BIOPHYSICAL CHARACTERIZATION OF PROTEIN FOLDING AND MISFOLDING

A Dissertation

by

JASON PETER SCHMITTSCHMITT

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2003

Major Subject: Biochemistry

BIOPHYSICAL CHARACTERIZATION OF PROTEIN FOLDING AND MISFOLDING

A Dissertation

by

JASON PETER SCHMITTSCHMITT

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

J. Martin Scholtz (Chair of Committee)	C. Nick Pace (Member)
Andy C. LiWang (Member)	Gary Kunkel (Member)
	Gregory D Reinhart (Head of Department)

December 2003

Major Subject: Biochemistry

ABSTRACT

Biophysical Characterization of Protein Folding and Misfolding.

(December 2003)

Jason Peter Schmittschmitt, B.S., University of Kentucky

Chair of Advisory Committee: Dr. J Martin Scholtz

The HPr proteins were characterized as folding by a two-state folding mechanism. Here, we present a comparison of the equilibrium and kinetic folding for the HPr protein from *Bacillus subtilis*, *E coli* and a key variant from these proteins. For the wild-type protein we find that ΔG_{HX} is greater than ΔG_{UDC}, suggesting that the HPr does not fold by a simple two-state mechanism. This discrepancy is revealed by testing the two-state nature of the folding reaction of HPr with mutation. We show that removing a single charge side chain (Asp 69) converts the HPr protein back to a simple two-state mechanism.

Ribonuclease Sa and two charge-reversal variants can be converted into amyloid *in vitro* by the addition of 2,2,2-triflouroethanol (TFE). We report here amyloid fibril formation for these proteins as a function of pH. The pH at maximal fibril formation correlates with the pH dependence of protein solubility, but not with stability, for these variants. Additionally, we show that the pH at maximal fibril formation for a number of

well-characterized proteins is near the pI, where the protein is expected to be the least soluble. This suggests that protein solubility is an important determinant of fibril formation.

This dissertation is dedicated to my grandparents, Sophie and Edward Zinda.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. J. Martin Scholtz, for his guidance and for giving me considerable latitude in choosing the direction of my research. I would also like to thank past members of the lab, Dr. Eric Nicholson and Dr. Ron Peterson, for their generosity with their time and expertise. The members of the Pace Scholtz Labs, past and present, have all contributed to my graduate work, either by engaging in helpful discussions or by giving me the push I needed to finish my research. The staff of the Medical Biochemistry and Genetics Department have been a valuable resource over the years. Finally, thanks to my friends and family for all their support over the years.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Why do we study protein folding? Understanding how proteins fold could be the key to understanding life. Proteins are involved in just about every aspect of the maintenance of cells and are the targets of may drugs. In addition, there are about 20 known protein-misfolding diseases including, Alzheimer's and Huntington's. A physical understanding of how a protein folds or misfolds will help us to develop drugs that recognize target proteins or fix proteins that have misfolded. For example, antibodies, which are proteins, have for many years been seen as useful therapeutics for a number of human diseases ranging from rheumatoid arthritis to leukemia because they are designed to target particular cells and attract other parts of the immune system to the site. There are a dozen antibodies that are approved as therapeutics by the U.S. Food and Drug Administration, and many more under development. Although the common person may not know what proteins are, they affect every aspect of their life.

One of the first things I learned as an undergraduate student is that protein structure = function. This addresses the idea that for a protein to work, it must fold into a structure. To understand this paradigm, we must have a clear comprehension of how a

---

This dissertation follows the style and format of the *Journal of Molecular Biology*.

protein folds and what forces stabilize the folded structure. While much is known about primary structure and the final native state of folded proteins, the folding pathway itself is still not well understood. In the late 1960's, Cyrus Levinthal performed a simple calculation to determine how long it might take for a protein to fold into its native structure. In a standard illustration of Levinthal's paradox, if we limit each connection between an amino acid residue to three possible states, then a polypeptide chain of 101 amino acids could exist in $3^{100} = 5 * 10^{47}$ configurations. Even if a protein were capable of sampling 10^{13} configurations per second, it would take 10^{27} years to try them all. Therefore, Levinthal concluded that proteins must fold by specific pathways.[1; 2] This introduced the hypothesis of "kinetic control" and lead to a search for folding pathways, there are several theories on the role and nature of protein folding pathways, which can be separated into two groups: the hierarchical model which states that preformed secondary structure fold into tertiary structures[3; 4], and the hydrophobic collapse model in which parts of the protein are brought together (or nucleate) to form the beginnings of tertiary structure and then secondary structures form.[5; 6]

The hierarchical model has intellectual appeal, because it reduces the folding problem to a simple understanding of the individual secondary structures and how they assemble. A good metaphor for this folding model is "prefabricated construction", where all the individual parts (α helices, β strands) are fabricated first and then assembled into the final tertiary structure. This model gained support when it was first demonstrated that relatively small peptides could form stable secondary structures. Previously, most peptides removed form proteins were not stable enough to form secondary structural

elements without tertiary contacts. Often, when secondary structural are isolated they tend to adopt helical structures, turns and less frequently β structures like β hairpins[3]. These structures might represent the starting point for folding. Unfortunately, the largest opponent is that most peptides removed from proteins do not form structure.

Support for the hydrophobic collapse model came from early studies showing that the energetics of removal of water from a hydrophobic group to be considerable. [7] However, the idea that a conformational search is facilitated within a nonspecific hydrophobic globule presents a problem. The excess of interactions hinders reorganization of the protein. A redefined model suggests that the secondary structural elements formed during the collapse, thus limiting the need for large structural rearrangements.

