as a function of the order parameter, enthalpy. The low enthalpy state (native) at low temperature shifts in a continuous fashion to the high enthalpy state (unfolded) at high temperature.

Therefore, this makes DSC a vital approach in addition to the multiprobe analysis to unequivocally determine the folding process.WT gpW exhibits all of the qualitative as well and quantitative signals for a protein with a marginal folding barrier. Specifically, these criteria are: (A) the transition is broad spanning a large range of temperatures; (B) the baselines from a two-state fit cross in the transition region; (C) two-state folded baseline has a steeper slope than the empirically determined native baseline; (D) the barrier height extracted from the VB model is less than 3 RT.

5. Kinetic Characterization of gpW

5.1 Introduction

Protein folding rates occur across several orders of magnitude (i.e. microseconds to hours). The existence of many small globular proteins that following simple kinetic profiles in the millisecond range (Jackson, S. E. 1998, Akmal, A. 2004 Proteins) to the ever growing emergence of those that fold on the submillisecond timescale (Huang, G. S. 1995, Naganathan, A. N. 2005, Muñoz, V. 2007), begs an important question: Why do small proteins fold so fast? An example of engineered faster folding counterparts of a natural protein (Yang, W. Y. 2003, Yang, W. 2004, Ma, H. 2005) approaching the diffusion controlled speed limit (Kubelka, J. 2004) further justifies this question. The answer may be as straightforward as biology has simply selected for this behavior.

Energy landscape theory predicts several folding trajectories that all reach the final native ensemble. It diminishes the idea of a step-by-step pathway of well-defined structural assembly, but doesn't eliminate the possibility of lower energy options for a protein to reach its native state which would correspondingly bias the kinetic flux. The implication of this idea is that proteins may indeed fold without crossing a free energy barrier. The discovery of downhill folding (Garcia-Mira, M.

2002) successfully connects theory with experiment and has opened up a new realm of identifying folding behavior from both a thermodynamic and kinetic perspective.

Historically, the assumption is that natural single domain proteins are intrinsically predisposed to two-state folding where high energy intermediates are not amendable to experimental observation. The appearances of single exponential relaxation rates are typically thought of as a landmark for two-state-like behavior. Equal sensitivity to chemical denaturants (i.e. *m* value) between equilibrium and kinetics should also be observed for a two-state system. Typical two-state kinetic experiments are performed using rapid-mixing techniques by observing changes in the relaxation rate at varying concentrations of chemical denaturant (i.e. urea or GdnHCl) where a plot of the observed relaxation rate versus chemical denaturant shows a characteristic chevron shape.

A growing list of ultrafast folding proteins has sparked an increasing popularity in temperature-jump (T-jump) techniques to resolve the ultrafast folding kinetics. The observed relaxation rates of the downhill folders can be reasonably fit to a single exponential suggesting that this is not a sufficient criterion for two-state-like behavior. Global downhill folders are more precisely fit to a stretched single exponential while those with marginal barriers are better fit to a bi-exponential. Further, T-jump experiments at different chemical denaturant concentrations mimic the chevron-like shape observed for millisecond folding two-state proteins. However, the slopes are more shallow the faster the relaxation rate at its chemical midpoint and

can show curvature at low denaturant concentrations (Naganathan, A.N. 2006), which has previously been attributed to the presence of an on-pathway intermediate possessing a different rate dependence on denaturant concentration (Gianni, S. 2003). Another feature of proteins with marginal barriers is the temperature dependence of the relaxation rate. At low temperature, this dependence is weak, but as temperature is increased, a change in the slope resembling a shallow 'L' is observed in comparison with the more pronounced opposing dependency near the T_m for two-state folders (Naganathan, A. N. 2007). These observations, in addition to establishing that not only do protein folding rates correlate with size, but that these rates are directly related to the free energy barrier height (Naganathan, A.N. 2005) confers an immense implication. Size scaling and folding rates therefore suggest that most, if not all, small proteins folding in the microsecond timescale are inherently in the downhill regime under native conditions. To address the issue of experimental kinetic observations and the common two-state treatment of the data, a one-dimensional free energy surface model has been developed as a variant of Zwanzig's one-dimensional statistical-mechanical model (Zwanzig, R. 1995).

5.2 One-Dimensional Free Energy Surface Kinetic Model

Zwanzig's model for protein folding is one-dimensional with 'correctness' as the order parameter. The number of incorrect residues (S) is defined in reference to the native state where any residue can be in the correct or incorrect conformation and

is directly linked to the sampling of conformational entropy. The one-dimensional free energy surface (1-D FES) model (Naganathan, A. N. 2007) substitutes Zwanzig's reaction coordinate (S) with the parameter nativeness (*n*), which is defined as the average probability of finding any residue in the native conformation. This is a manipulation of Zwanzig's (*N*-*S*)/*N* parameter allowing for a continuous mode of distribution along the reaction coordinate. Therefore, *n* is intuitively a probability allowing conformational entropy ($\Delta S^{conf}(n)$) to be calculated from the Gibbs entropy formula (equation 5.1).

$$\Delta S_{res}^{conf}(n) = -R[n\ln(n) + (1-n)\ln(1-n)] + n\Delta S_{res}^{n=1} + (1-n)\Delta S_{res}^{n=0} \text{ for } 0 < n < 1 \quad (5.1)$$

with

$$\Delta S_{res}^{conf}(0) = \Delta S_{res}^{n=0} = S_{res}^{n=0} - S_{res}^{n=0}$$

$$\Delta S_{res}^{conf}(1) = \Delta S_{res}^{n=1} = 0 \quad (5.2)$$

$$\Delta S^{conf}(n) = N\Delta S_{res}^{conf}(n)$$

where $\Delta S_{res}^{n=0}$ is the variance between the conformational entropy of all possible non-native and native conformations of a particular residue. The folding stabilization energy ($\Delta H^o(n)$) is an exponential function (equation 5.3) where H_{res}^o is the stabilization energy per residue.

$$\Delta H^o(n) = N\Delta H_{res}^o \frac{1+e^{(k_{\Delta H}n)-1}}{1-e^{k\Delta H}} \quad (5.3)$$

and the one-dimensional free energy surface is obtained by adapting to the familiar Gibb's free energy equation (equation 5.4).

$$\Delta G(n) = \Delta H(n) - T\Delta S^{conf}(n) \tag{5.4}$$

Keeping consistent with energy landscape theory, the free energy barrier appears as a function of the independent changes in folding enthalpy and conformational entropy with the height of the barrier simply tuned by modifying $k_{\Delta H}$ allowing the one-dimensional surface to model the profile from two-state to downhill.

Like the stabilization energy, an exponentially decaying heat capacity function ($\Delta C_p(n)$) is introduced to account for the temperature effect (equation 5.5) which is empirically correlated with a linear increase with protein size.

$$\Delta C_p(n) = N\Delta C_{p_{res}} \frac{1+e^{(k_{\Delta C_p} n)-1}}{1-e^{k\Delta C_p}} \tag{5.5}$$

For this function, the exponent dictates the heat capacity at the top of the free energy barrier if one is present (two-state) according to the curvature of $\Delta C_p(n)$. The total entropy can then be expressed as in equation 5.6 using 385 K as the entropy convergence temperature (Roberston, A. D. 1997) at which solvation effects on entropy are factored out.

$$\Delta S(T,n) = \Delta S^{conf}(n) + \Delta C_p(n)\ln(T/385) \tag{5.6}$$

The total change in folding enthalpy is then expressed at the midpoint temperature (equation 5.7):

$$\Delta H(T,n) = \Delta H^{o}(n) + \Delta C_{p}(n)T - T_{m} \tag{5.7}$$

Combining the total entropy and total enthalpy as a function of temperature and n, the one-dimensional free energy surface can be rewritten as (equation 5.8):

$$\Delta G(T,n) = \Delta H(T,n) - T\Delta S(T,n) \tag{5.8}$$

Relaxation kinetics as a result of fluctuations in the free energy surface are treated in a Kramers-like treatment following the matrix method for diffusion kinetics (Lapidus, L. 2002). The effective diffusion coefficient is then expressed as (equation 5.9):

$$D(T) = k_{o} \cdot e\left(\frac{-NE_{a,res}}{RT}\right) \tag{5.9}$$

where k_o is assumed to be temperature independent, N is the number of residues, and $E_{a,res}$ accounts for all changes in solvent viscosity and internal friction. It should be noted that modifications of the model for effects with chemical denaturants do exist, but are not included here.

5.2.1 Free Energy Barrier Heights from Kinetics

From the free energy surface generated from the 1-D FES model, the barrier height is calculated by partitioning the ‘nativeness’ reaction coordinate into thirds. At

two-thirds the distance between the two macroscopic minima, a dividing line is placed corresponding to the transition state ensemble and is assigned a width of 0.22 nativeness for thermal denaturation. The position of the transition state ensemble is fixed to quantitatively differentiate between a scenario with two macroscopic states and a profile exhibiting a single free energy minimum. The invariant width of the transition state ensemble is a way to ensure built-in fluctuations less than 1RT, which maximize the conformity between the folding-unfolding barrier heights and populations on each side of the barrier. The barrier heights can then acquired from the weighted probability of the native or unfolded state and the transition state ensemble.

5.2.2 Robustness of 1-D FES Model

To test the strength of the 1-D FES model for its ability to reproduce the experimental temperature dependent rates from a set of nine proteins, simulations were carried out applying a $\Delta S_{res}^{n=0}$ of 16.5 J mol^{-1} K^{-1} per residue at 385 K, $\Delta C_{p\,res}$ of 50 J mol^{-1} K^{-1} per residue and a $k_{\Delta C_p}$ of 4.3. To review, these parameters correspond to the entropy change of a residue going from its unfolded to folded conformation, the residue specific heat capacity change due to unfolding, and the heat capacity at the top of the free energy barrier (if present), respectively. The entropy cost of fixing the backbone conformation of a certain residue was obtained from Robertson and Murphy's empirical estimate (Robertson, A. D. 1997) while the residue specific heat capacity was obtained by applying the VB model to the calorimetry profiles of 14

different proteins. The obtained value of 50 J mol^{-1} K^{-1} per residue for $\Delta C_{p_{res}}$ agrees remarkably well to the empirical estimate of 58 J mol^{-1} K^{-1} per residue from the linear scaling of ΔCp with size. Simulations of T-jump kinetics for a pool of *ad hoc* proteins with differing barrier heights ranging from -0.6 to 8.2 kJ mol^{-1} at the midpoint temperature were then completed to illustrate the complete thermodynamic description of the relaxation rate and the effect of using a temperature independent diffusion coefficient. The resultant relaxation rate versus temperature plots duplicated the chevron-like behavior with an expected gradient of curvatures scaling with the barrier height. This shape is an indicator that the diffusion coefficient is indeed temperature dependent, as the more characteristic and pronounced 'V' would be apparent if D_{eff} was independent of temperature. An Arrhenius plot of the relaxation rate versus 1000/T using a temperature dependent D_{eff} with an activation energy of 45 kJ mol^{-1} showed the characteristic near-linear shape as the barrier approached that of a global downhill folder (i.e. 0) with a slight kink in at the midpoint temperature as an effect of the temperature dependent diffusion coefficient (i.e. faster relaxation at higher temperature, slower at low temperature). It is therefore convincing that an estimation of the barrier height can be directly extracted from the extent of changes in the temperature dependence of the relaxation rate and the diffusion coefficient from the temperature dependence.

The ability of the 1-D FES model to simulate the behavior between a two-state-like system from one with a negligible barrier is promising, but a model is only

as good as its ability to match experimental observations. After adjusting the per residue entropy and heat capacity to scale with the size of each protein in a nine protein sample database, a fit could be achieved requiring only four parameters, T_m, k_{DH} (free energy surface properties), k_o, and $E_{a,res}$ (diffusion coefficient). The fit to the experimental data was on par with that of the two-state fit originally used by the individual authors. However, the 1-D FES model offers an unparalleled advantage in comparison due to its ability to assess the temperature dependent barrier height and diffusion coefficient. The barrier heights, unattainable from the original kinetics experiments, were small to negligible in free energy, in agreement with the qualitative observation of a nearly-flat Arrhenius plot with a kink at T_m becoming more pronounced the higher the barrier height. Thus, the 1-D FES model provides a compelling method to obtain the elusive diffusion coefficient and extract the free energy barrier height from experimental kinetic data to compare with information gained in equilibrium.

5.3 gpW Infrared T-Jump Kinetics

With folding rates being correlated to protein size, the first expectation was that gpW would fold in a two-state fashion in which the kinetics would be resolved in the millisecond time scale. In parallel with the multiprobe analysis (Chapter 3) and calorimetric characterization (Chapter 4), unfolding kinetics monitored via a stopped-flow device coupled to a circular dichroism spectrophotometer were performed. The

result was unexpected, as the f-CD signal was fully recovered within the 2 millisecond dead-time of the instrument. This was the first kinetic evidence that gpW folds by ultrafast kinetics and is a great candidate to be a downhill folder.

After exploratory kinetic experiments, the equilibrium characterization was completed indicating thermodynamic signatures of gpW folding over a marginal barrier and confirmed by the calculated barrier height from applying the VB model to DSC data. To explore the agreement between thermodynamics and kinetics, nanosecond laser induced temperature-jump experiments were completed by detecting fluctuations of the backbone structure using infrared absorption at 1632 cm^{-1}. Temperature jumps of ~11 K from starting temperatures in the range of 300 – 360 K were achieved. The relaxation rates were fit to a single exponential decay with relaxation times (τ) between 33 and 1.7 μs. Figure 5.1 highlights the relaxation times of WT gpW ±10 K from the thermodynamically determined melting temperature (T_m) of 340 K. In this window, the relaxation times range from 3 μs at 350 K to 15.7 μs at 330 K.

The relaxation rate versus 1/T plot (Figure 5.2) illustrates a profile consistent with a protein folding in the microsecond regime over a marginal barrier based on the flattened shape of the curve. This near-linear curve shows a slight kink starting at the characteristic temperature with a concomitant increase in slope at increasing temperature indicating a slight barrier involved in the folding of WT gpW. The smooth trends in the plot and the small fitting errors (error bars are smaller than the

size of the circles) highlight the high data quality and permitting the observation of the change in slope.

In figure 5.3A, the amplitude of the relaxation decays are plotted as a function of the final temperature after the T-jump. The kinetic amplitude is a direct reflection of the shift in population after jump in temperature and is a kinetic parallel to the first derivative of the equilibrium unfolding curve. Thus, the kinetic amplitudes serve as an independent measure of to the agreement between equilibrium and kinetics rather than the more familiar enforcement in a two-state analysis. Figure 5.3B illustrates the agreement between the kinetic amplitudes (blue closed circles) and the FTIR data (red curve) with an agreeable overall shape and coincident melting temperatures at 340 K. Equilibrium and kinetic data should be consistent regardless of folding behavior (two-state or downhill) and thus, is an important experimental control. Further, having the rate and amplitude data allows the application of the one-dimensional free energy surface model to obtain simulations of the complete folding kinetics independently.

Using the 1-D FES model described above, the kinetic data can be fit to obtain an estimate of its one dimensional folding free energy surface. The four parameter fit of the model reproduces the experimental data very well. This can be seen by the simulated exponential decays (Figure 5.4), the change in the relaxation rate with temperature (red curve in figure 5.2), and the kinetic amplitudes (red curve in figure5.3). The fixed parameters used in the model were: ΔS_{res} = 16.5 J/(mol·K·res) and $\kappa\Delta C_p$ = 4.3 (from Naganathan, A.N. 2007), and $\Delta C_{p,res}$ = 40 J/(mol·K·res)

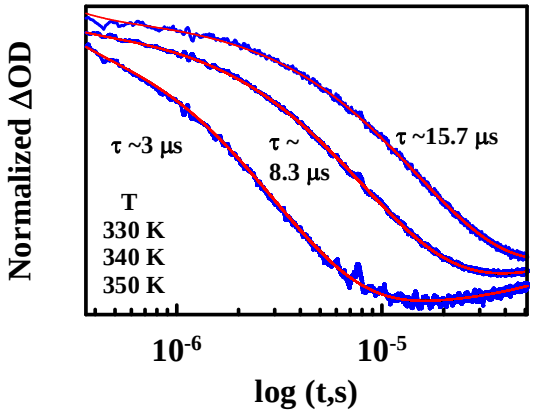

Figure 5.1 IR relaxation decays for WT gpW after ~11 K T-jumps. Red curves correspond to fits to a single exponential decay. The slight drift at the longest times corresponds to the beginning of recooling after the T-jump.

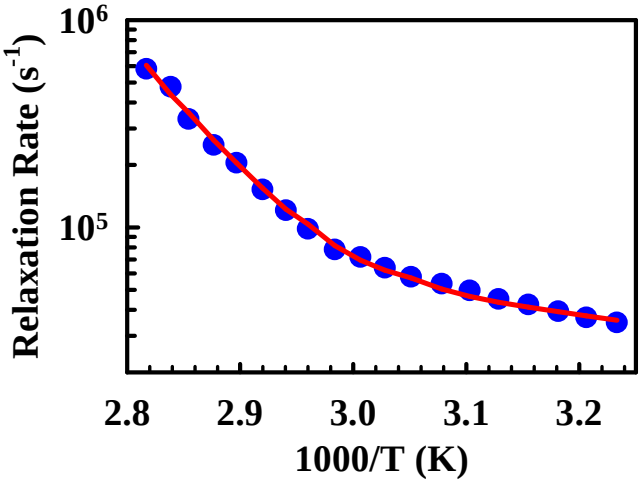

Figure 5.2 Relaxation rates from single exponential fits of the relaxation decays versus 1/T (blue closed circles); fit to the one-dimensional free energy surface model (red curve)

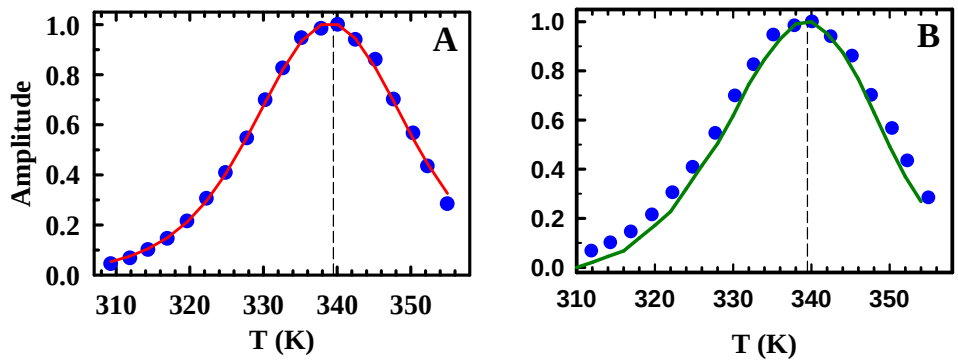

Figure 5.3 WT gpW (A) Kinetic amplitude of the T-jump relaxation at the final temperature after the jump (blue filled circles); fit to free energy surface model (B) Agreement between the kinetic amplitude (blue filled circles) and the first derivative of the FTIR unfolding curve (red curve). Black dashed lines in both plots indicate the Tm ≈ 340 K.

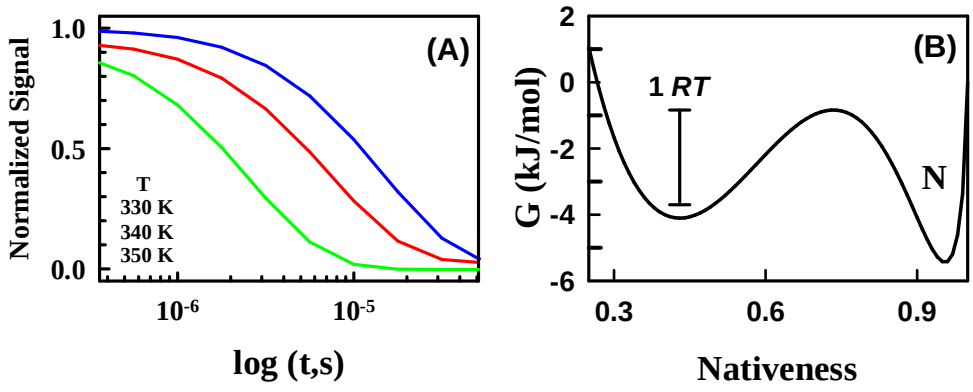

Figure 5.4 (A) Simulated relaxation decays for WT gpW from the free energy surface model for comparison with experimental decays (Figure 5.1) (B) One-dimensional folding free energy surface at the T_m = 340 K.

from the double perturbation analysis. The four fitted parameters were: ΔH^{340K} = 4.75 kJ/(mol·res), $\kappa\Delta H$ = 1.22 (properties of the free energy surface), and D_o = 7108 $n^2 \cdot s^{-1}$ (at 333 *K*), E_a = 59.2 *kJ/mol*. The activation energy (E_a) of the diffusion coefficient from the fitting is in close agreement with the empirical estimate of ~1 *kJ/mol* per residue obtained from the other ultrafast folding proteins (Naganathan, A.N. 2007) (61 residues = ~61 kJ/mol for gpW).

A free energy barrier of ~2.8 kJ/mol at 340 *K* is obtained from the fit to the rate and amplitude data signaling a marginal barrier (~1*RT*) at its maximal value (T_m) (Figure 5.5). The free energy barrier obtained from kinetics is within the range obtained by fitting the DSC data to the VB model. Therefore, the agreement between thermodynamics and kinetics indicates a well-controlled result and provides confirmation that the two approaches are valid.

As was done for the wild-type construct of gpW, a complete kinetic characterization for each mutant was completed for comparison of thermodynamic parameters obtained in equilibrium. Table 5.1 provides a summary of the relaxation times for each mutant with a spread distributed plus and minus ten degrees around the midpoint temperature. Also shown in table 5.1 are the kinetic barrier heights for each mutant for comparison to wild-type.

Mutant	T_m (K)	Relaxation Time (μs)	Barrier Height (kJ/mol)
E5H E6G A9H	320.5	32.3 > 16.9 > 7.1	2.5
R11H A12G	324.1	44.4 > 30.9 > 9.3	4.8
K45H A48G			5.5
E49H	319.3	60.2 > 30.9 > 11.4	
E34H V23H	332.8	17.3 > 11.0 > 6.3	5.1
E34H Q27H	339.2	10.4 > 5.5 > 4.8	7.3
Wild Type	340.0	15.7 > 8.3 > 3.0	2.8

Table 5.1 Comparison of relaxation times for each bi-histidine mutant with wild-type gpW. Relaxation times are shown from ten degrees below the T_m, at the T_m, and 10 degrees above the T_m. Melting temperature was determined from the maximum value of the kinetic amplitude.

The IR T-jump relaxation decays were fit to a single exponential identically as was done for the wild-type construct (data not shown). Corresponding relaxation times for each helical mutant were a bit longer, but in the same order of magnitude as wild-type. The kinetic barrier heights obtained from fits to the one-dimensional free energy surface model were similar in magnitude to the kinetically determined barrier height of wild-type, and all were within ~1 RT of each other (all remained below 3 RT). The melting temperatures obtained from fitting to the 1-D FES model reflect the differences in thermal stability observed in equilibrium f-CD thermal melt experiments.

The kinetic barrier height acquired from the 1-D FES model for the E5H E6G A9H mutant shows a nearly identical kinetic barrier height to WT. This similarity

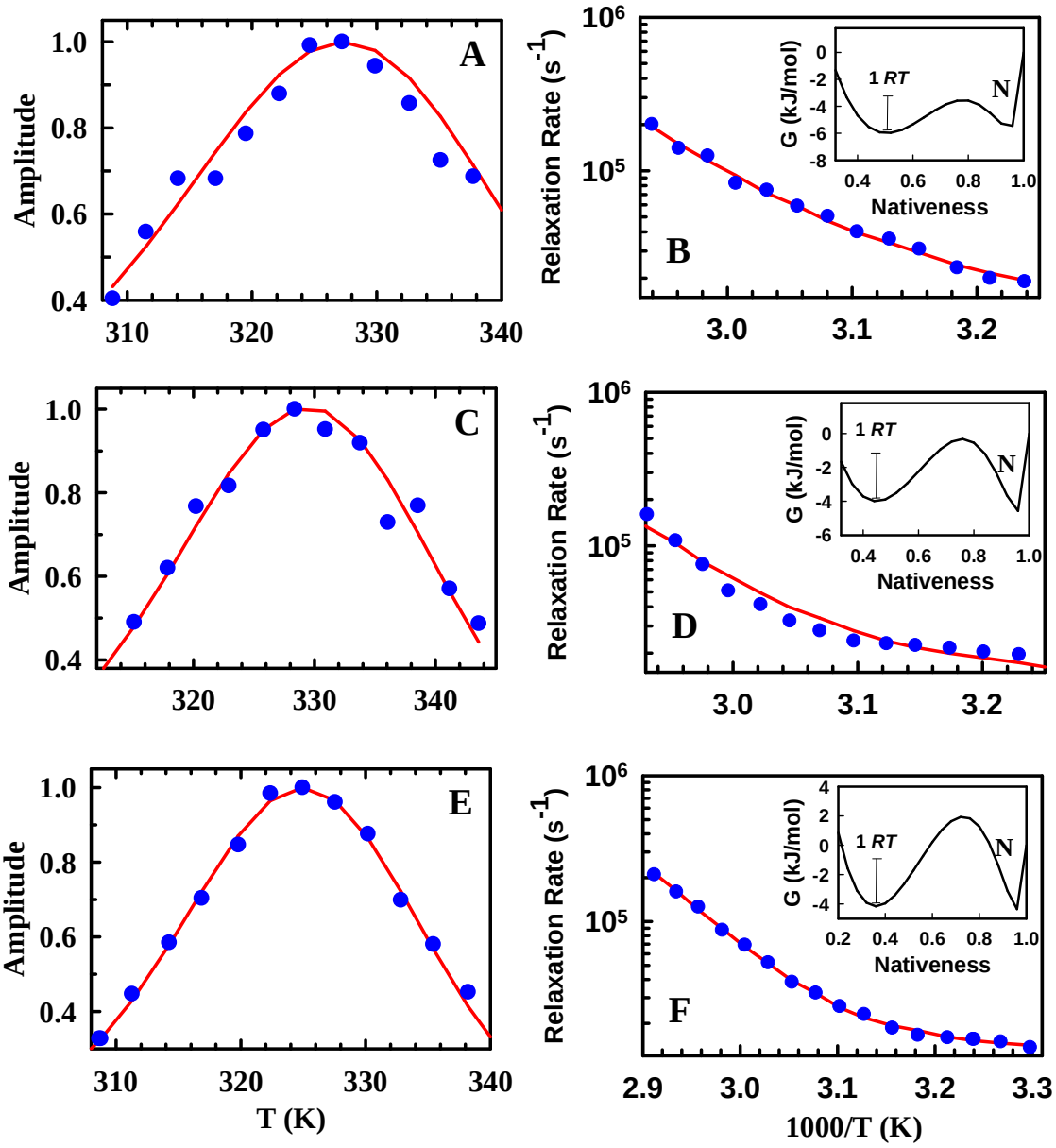

Figure 5.5 Panels A, C, and E are the kinetic amplitudes (blue circles) and fits to the 1-D FES model (red curves) for each of the helical mutants E5H, R11H, and K45H respectively. Panels B, D, and F illustrate the relaxation rate versus 1/T (blue circles) and fits to the 1-D FES model (red curves) with 1-D free energy profiles in the insets.

serves as a reference to the overall structure discussed earlier. Since it appears that tertiary packing is deficient in this mutant, it is not surprising to see a predicted barrier height of this magnitude. Upon comparison of the WT barrier heights, it is reasonable to assume that there may be a ≈ 1 RT standard deviation. Historically, the kinetically determined barrier height is high in comparison to thermodynamics, so the E5H mutant may well fold globally downhill by populating some degree of secondary structure, but no tertiary structure. A similar scenario could be true for the K45H mutant, which has a barrier height of 5.5 kJ/mol, although a greater degree of tertiary structure is observed for this mutant indicating that arrangement of global structure may be the cause of an increased barrier. Figure 5.5, panels A, C, and E, illustrate the kinetic amplitude (blue circles) obtained from the experimental relaxation decays and the fit (red curve) by the 1-D FES model for each of the helical mutants (E5H, R11H, K45H). Panels B, D, and F show the relaxation rate versus 1/T plots (blue circles) and the corresponding fits from the 1-D FES model for each helical mutant respectively. This serves as verification of ultrafast folding behavior while the insets in these panels illustrate the one-dimensional free energy profile. In each case, the expected change in slope of the plot near the midpoint transition temperature is indicative of folding over marginal barriers.

The kinetically determined T_ms of the β-sheet mutants show higher thermal stability in agreement with equilibrium measurements. However, the barrier heights are nearly two and three times higher compared to wild type for E34H V23H and

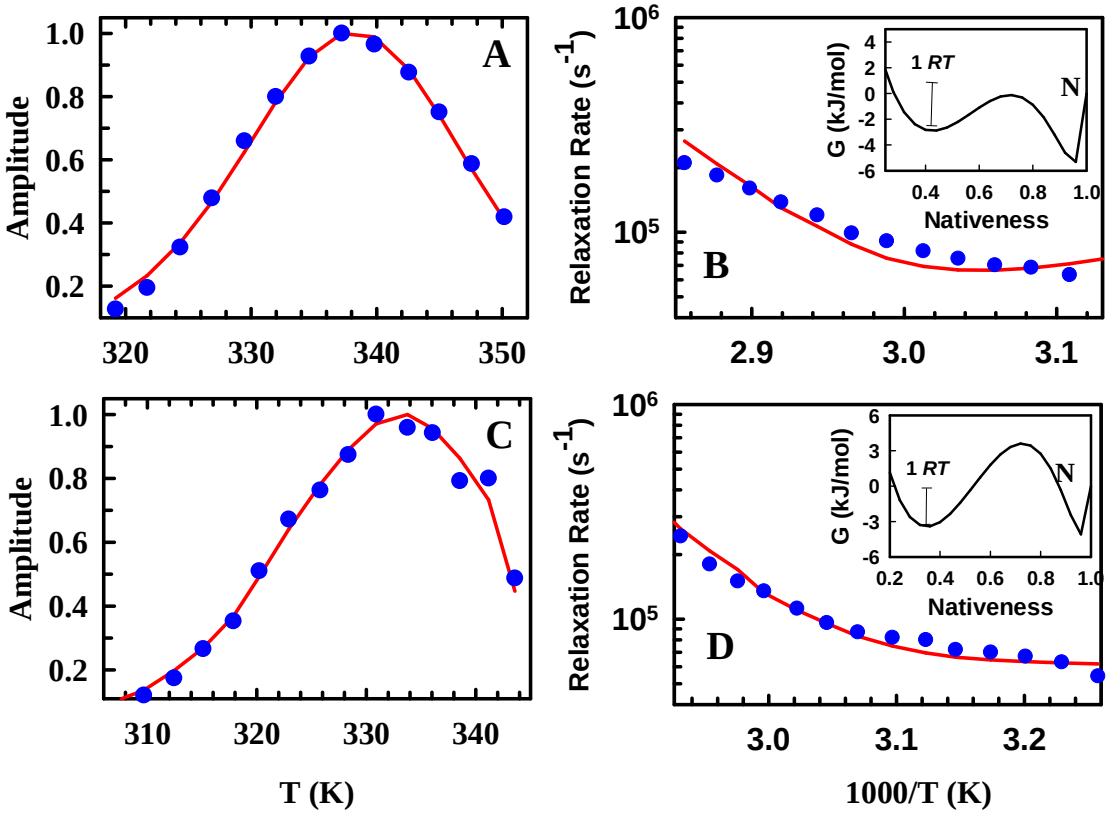

Figure 5.6 Panels A and C show kinetic amplitudes (blue circles) and fits to the 1-D FES model (red curves) for each of the beta mutants V23H and Q27H respectively. Panels B and D illustrate the relaxation rate versus 1/T (blue circles) and fits to the 1-D FES model (red curves) with 1-D free energy profiles in the insets.

E34H Q27H respectively, though still within the 3 RT limit for folding over a marginal barrier. Panels A and C in figure 5.6 illustrate the kinetic amplitudes and fits to the 1-D FES model (representative symbols identical to figure 5.5). Corresponding relaxation rates and one-dimensional free energy profiles are shown in panels B and D and are in agreement with experimental signatures of folding over marginal barriers in the microsecond timescale.

Comparing the folding barriers from fitting the kinetic data to the 1-D FES model reveals an interesting confirmation of the importance of the contacts made near the β-hairpin. E34H V23H is located near the N-terminal end of β1 on the opposite side of the hairpin. In contrast, mutantE34H Q27H is located very close to the hairpin on β1. With the higher barrier height calculated for the latter mutant, it is possible that this mutation disrupts native contacts at the interface of the β-hairpin and the first α-helix. Indeed, inspection of the WT contact map (Figure 3.2) shows strong long-range contacts formed between these two regions. Therefore, tertiary packing in this region could be cause for the increased barrier height forcing (in energy landscape terms) a more extensive conformational search to achieve native-like packing between these two structural elements.

5.4 Conclusions

The unanticipated ultrafast folding kinetics provides a first-approximation guide to the folding behavior of gpW. Relaxation kinetics were measured via a nanosecond laser-induced temperature-jump apparatus with structural alterations in the peptide backbone monitored with an infrared detector at 1632 cm^{-1}. The relaxation times were well-fit by a single exponential decay resulting in relaxation rates ranging from 1.7 μs at high temperature to 33 μs at low temperature (300 – 350 K). A plot of the relaxation rate versus 1/T revealed the ultrafast folding behavior with a shape correlating to that of other ultrafast folding proteins with marginal (3RT) to absent free energy barriers. The kinetic amplitude obtained is a direct reflection of the equilibrium unfolding curves; a first approximation of the agreement between thermodynamics and kinetics. Using the one-dimensional free energy surface model, the kinetic data was reproduced with precise accuracy. Simulated relaxation times were in accord with experimental data. Additionally, the relaxation rate versus 1/T and the kinetic amplitudes were duplicated with the model allowing the calculation of the free energy barrier at the midpoint temperature to be calculated. This barrier, ~2.8 kJ/mol, is within the one standard deviation range of the barrier height extracted from the variable-barrier model, validating the description of the free energy surface from equilibrium and kinetics.

Barrier heights of helical mutants are within 1.5 RT of the value for WT gpW and have melting temperatures in agreement with equilibrium techniques.

Quantitatively, both WT and mutant gpW barrier heights signal that these mutants fold over marginal barriers. Additionally, the folding barriers are a reflection of the expansive conformational search imposed by the mutant destabilization.

Mutants in the β-sheet fold over barriers higher than for wild-type, but remain marginally within the 3 RT limit. For these mutants, a picture of the native contacts necessary to minimize the folding barrier is achieved. For the E34H V23H mutant, the barrier height is within 1 RT of WT, while E34H Q27H is nearly 2 RT higher compared to WT. This is indicative of a region sensitive to tertiary interactions; a conclusion supported by strong contacts between these two structural elements as seen in the contact map.

In summary, WT gpW and each of the bi-histidine mutants are consistent with the expected behavior of a protein folding over a marginal barrier. Fitting of the kinetic amplitudes and relaxation rates as a function of temperature produced the qualitative symbols of downhill folding and allowed the calculation of folding free energy barriers that were each within the 3 RT marginal barrier limit. Thus, fitting of the experimental data to the phenomenological free energy surface model provides an unbiased calculation of kinetic behavior with only 4 fitting parameters and without enforcement of experimental data to a specific model. By using the 1-D FES model, an independent assessment of kinetic behavior is obtained for direct comparison to thermodynamic behavior where agreement is expected regardless of folding mechanism.

6. gpW Functional Perspectives

6.1 Introduction

Elucidation of the atomic-level details and structural ensembles along a reaction coordinate during protein folding provides a great deal of valuable information *in vitro*. However, this still leaves many questions unanswered and forces scores of assumptions to be made. Chief among the list of important questions is in regards to the biological relevance of *in vitro* experimentation to *in vivo* systems. The ultimate goal is to simulate conditions within the cell in order to draw biologically applicable conclusions. This is not always possible, or even practical, making meaningful inferences difficult.

In the previous chapters, a discussion of the monomeric folding behavior of the viral protein gpW has been made. In comparison to the first experimentally discovered downhill folder, BBL (and other PBSD homologs), gpW is an independently folding domain and not a subunit of a larger complex. Therefore, a certain consciousness of the 'big-picture' impact of downhill folding can be given further treatment.

Here, a review of the macromolecular organization of bacteriophage lambda will be given. Following, the role of gpW in that assembly will be outlined from experimental results aimed at probing mature virion construction. This will all serve

as a precursor to discussion on a possible biological role of gpW governed by downhill folding.

6.2 Mature Virions of Bacteriophage λ

Bacteriophage lambda (λ), as the name suggests, is a bacterial parasite. Common among viruses, initial infection of a host is followed by taking command of the cellular machinery for an extraordinarily selfish practice, reproduction. However, in order for the virus to proliferate, the very complex process of virion assembly must be completed. In the case of bacteriophage λ, mature infectious virions are composed of a head where DNA is packaged and a tail responsible for host transfection. Complete head construction is achieved by assembly of six gene products (gpW, gpD, gpB, gpW, gpFII, gpB, pX1, and pX2) (Georgopoulos, C. 1983), but the exact mechanism of assembly is highly ambiguous. Final assembly of bacteriophage λ includes the products of three independent phases: viral DNA replication, prohead assembly, and polymerization of tails. The most important process with respect to gpW is prohead assembly.

The λ procapsid, also referred to at the prohead, is an aggregate structure with an icosahedral geometry. The procapsid is composed primarily of gpE. At the base of the prohead, a doughnut-shaped portal complex is formed through which the DNA enters upon packaging and exits during infection. Assembly of the procapsid follows a step-wise pathway including the gene products, E, B, C, gpNu3, and host groEL/groES proteins. First, gpB monomers oligomerize into a 25 S dodecameric

ring with a hole ~25 Å in diameter. (Kochan, J. 1984). The groEL/groES chaperonins are likely useful in facilitating gpB folding, but the role of nuclease gpNu3 in ring formation is not known. Next, gpC adjoins to the preconnector yielding a 30 S initiator structure (which may include gpNu3), but neither the structure nor the stochiometry is known for this protein. The resulting initiator complex serves as a nucleation site for co-polymerization of gpNu3 and gpE, yielding an immature procapsid. GpNu3 directs polymerization of gpE into the icosahedral shell where the gpB / gpC initiator complex attaches at the portal vertex (Catalano, C.E. 2005).

The portal complex is both intricately constructed and processed. The main initiator complex protein is gpB*, is generated from proteolysis of at least 20 N-terminal residues of gpB. Secondly, pX1 and pX2 are formed which are thought to be products of a covalent fusion between gpC and gpE (Hendrix, R. W. 1974). The origin of these proteins both in chemical and functional terms, and mechanism, remain mysterious, but gpC is thought to act as the protease responsible for gpB* and pX proteins from sequence homology to a family of serine proteases (Baird, L. 1991). Maturity of proheads is finalized after gpNu3 is proteolytically cleaved leaving ≈420 copies of gpE, 12 copies of gpB*, and ≈6 copies of the pX proteins (Hendrix, R. W. 1975).

Bacteriophage λ DNA packaging can be described as a three-step process: initiation, propagation, and termination (Catalano, C.E. 2005). Initiation of DNA

packaging includes: (A) packaging machinery recognition and congregation at the concatemer genome junction (*cos* site) (B) nicking of duplex DNA at *cosN* by gpA (C) DNA strand separation. DNA translocation into the procapsid is described by steps which: (A) promote the stable terminase bound concatameric DNA to transition into a packaging motor and (B) translocation of the DNA into the capsid. The final step in DNA packaging involves steps which involve (A) recognition of next *cos* sequence by translocating complex, (B) DNA nicking, strand separation, and release of next concatemer sequence, and (C) assembly of head connector proteins and tail addition to produce an infectious virus.

Head expansion occurs upon DNA translocation where the capsid volume roughly doubles. The expansion is thought to be stabilized by gpD which adds as trimers to the outer surface and in similar numbers to gpE, the major capsid protein, resulting in the final icosahedral head structure (Murialdo, H. 1991). The head completion process includes addition of the accessory proteins gpFII and gpW. These proteins are thought to be responsible for preventing packaged DNA from being lost (Perucchetti, R. 1988) and as the primary mediators of the addition of tails (Casjens, S. 1974).

λ tail assembly consists of an assembly pathway made up of three steps: (A) formation of the initiator (from products of genes J, I, L, K, H, G, and M), (B) polymerization of the tail tube by the major tail protein, gpV and (C) termination and maturation of the tail by the addition of gpU and gpZ along with gpH cleavage. The

initiator structure is primarily composed by gpJ which forms the tail fiber responsible for direct interaction with the host cell (Katsura, I. 1983). Attached to the tail fiber is a short (3 discs) conical structure which leads into the main structural feature of the tail. Forming more than 80% of the tail, gpV appears to assemble in disc-like structures developing a tube-like structure of 135 nm in length. In all, 32 of these discs polymerize to create the tube of the tail. In an unknown regulatory mechanism, the tail is terminated once at the appropriate length and is followed by gpU attachment to the proximal end. The tail is finally completed by the addition of gpZ, though neither the mechanism nor the location of attachment is known. At this stage, the heads and tails are independently mature and can be attached to form an infectious virion.

6.3 λ Head-to-Tail Assembly

The termination step of bacteriophage λ DNA translocation is worth further consideration as gpW is involved at the intersection of the head and tail before construction of mature virions. The products of the F and W genes are believed to govern head completion. Indeed, complementation studies of F^- and W^- mutants were marked by the inability of the phage to form complete virions (Casjens, S. 1971). DNA filled head structures were observed by electron microscopy and active tails were present, but the heads and tails did not join together.

The addition of gpW is required before gpFII addition, which is in turn required for complete phage assembly of infectious virions. gpW addition to DNA

filled heads imparts greater DNase resistance of the packaged DNA, hinting at a probable role as a stopcock at the head connector or even interacting directly with DNA (Casjens, S. 1974, Perucchetti, R. 1988). The final step of head morphogenesis is the addition of gpFII, a monomeric independently folding single-domain protein of primarily loosely arranged beta structure, a single turn α-helix, and two large unstructured loops (residues 1-24 and 46-62 – total number of residues is 117) (Maxwell, K.L. 2002). GpFII serves as the attachment site for tail addition. As a matter of fact, λ phage lacking gpFII fail to produce mature virions even though DNA filled heads and active tails are still formed (Tsui, L. 1980). Based on the geometry and symmetry of the connector and tail, it is likely that gpFII also assembles into a hexameric ring structure (Casjens, S. 1974). The orientation or specific residues involved in the joining of heads and tails is unknown. An argument has been made proposed for the N-terminus being the binding site for heads (interaction with gpW) with the C-terminus being the binding site for tails (attaching to gpU) (Juhala, R.J. 2000), but this is debated by Maxwell though no alternative scenario was proposed (Maxwell, K.L. 2002).

6.4 Molecular Rheostats

Biological systems are no strangers to molecular switches. The existence of partially disordered proteins such as insulin and protease inhibitors, as well as "natively unfolded" proteins who's folding is coupled to binding highlights that

folding can also be a regulatory mechanism (Sosnick TR 2006). Regulatory pathways commonly employ mechanisms controlled by allosteric transitions which function as simple quasi on / off switches depending on structural conformation. The most popular examples are the R- to T-state transitions of the oxygen transport protein hemoglobin and the ATP-coupled chaperonin activity of the groEL / groES complex. Another class of these switches can be treated as intermediates to the on / off scenario; the so-called molecular rheostats. This particular class of switch can be thought of as analogs to electronic potentiometers. An example of such a system is the active site CXXC motif in thiol:disulfide oxireductases which have been shown to differentially control the flux of electrons involved in redox catalysis reactions. Signal transduction pathways are similarly controlled by concentration gradients of inducers and repressors. The list goes on with examples of transcription factors (Rizzino, A. 2008), structural moderators (Li, Q. 2007), and sensors (Moskvin, O.V. 2007) which are thought to commonly involve several chemical and physical interactions.

With respect to protein folding, the classical two-state mechanism has long been associated as an on / off affair in which the native and unfolded states are the active and inactive ensembles, respectively. With the discovery of downhill folding, such a simple explanation is no longer relevant. It would be fair to say that evolution has selected for certain behaviors depending on the nature of the structural transition. Two-state folders may act as binary switches as a function of an energy barrier. This leaves the door open to explain the function of downhill folders with an equally

simple rationalization. It is likely that the importance of downhill folding is entrenched in the continuous structural landscape imparted by the single free energy well. Analogous to a concentration gradient, a signal could be tightly regulated by the degree of structure formation present in the folding ensemble. Similarly, a coordination event between a multi-enzyme or multi-domain complex could be controlled by the oscillatory fluctuations in the structure at a common interface within the complex.

In the case of gpW, the exact function and its mechanism of action remain poorly defined. Its high pI (10.8) suggests a likely direct role in DNA packaging and stabilization, but its interaction with both gpB* and gpFII suggests a similarly important role as a moderator of protein-protein interactions. A monomer in solution, it is believed to oligomerize during head assembly (possibly into a hexamer) into a ring with six-fold rotational symmetry. In fact, when modeled as a hexamer with six-fold rotational symmetry, a ring structure similar to DNA sliding clamps is revealed (Maxwell, K.L. 2001). Support for this is provided by a monomer-to-hexamer transition of the tail terminator protein gpU upon incubation with magnesium ions (Katsura, I. 1977, Edmonds, L. 2007). Modeling of the NMR solution structure as a hexamer based on the solved crystal structure of a hexameric ring formed by a homologous gene product from bacteriophage P2 reveals a nearly identical oligimeric orientation to the homolog crystal structure (Edmonds, L. 2007). In fact, these same authors propose similar behavior for gpFII and go as far as to suggest that a conformational change could not only be a regulatory control in the assembly

process, but also in controlling multimerization after the observation that gpW and gpFII do not interact *in vitro* even at concentrations as high as 100 fold greater than expected in the cell (Maxwell K.L. 2002). It seems likely that gpW carries out multiple functions important to DNA packaging and injection as well as a scaffold protein in the bacteriophage λ connector region coordinating the assembly of packaged heads and active tails to form infectious virions.

6.5 Tuning Stability of gpW Bi-His Mutants

Given that gpW has been well-established as an effectively downhill folder, a distinct advantage is gained linking the folding behavior and the ability to focus a unique experimental lens on differing structural ensembles. The tunable nature of a one-state folding protein as a function of environment lead to characterizing the behavior of individual protons of the protein BBL by chemical shift perturbation (Sadqi, M. 2006). This served as the necessary proof to justify the independent unfolding nature of individual atoms. While this process is still cooperative in the folding / unfolding sense, it is not so in classical respects to the transition between the ground states.

gpW was first anticipated to exhibit two-state behavior based on the size and contact order correlations (Chapter 3). This expected behavior was to be used as a model to implement a two-dimensional mutational analysis which would expand the

already established one-dimensional ϕ- and ψ-value analyses (Chapter 1). These two methods are limited in their ability to correlate structure formation with energetic changes due to single point mutations or metal binding at a bi-histidine site, respectively. Bi-histidine mutants were generated (Chapters 2 and 3) to exploit the natural metal binding capacity of a pair of histidine residues oriented in three dimensional space such that formation of secondary structure could be stabilized upon metal binding. The first dimension of the analysis would be the bi-histidine pair, and would report on local structure. The second dimension would involve a second point mutation at a site far away in three-dimensional space. This mutation would report on global stability as a function of local perturbations and would allow the ability to report on structure of the transition state ensemble at a global site as a function of local stability.

Since gpW does not exhibit two-state behavior as was initially expected, application of the two-dimensional mutational analysis is inappropriate as there is an absence of a high energy transition state. However, since the structural ensembles of downhill folders can be pseudo-regulated, these mutants are still useful in analyzing structure formation. Building on this idea, the bi-histidine mutants can serve as a macroscopic model (compared to atom-by-atom unfolding monitored by NMR) of this idea of tunable stability.

Mutations were designed to take advantage of the natural metal binding behavior of a bi-histidine site. For helical mutants, a glycine was introduced

upstream of the first histidine point mutation to destabilize the helix. The introduction of glycine imparts greater flexibility of the peptide backbone and reduces the intrinsic helical content. The bi-histidine sites were designed to function as bidentate coordination sites for divalent metal binding (Figure 6.1). For helical mutants with i, i+4 alignment of sidechain orientations, cobalt was determined to be the best ligand while zinc was found to bind more efficiently for beta mutations (Krantz, B. A. 2004). This is most likely a result of the different coordination geometries of the octahedral Co^{2+} and tetrahedral Zn^{2+}. The addition of either cobalt or zinc is titrated in known quantities with the benefit of being able to tune the stability of either the α-helix or β-sheet in a continuous and predictable manner. This approach allows the ability to experimentally characterize the continuous structural ensemble both thermodynamically and kinetically as a function of metal concentration.

Borrowing from the experimental practices of the ψ-value analysis in modulating the flux of the transition state ensemble upon metal titration (Krantz, B. A. 2001, Krantz, B. A. 2004, Sosnick, T. R. 2004, Sosnick, T. R. 2006), the same concept can be applied to predictably tune the stability of the gpW mutants. After first determining the experimental range of metal concentrations where binding occurs (K_d), stability of the gpW mutants can be modulated in a continuous fashion by repetitious increases in metal concentration. Up to the saturation point, the

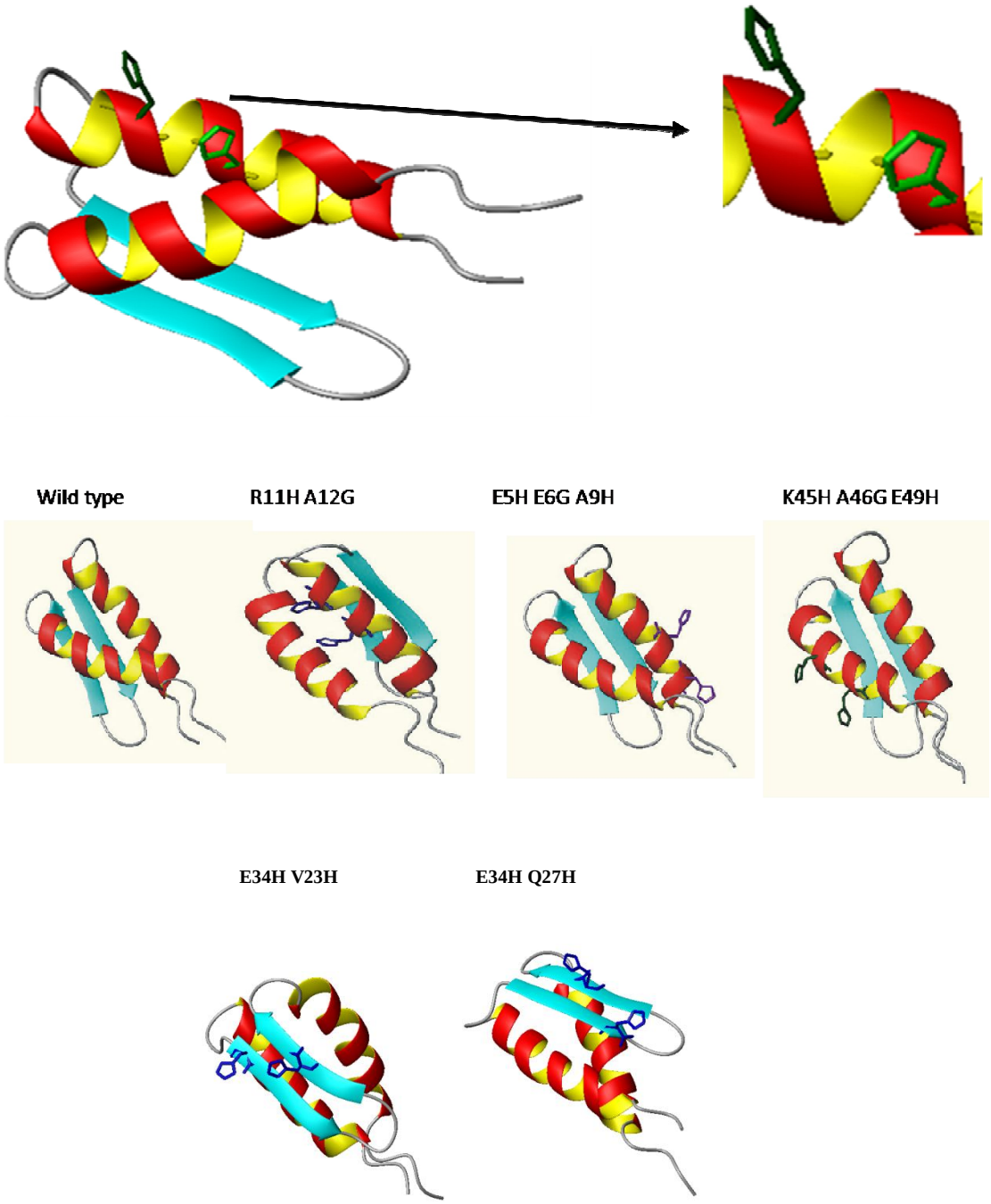

Figure 6.1 gpW mutant structures highlighting the metal binding bi-histidine sites located in the helical and beta-sheet structures.

expected result is to see a direct coupling of stability with metal concentration. In essence, upon metal binding, the magnitude of the free energy perturbation as a function of mutation should decrease. Thus, metal binding serves as a way to rescue stability, perhaps up to and beyond wild-type levels in a predictable and continuous manner as a function of metal ion concentration.

As was discussed earlier, each of the bi-histidine mutants, with an emphasis on the helical mutants, showed a large decrease in thermal stability demonstrating that these sites are important in forming local as well as global interactions. With addition of either Co^{2+} or Zn^{2+}, metal binding is expected to have the effect of lowering the free energy of fixing the mutant sidechains into their native conformation and increasing thermal stability. This is accomplished upon binding of metal ions to the bi-histidine site reducing the free energy of fixing the residues involved in the i, i+4 site. Essentially, metal binding acts as an inducer of helix nucleation. The ultimate hope would be that the T_m would be rescued to near wild-type levels as a function of increasing metal concentrations. However, the magnitude of stability is ultimately dependent several factors: (A) a positive binding event (B) the effective range of concentrations necessary to obtain a measurable dissociation constant and (C) the concentration of metal leading to binding site saturation.

Figure 6.2 provides a summary of initial metal binding experiments with mutants E5H E6H A9H, R11H A12G (alpha), and E34H V23H (beta). Panels A and D correspond to E5H E6G A9H with titrations of 0, 25, and 100 μM concentrations

of Co^{2+}. The thermal melting temperatures show a slight variation with the unfolding curves overlaid (Panel A). A plot of the first derivative illustrates up to 3 – 4 degrees of induced stabilization at 100 μM cobalt. Cobalt was gradually added over a wider range of metal concentrations for the R11H A12G mutant (0, 10, 100, 250, 1000 μM), but the melting temperatures were sharply coincident (Panels B and E). For the E34H V23H mutant,thermal melts were performed in the absence of metal and with the addition of both 250 μM Co^{2+} and Zn^{2+} (Panels C and F). However, neither metal had an observable effect on T_m. It is important to note also that reversibility was significantly compromised with the addition of metal ions. It appears that the addition of metal ions enhances the already complex behavior of these low stability mutants and implies that a thermodynamic analysis would be difficult due to low reversibility.

6.6 Conclusions

It may appear from the absence of increased thermal stability that these mutants are not capable of binding metal. These results are suspected of being caused by one, or a combination of factors. Metal binding is suspected of occurring as determined by evidence observed while purifying each mutant. Upon cleavage from the SUMO fusion system, the Ulp1 protease, SUMO fusion construct, and free gpW mutants were once again passed through the HisTrap column to liberate the mutants from the his-tagged protease and fusion protein. Upon injection, the

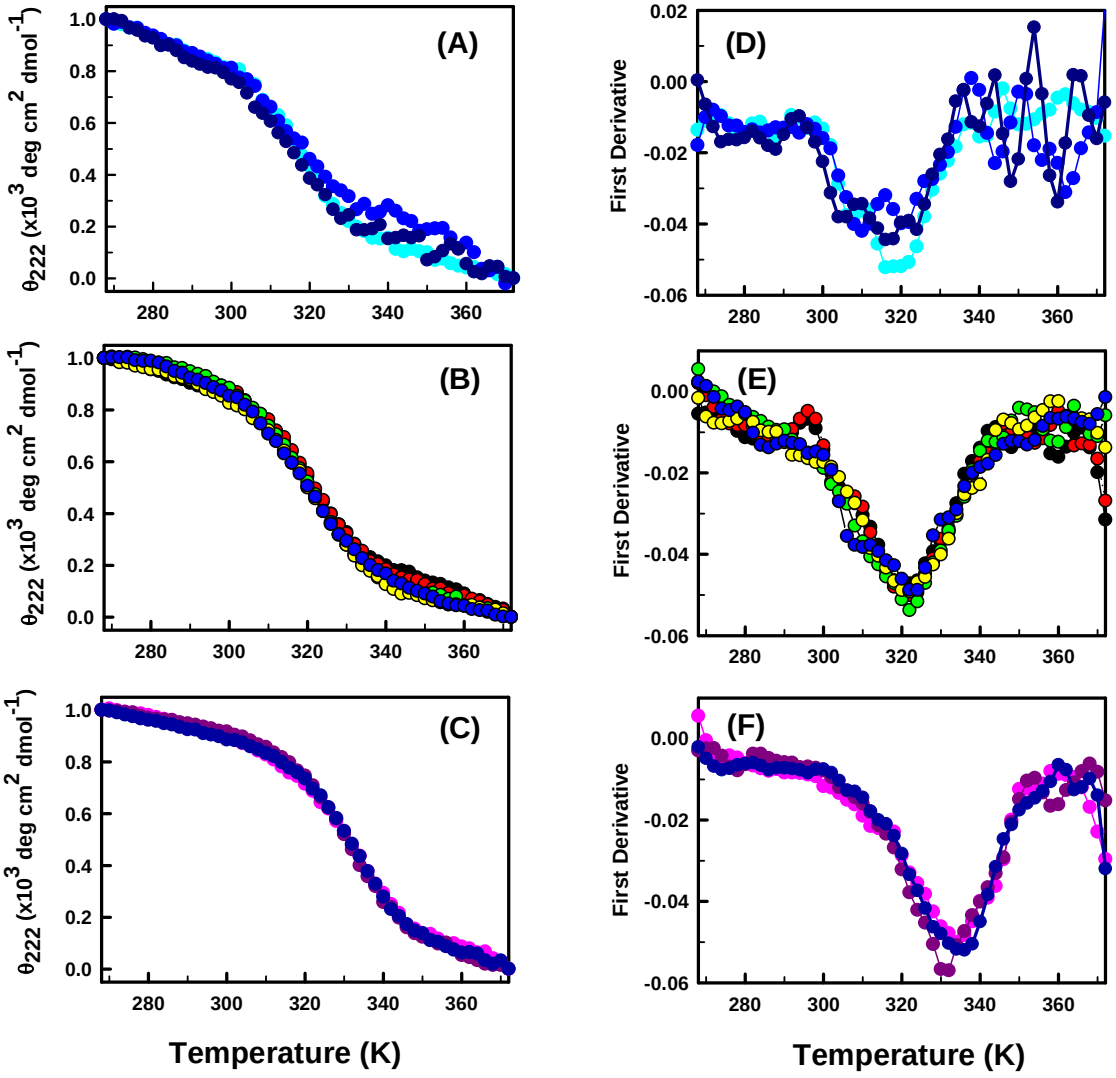

Figure 6.2 Normalized unfolding monitored by f-CD at 222 nm (A, B, and C) and first derivative showing melting temperatures (D, E, and F). Corresponding unfolding curves and derivative for mutants E5H E6G A9H, R11H A12G, and E34H V23H are shown paired horizontally with metal concentrations skewing with darkness of color (i.e. darker equal higher metal concentration).

free mutant did not readily elute as expected. Instead, imidazole had to be mixed with the running buffer at nearly 100 mM to elute the mutants from the Ni^{2+} charged resin. This was a positive sign of metal binding, but also may be the cause for the lack of increased stability upon the titration of cobalt in f-CD experiments. It is conceivable that there could be leaching of nickel ions into the eluate of gpW mutants thereby occupying the binding sites. This would have to be a saturation event as there was no apparent stability with increasing metal concentrations.

This becomes an issue if the binding constant is very low. At a sufficiently low dissociation constant, the equilibrium between bound and unbound would appear to favor the bound state. This of course must take into account that Ni^{2+} concentration would have to be present at equal concentration to the protein assuming 1:1 stochiometry. Therefore, the metal could be bound so tight, or the affinity of nickel could be sufficiently high compared to cobalt, that the addition of Co^{2+} has no impact on thermal stability.

Another scenario of a related nature is the possibility that the metal is bound in both the folded state and unfolded state. Coordination of the metal could be strong enough such that residual structure could be present in the unfolded state at the site of binding. This would in effect impart “native-like” stability tying in to the potentially low dissociation constant.

Several approaches were attempted to minimize the presence of metal ions in the protein solutions. First, samples of each mutant were subjected to treatment with

Chelex (Bio Rad) via batch method. However, upon centrifuging the slurry, the protein concentration in solution decreased by 95% as detected by UV absorbance. This implies that either the Chelex is responsible for precipitating the proteins, or the protein is bound to the Chelex mesh as a function of an extremely low dissociation constant. In other words, if the metal was bound to gpW sufficiently tight, yet the metal ions could also bind to the Chelex, the protein would not be found free in solution. This implies the possibility that the metal ion may only be bound to one histidine residue or a binding event involving single histidine residues from two protein molecules. This scenario is unlikely due to steric effects. The binding mechanism of Chelex may not require a free binding site on the metal, but may be adequately bound by some other means. This seems the more likely scenario.

A second attempt was made by dialyzing gpW mutants in a low pH buffer which should result in the ionization of the histidine sidechains perhaps causing the metal ions to be released. In parallel, samples were dialyzed in the same low pH buffer with the inclusion of Chelex in the dialysis buffer, but separated from the protein by a dialysis cassette. In both of these instances, dialysis resulted in the complete loss of protein as judged by UV absorbance. The loss in yield of gpW upon dialysis has been a continuous issue, perhaps due to protein adsorption to the dialysis membrane.

A final attempt at minimizing metal ion exposure included purifying the SUMO-gpW fusion with ion-exchange chromatography instead of taking advantage

of the convenient his-tag. There is very little information about the exact sequence of the SUMO protein making a calculation of the pI difficult. From the bioinformatics information available, it would appear that the pI for SUMO is much lower than gpW making ion-exchange a feasible means to separate the two proteins. Initial cation-exchange method development confirms that the SUMO construct and mutant gpW proteins can be separated, but further refinement of the method is necessary to eliminate co-elution of gpW with other proteins. A second method that should be explored is gel-filtration chromatography by taking advantage of the size difference of the two proteins.

Regardless of the unsuccessful effort to increase stability upon metal binding, this negative result serves as a positive control that the multiprobe analysis and kinetic experiments are not affected by metal binding.

7. Summary and Future Directives

With roots from an enzymological perspective, the protein folding problem has historically been explained with simple chemical models under the guise of well-defined step-by-step processes. The discovery of experimental signatures that could not be explained with such simple models have long been ignored and classified as experimental artifacts or with quantitatively unreasonable explanations.

Energy landscape theory provides a more comprehensive point of view with the idea of protein ensembles populating multiple trajectories in the conformational search to minimize free energy. An important issue is revealed which contrasts the classical chemical models in that there is a non-synchronous compensation between minimizing conformational entropy with increasing stabilization energy. The implication here is that folding behavior cannot be interpreted without a more quantitative approach.

The experimental identification of downhill folding has effectively bridged the gap between theory and experiment. What was once just a prediction now sees a growing class of proteins that exhibit ultrafast folding kinetics and marginal free energy barriers to folding. With the combination of both qualitative experimental signatures and quantitative physical models, an unequivocal approach to partition the folding behaviors of small proteins has been achieved. These methods are unbiased making an independent assessment of the folding behavior an objective process as opposed to employing the tactic of constraining data to fit a particular model. The result has led to several surprises including the reclassification of many of the small single-domain proteins which have been experimentally characterized to date.

The advantage of this unbiased approach is the ability to expand the pool of experimental results, thus improving on empirical correlations such as contact order and folding rates based on protein size.

The model of this work, gpW was expected to exhibit two-state folding behavior based on these empirical correlations. However, probing multiple structural features with several spectroscopic techniques provided the first evidence that gpW did not abide by these assumptions. Further, by applying the variable-barrier and the one-dimensional free energy surface models to experimental data, this protein could unquestionably be characterized as folding in the downhill regime with a marginal barrier height of less than 1 RT. This highlights the necessity of applying such statistical-mechanical and physical models.

Unlike predecessors, gpW is a natural single-domain protein which performs a specific function. A viral protein necessary for the virulent activity of bacteriophage lambda, its absence prevents the association of DNA-packaged heads and active tails. Although the details remain unclear, gpW functions at the interface of the head-to-tail connector region. Its chemical properties suggest a direct role in DNA binding and/or packaging, but gpW is also suspected of serving as a scaffold in the association of a multi-protein complex.

To explore the continuous nature of the downhill folding / unfolding transition, several mutants were engineered to incorporate a bi-histidine metal binding site to serve as probes of tunable stability. Each mutant exhibited all the experimental hallmarks of a marginal barrier, as with wild-type, although complications arose in the metal binding activity of the tested mutants.

Interestingly, the first experimentally characterized downhill folder, BBL, is an individual domain from a multi-domain complex. It is possible that this domain coordinates the folding or association activity of each domain through a continuous flux in structural formation. This is perhaps a trend common for all downhill folders. While the folding behavior of an evolutionary relative of BBL has pointed towards common behavior among this family, the folding behavior of homologously functioning proteins with respect to gpW has yet to be identified. This is a probable extension that should be explored to further investigate the link between folding behavior and function.

Metal binding experiments remain an interesting area of pursuit to study this link in gpW. A refinement of the purification process to eliminate encounters with metal ions would likely lead to the ability to predictably tune the stability of this protein as a proof of principal to the molecular rheostat concept. Correspondingly, the structural response to stabilizing a local interaction makes for an interesting model to study global stability. The E5H E6G A9H mutant would serve as a particularly good model to investigate this idea since it was this construct which exhibited the most catastrophic decrease in tertiary structure. Extracting barrier heights from DSC and kinetics experiments would similarly be exciting to explore the correlation between the increased barrier heights conceivably associated with packing of the first helix and beta hairpin.

Bibliography

1. Akmal, A. and V. Munoz (2004). The nature of the free energy barriers to two-state folding. Proteins 57(1): 142-52.

2. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. Science 181(96): 223-30.

3. Baird, L., B. Lipinska, S. Raina and C. Georgopoulos (1991). Identification of the Escherichia coli sohB gene, a multicopy suppressor of the HtrA (DegP) null phenotype. J Bacteriol 173(18): 5763-70.

4. Bennett, W. F., N. F. Paoni, B. A. Keyt, D. Botstein, A. J. Jones, L. Presta, F. M. Wurm and M. J. Zoller (1991). High resolution analysis of functional determinants on human tissue-type plasminogen activator. J Biol Chem 266(8): 5191-201.

5. Bertini, I., J. A. Cowan, C. Luchinat, K. Natarajan and M. Piccioli (1997). Characterization of a partially unfolded high potential iron protein. Biochemistry 36(31): 9332-9.

6. Bodenreider, C. and T. Kiefhaber (2005). Interpretation of protein folding psi values. J Mol Biol 351(2): 393-401.

7. Bowie, J. U. and R. T. Sauer (1989). Identifying determinants of folding and activity for a protein of unknown structure. Proc Natl Acad Sci U S A 86(7): 2152-6.

8. Bryngelson, J. D., J. N. Onuchic, N. D. Socci and P. G. Wolynes (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21(3): 167-95.

9. Bulaj, G. and D. P. Goldenberg (2001). Phi-values for BPTI folding intermediates and implications for transition state analysis. Nat Struct Biol 8(4): 326-30.

10. Burton, R. E., G. S. Huang, M. A. Daugherty, P. W. Fullbright and T. G. Oas (1996). Microsecond protein folding through a compact transition state. J Mol Biol 263(2): 311-22.

11. Callen, H. B. (1985). Thermodynamics and an Introduction to Thermostatistics. New York, Wiley.

12. Casjens, S. (1971). The Morphogenesis of the Phage Lambda Head: the Step Controlled by Gene F. The Bacteriophage Lambda. A. D. Hershey. Cold Spring Harbor, Cold Spring Harbor Laboratory: 725-746.

13. Casjens, S. (1974). Bacteriophage lambda FII gene protein: role in head assembly. J Mol Biol 90(1): 1-20.

14. Catalano, M. F. a. C. E. (2005). Viral Genome Packaging Machines: Genetics, Structure, and Mechanism. New York, New York, Eurekah.com and Kluwer Academic / Plenum Publishers.

15. Cline, D. J., C. Thorpe and J. P. Schneider (2003). Effects of As(III) binding on alpha-helical structure. J Am Chem Soc 125(10): 2923-9.

16. Crane, J. C., E. K. Koepf, J. W. Kelly and M. Gruebele (2000). Mapping the transition state of the WW domain beta-sheet. J Mol Biol 298(2): 283-92.

17. Cunningham, B. C. and J. A. Wells (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 244(4908): 1081-5.

18. Dragan, A. I. and P. L. Privalov (2002). Unfolding of a leucine zipper is not a simple two-state transition. J Mol Biol 321(5): 891-908.

19. Edmonds, L., A. Liu, J. J. Kwan, A. Avanessy, M. Caracoglia, I. Yang, K. L. Maxwell, J. Rubenstein, A. R. Davidson and L. W. Donaldson (2007). The NMR structure of the gpU tail-terminator protein from bacteriophage lambda: identification of sites contributing to Mg(II)-mediated oligomerization and biological function. J Mol Biol 365(1): 175-86.

20. Feng, H., N. D. Vu, Z. Zhou and Y. Bai (2004). Structural examination of phi-value analysis in protein folding. Biochemistry 43(45): 14325-31.

21. Ferguson, N., P. J. Schartau, T. D. Sharpe, S. Sato and A. R. Fersht (2004). One-state downhill versus conventional protein folding. J Mol Biol 344(2): 295-301.

22. Ferguson, N., T. D. Sharpe, P. J. Schartau, S. Sato, M. D. Allen, C. M. Johnson, T. J. Rutherford and A. R. Fersht (2005). Ultra-fast barrier-limited folding in the peripheral subunit-binding domain family. J Mol Biol 353(2): 427-46.

23. Fersht, A. R. (1988). Relationships between apparent binding energies measured in site-directed mutagenesis experiments and energetics of binding and catalysis. Biochemistry 27(5): 1577-80.

24. Fersht, A. R., J. W. Knill-Jones, H. Bedouelle and G. Winter (1988). Reconstruction by site-directed mutagenesis of the transition state for the activation of tyrosine by the tyrosyl-tRNA synthetase: a mobile loop envelopes the transition state in an induced-fit mechanism. Biochemistry 27(5): 1581-7.

25. Fersht, A. R., A. Matouschek and L. Serrano (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 224(3): 771-82.

26. Fersht, A. R. and S. Sato (2004). Phi-value analysis and the nature of protein-folding transition states. Proc Natl Acad Sci U S A 101(21): 7976-81.

27. Freire, E. (1995). Protein Folding and Stability. Totowa, NJ, Humana.

28. Fung, A., P. Li, R. Godoy-Ruiz, J. M. Sanchez-Ruiz and V. Munoz (2008). Expanding the Realm of Ultrafast Protein Folding: gpW, a Midsize Natural Single-Domain with alpha+beta Topology that Folds Downhill. J Am Chem Soc.

29. Garcia-Mira, M. M., M. Sadqi, N. Fischer, J. M. Sanchez-Ruiz and V. Munoz (2002). Experimental identification of downhill protein folding. Science 298(5601): 2191-5.

30. Garel, J. R. and R. L. Baldwin (1973). Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. Proc Natl Acad Sci U S A 70(12): 3347-51.

31. Georgescu, R. E., M. M. Garcia-Mira, M. L. Tasayco and J. M. Sanchez-Ruiz (2001). Heat capacity analysis of oxidized Escherichia coli thioredoxin fragments (1--73, 74--108) and their noncovalent complex. Evidence for the burial of apolar surface in protein unfolded states. Eur J Biochem 268(5): 1477-85.

32. Georgopoulos, C., Tilly, K., Casjens, S. (1983). Lambdoid Phage Head Assembly. Lambda II. R. W. Hendrix, Roberts, J. W., Stahl, F. W., Weisberg, R. A. Cold Spring Harbor, Cold Spring Harbor Laboratory: 279-304.

33. Goch, G., M. Maciejczyk, M. Oleszczuk, D. Stachowiak, J. Malicka and A. Bierzynski (2003). Experimental investigation of initial steps of helix propagation in model peptides. Biochemistry 42(22): 6840-7.

34. Goldenberg, D. P. (1999). Finding the right fold. Nat Struct Biol 6(11): 987-90.

35. Goldenberg, D. P., R. W. Frieden, J. A. Haack and T. B. Morrison (1989). Mutational analysis of a protein-folding pathway. Nature 338(6211): 127-32.

36. Gomez, J., V. J. Hilser, D. Xie and E. Freire (1995). The heat capacity of proteins. Proteins 22(4): 404-12.
37. Grantcharova, V. P., D. S. Riddle, J. V. Santiago and D. Baker (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. Nat Struct Biol 5(8): 714-20.

38. Gruebele, M. (2008). Fast Protein Folding. Protein Folding, Misfolding and Aggregation. V. Munoz. Cambridge, RSC Publishing: 106-138.

39. Gruebele, M. and P. G. Wolynes (1998). Satisfying turns in folding transitions. Nat Struct Biol 5(8): 662-5.

40. Guzman-Casado, M., A. Parody-Morreale, S. Robic, S. Marqusee and J. M. Sanchez-Ruiz (2003). Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of Thermus thermophilus ribonuclease H. J Mol Biol 329(4): 731-43.

41. Heinz, D. W., W. A. Baase and B. W. Matthews (1992). Folding and function of a T4 lysozyme containing 10 consecutive alanines illustrate the redundancy of information in an amino acid sequence. Proc Natl Acad Sci U S A 89(9): 3751-5.

42. Hendrix, R. W. and S. R. Casjens (1974). Protein fusion: a novel reaction in bacteriophage lambda head assembly. Proc Natl Acad Sci U S A 71(4): 1451-5.

43. Hendrix, R. W. and S. R. Casjens (1975). Assembly of bacteriophage lambda heads: protein processing and its genetic control in petit lambda assembly. J Mol Biol 91(2): 187-99.

44. Huang, G. S. and T. G. Oas (1995). Submillisecond folding of monomeric lambda repressor. Proc Natl Acad Sci U S A 92(15): 6878-82.

45. Irun, M. P., M. M. Garcia-Mira, J. M. Sanchez-Ruiz and J. Sancho (2001). Native hydrogen bonds in a molten globule: the apoflavodoxin thermal intermediate. J Mol Biol 306(4): 877-88.

46. Itzhaki, L. S., D. E. Otzen and A. R. Fersht (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J Mol Biol 254(2): 260-88.

47. Jackson, S. E. (1998). How do small single-domain proteins fold? Fold Des 3(4): R81-91.

48. Jackson, S. E., N. elMasry and A. R. Fersht (1993a). Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: a critical test of the protein engineering method of analysis. Biochemistry 32(42): 11270-8.
49. Jackson, S. E. and A. R. Fersht (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30(43): 10428-35.

50. Jackson, S. E., M. Moracci, N. elMasry, C. M. Johnson and A. R. Fersht (1993b). Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. Biochemistry 32(42): 11259-69.

51. Juhala, R. J., M. E. Ford, R. L. Duda, A. Youlton, G. F. Hatfull and R. W. Hendrix (2000). Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol 299(1): 27-51.

52. Katsura, I. (1983). Tail Assembly and Injection. Lambda II. R. W. Hendrix, Roberts, J. W., Stahl, F. W., Weisberg, R. A. Cold Spring Harbor, Cold Spring Harbor Laboratory: 331-346.

53. Katsura, I. and A. Tsugita (1977). Purification and characterization of the major protein and the terminator protein of the bacteriophage lambda tail. Virology 76(1): 129-45.

54. Kaya, H. and H. S. Chan (2000). Polymer principles of protein calorimetric two-state cooperativity. Proteins 40(4): 637-61.

55. Kelly, J. W. (2006). Structural biology: proteins downhill all the way. Nature 442(7100): 255-6.

56. Kholodenko, V. and E. Freire (1999). A simple method to measure the absolute heat capacity of proteins. Anal Biochem 270(2): 336-8.

57. Kim, P. S. and R. L. Baldwin (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem 51: 459-89.

58. Kim, P. S. and R. L. Baldwin (1990). Intermediates in the folding reactions of small proteins. Annu Rev Biochem 59: 631-60.

59. Kise, K. J., Jr. and B. E. Bowler (2002). Induction of helical structure in a heptapeptide with a metal cross-link: modification of the Lifson-Roig helix-coil theory to account for covalent cross-links. Biochemistry 41(52): 15826-37.

60. Kochan, J., J. L. Carrascosa and H. Murialdo (1984). Bacteriophage lambda preconnectors. Purification and structure. J Mol Biol 174(3): 433-47.

61. Krantz, B. A., R. S. Dothager and T. R. Sosnick (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. J Mol Biol 337(2): 463-75.

62. Krantz, B. A. and T. R. Sosnick (2001). Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. Nat Struct Biol 8(12): 1042-7.

63. Kubelka, J., J. Hofrichter and W. A. Eaton (2004). The protein folding 'speed limit'. Curr Opin Struct Biol 14(1): 76-88.

64. Kuhlman, B., G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard and D. Baker (2003). Design of a novel globular protein fold with atomic-level accuracy. Science 302(5649): 1364-8.

65. Levinthal, C. (1968). Are there pathways for protein folding? J. Chem. Phys. 65: 44-45.

66. Li, Q., M. R. Nance, R. Kulikauskas, K. Nyberg, R. Fehon, P. A. Karplus, A. Bretscher and J. J. Tesmer (2007). Self-masking in an intact ERM-merlin protein: an active role for the central alpha-helical domain. J Mol Biol 365(5): 1446-59.

67. Lopez, M. M., D. H. Chin, R. L. Baldwin and G. I. Makhatadze (2002). The enthalpy of the alanine peptide helix measured by isothermal titration calorimetry using metal-binding to induce helix formation. Proc Natl Acad Sci U S A 99(3): 1298-302.

68. Lumry, R. and R. Biltonen (1966). Validity of the "two-state" hypothesis for conformational transitions of proteins. Biopolymers 4(8): 917-44.

69. Ma, H. and M. Gruebele (2005). Kinetics are probe-dependent during downhill folding of an engineered lambda6-85 protein. Proc Natl Acad Sci U S A 102(7): 2283-7.

70. Makhatadze, G. I., V. N. Medvedkin and P. L. Privalov (1990). Partial molar volumes of polypeptides and their constituent groups in aqueous solution over a broad temperature range. Biopolymers 30(11-12): 1001-10.

71. Matouschek, A., J. T. Kellis, Jr., L. Serrano and A. R. Fersht (1989). Mapping the transition state and pathway of protein folding by protein engineering. Nature 340(6229): 122-6.

72. Matouschek, A., L. Serrano and A. R. Fersht (1992). The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. J Mol Biol 224(3): 819-35.
73. Maxwell, K. L., A. A. Yee, C. H. Arrowsmith, M. Gold and A. R. Davidson (2002). The solution structure of the bacteriophage lambda head-tail joining protein, gpFII. J Mol Biol 318(5): 1395-404.

74. Maxwell, K. L., A. A. Yee, V. Booth, C. H. Arrowsmith, M. Gold and A. R. Davidson (2001). The solution structure of bacteriophage lambda protein W, a small morphogenetic protein possessing a novel fold. J Mol Biol 308(1): 9-14.

75. Miller, J. H., C. Coulondre, M. Hofer, U. Schmeissner, H. Sommer, A. Schmitz and P. Lu (1979). Genetic studies of the lac repressor. IX. Generation of altered proteins by the suppression of nonsence mutations. J Mol Biol 131(2): 191-222.

76. Moskvin, O. V., S. Kaplan, M. A. Gilles-Gonzalez and M. Gomelsky (2007). Novel heme-based oxygen sensor with a revealing evolutionary history. J Biol Chem 282(39): 28740-8.

77. Mossessova, E. and C. D. Lima (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol Cell 5(5): 865-76.

78. Munoz, V. (2007). Conformational dynamics and ensembles in protein folding. Annu Rev Biophys Biomol Struct 36: 395-412.

79. Munoz, V. and J. M. Sanchez-Ruiz (2004). Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. Proc Natl Acad Sci U S A 101(51): 17646-51.

80. Murialdo, H. (1991). Bacteriophage lambda DNA maturation and packaging. Annu Rev Biochem 60: 125-53.

81. Naganathan, A. N., U. Doshi, A. Fung, M. Sadqi and V. Munoz (2006). Dynamics, energetics, and structure in protein folding. Biochemistry 45(28): 8466-75.

82. Naganathan, A. N., U. Doshi and V. Munoz (2007). Protein folding kinetics: barrier effects in chemical and thermal denaturation experiments. J Am Chem Soc 129(17): 5673-82.

83. Naganathan, A. N. and V. Munoz (2005). Scaling of folding times with protein size. J Am Chem Soc 127(2): 480-1.

84. Naganathan, A. N., R. Perez-Jimenez, J. M. Sanchez-Ruiz and V. Munoz (2005a). Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. Biochemistry 44(20): 7435-49.
85. Naganathan, A. N., J. M. Sanchez-Ruiz and V. Munoz (2005b). Direct measurement of barrier heights in protein folding. J Am Chem Soc 127(51): 17970-1.

86. Oliva, F. Y. and V. Munoz (2004). A simple thermodynamic test to discriminate between two-state and downhill folding. J Am Chem Soc 126(28): 8596-7.

87. Onuchic, J. N., Z. Luthey-Schulten and P. G. Wolynes (1997). Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 48: 545-600.

88. Ozkan, S. B., I. Bahar and K. A. Dill (2001). Transition states and the meaning of Phi-values in protein folding kinetics. Nat Struct Biol 8(9): 765-9.

89. Perucchetti, R., W. Parris, A. Becker and M. Gold (1988). Late stages in bacteriophage lambda head morphogenesis: in vitro studies on the action of the bacteriophage lambda D-gene and W-gene products. Virology 165(1): 103-14.

90. Plaxco, K. W., K. T. Simons and D. Baker (1998). Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 277(4): 985-94.

91. Raleigh, D. P. and K. W. Plaxco (2005). The protein folding transition state: what are Phi-values really telling us? Protein Pept Lett 12(2): 117-22.

92. Richardson, J. M., M. M. Lopez and G. I. Makhatadze (2005). Enthalpy of helix-coil transition: missing link in rationalizing the thermodynamics of helix-forming propensities of the amino acid residues. Proc Natl Acad Sci U S A 102(5): 1413-8.

93. Riddle, D. S., V. P. Grantcharova, J. V. Santiago, E. Alm, I. Ruczinski and D. Baker (1999). Experiment and theory highlight role of native state topology in SH3 folding. Nat Struct Biol 6(11): 1016-24.

94. Rizzino, A. (2008). Transcription factors that behave as master regulators during mammalian embryogenesis function as molecular rheostats. Biochem J 411(2): e5-7.

95. Robinson, C. R., Y. Liu, J. A. Thomson, J. M. Sturtevant and S. G. Sligar (1997). Energetics of heme binding to native and denatured states of cytochrome b562. Biochemistry 36(51): 16141-6.

96. Sadqi, M., D. Fushman and V. Munoz (2006). Atom-by-atom analysis of global downhill protein folding. Nature 442(7100): 317-21.

97. Sadqi, M., L. J. Lapidus and V. Munoz (2003). How fast is protein hydrophobic collapse? Proc Natl Acad Sci U S A 100(21): 12117-22.

98. Sanchez, I. E. and T. Kiefhaber (2003). Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. J Mol Biol 334(5): 1077-85.

99. Serrano, L., A. Matouschek and A. R. Fersht (1992b). The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. J Mol Biol 224(3): 805-18.

100. Serrano, L., A. Matouschek and A. R. Fersht (1992a). The folding of an enzyme. VI. The folding pathway of barnase: comparison with theoretical models. J Mol Biol 224(3): 847-59.

101. Sidhu, S. S. (2000). Phage display in pharmaceutical biotechnology. Curr Opin Biotechnol 11(6): 610-6.

102. Siedlecka, M., G. Goch, A. Ejchart, H. Sticht and A. Bierzyski (1999). Alpha-helix nucleation by a calcium-binding peptide loop. Proc Natl Acad Sci U S A 96(3): 903-8.

103. Smith, G. P. and V. A. Petrenko (1997). Phage Display. Chem Rev 97(2): 391-410.

104. Sosnick, T. R., R. S. Dothager and B. A. Krantz (2004). Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. Proc Natl Acad Sci U S A 101(50): 17377-82.

105. Sosnick, T. R., B. A. Krantz, R. S. Dothager and M. Baxa (2006). Characterizing the protein folding transition state using psi analysis. Chem Rev 106(5): 1862-76.

106. Thirumalai, D. (1995). From minimal models to real proteins - Time scales for protein-folding kinetics. Journal de Physique 5(11): 1457-1467.

107. Tsui, L. and R. W. Hendrix (1980). Head-tail connector of bacteriophage lambda. J Mol Biol 142(3): 419-38.

108. Viguera, A. R., J. C. Martinez, V. V. Filimonov, P. L. Mateo and L. Serrano (1994). Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. Biochemistry 33(8): 2142-50.

109. Weiss, G. A., C. K. Watanabe, A. Zhong, A. Goddard and S. S. Sidhu (2000). Rapid mapping of protein functional epitopes by combinatorial alanine scanning. Proc Natl Acad Sci U S A 97(16): 8950-4.

110. Weissman, J. S. and P. S. Kim (1993). Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. Nature 365(6442): 185-8.

111. Weissman, J. S. and P. S. Kim (1995). A kinetic explanation for the rearrangement pathway of BPTI folding. Nat Struct Biol 2(12): 1123-30.

112. Weissman, J. S. and P. S. Kim (1992a). Kinetic role of nonnative species in the folding of bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A 89(20): 9900-4.

113. Weissman, J. S. and P. S. Kim (1992b). The pro region of BPTI facilitates folding. Cell 71(5): 841-51.

114. Weissman, J. S. and P. S. Kim (1991). Reexamination of the folding of BPTI: predominance of native intermediates. Science 253(5026): 1386-93.

115. Wertman, K. F., D. G. Drubin and D. Botstein (1992). Systematic mutational analysis of the yeast ACT1 gene. Genetics 132(2): 337-50.

116. Yang, W. Y. and M. Gruebele (2003). Folding at the speed limit. Nature 423(6936): 193-7.

117. Yang, W. Y. and M. Gruebele (2004). Folding lambda-repressor at its speed limit. Biophys J 87(1): 596-608.

118. Zhou, Y., C. K. Hall and M. Karplus (1999). The calorimetric criterion for a two-state process revisited. Protein Sci 8(5): 1064-74.