In the past few years a "new view" of protein folding has emerged.[1] The "classical view" of protein folding suggested that proteins fold along a pathway with specific populated intermediates. The classical view looked at protein folding occurring along a reaction coordinate where the rate-limiting step was determined by the free energy of a single transition state. The new view holds that the "transition state" is actually an ensemble of many structures and there are multiple folding routes. Whether any or all of these models is correct is uncertain.

This dissertation consists of two distinctly different projects. The first was to perform a comparison of the equilibrium and kinetic folding for the HPr proteins from *Bacillus subtilis* and *E. coli*. These proteins exhibit super protection. Super protection is the lack of agreement between conformational stability determined by hydrogen

exchange ΔG_{HD} and traditional methods ΔG_{UDC}. It has been shown that with the proper corrections (discussed later in this chapter), agreement between ΔG_{HD} and ΔG_{UDC} can be achieved for a variety of different proteins.[8] We show that the HPr proteins do not fold in a simple two-state manner, but removing a single charged side chain (Asp 69) is enough to convert the folding transition for HPr to a simple two-state mechanism.

The second project investigates the relationship between protein solubility, stability and amyloid fiber formation using ribonuclease Sa (RNase Sa) as a model system. RNase Sa is a small acidic protein (96 residues, pI=3.5) with a mixed $\alpha + \beta$ structure. We have recently made variants that replaced solvent-exposed acidic residues with lysine residues. The mutant with three such charge reversals is denoted 3K, and that with five is 5K. By reversing the charge at these sites we are able to study the relationship between stability, solubility and fibril formation in a single protein over a wide pH range. We also describe a method to isolate an on-pathway intermediate of RNase Sa amyloid formation.

BACKGROUND

Measuring the conformational stability of a protein requires the determination of the equilibrium constant and ultimately the free energy of this reaction:

$$Native(N) \xleftrightarrow{Keq} Unfolded(U) \qquad (1)$$

We will explore the many methods that have been devised to measure conformational stability including solvent denaturations, thermal denaturations, hydrogen-deuterium exchange, and folding kinetics.

Solvent denaturations

For a solvent denaturation, we monitor a spectroscopic signal that represents the native population as a function of denaturant. Urea and guanidine chloride (GdnHCl) are the most commonly used denaturants for unfolding studies. The spectral probes most often used to monitor unfolding are tryptophan fluorescence [9; 10], and far-UV circular dichroism (CD).[11] Tryptophan fluorescence reports on the specific environment of the tryptophan (Trp) residue making it an excellent probe for local environmental fluctuations during unfolding, whereas far-UV CD is an excellent reporter of secondary structure. In order to define the native and unfolded states, the spectra of the native and unfolded protein must be determined. The native state is the spectrum of the protein in the absence of denaturant, whereas the unfolded state is the spectrum in the highest concentration of denaturant (10 M urea or 8 M GdnHCl). A single wavelength that shows a significant difference between native and unfolded states is chosen to monitor the unfolding reaction (Figure 1).

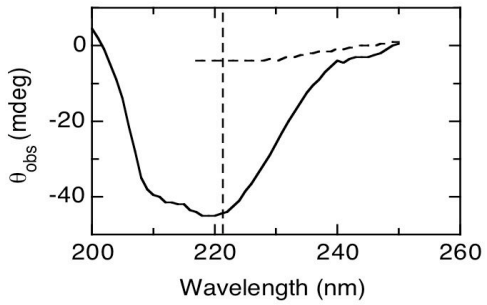

Figure 1. Representative circular dichroism spectra of a folded and unfolded protein. The solid line represents the far-UV CD spectrum of a folded protein. The dashed line represents the unfolded spectra of the same protein. The vertical dashed line represents the wavelength of greatest separation.

Figure 2 shows a typical equilibrium urea denaturation for *bs*HPr using CD at 222 nm to monitor the unfolding reaction. The curve is divided into three regions: the pre- transition region, the transition region and the post-transition region. The pre- and post-transition regions show how the denaturant affects the folded and unfolded protein and the transition region shows how denaturant effects the change in concentration of the native state with respect to the unfolded state. As with any thermodynamic measurement, it is essential that equilibrium is reached and the reaction is reversible.

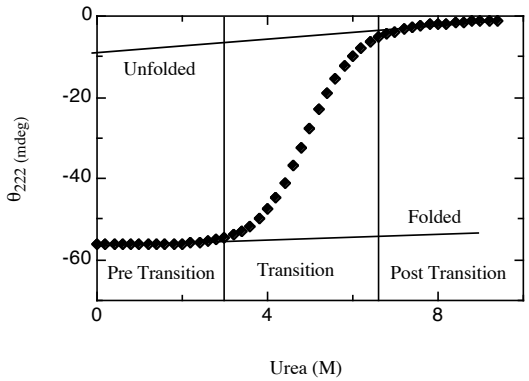

Figure 2. Representative urea denaturation curve. The unfolding reaction was monitored by far-UV CD at 222 nm. The three major regions of a denaturation curve are segregated by a vertical line and labeled pre-transition, transition, and post-transition. These regions were chosen arbitrarily and have no mathematical significance. The horizontal lines represent the folded and unfolded signals that will be used to convert the data into fraction unfolded (see Figure 3).

Many small proteins have been shown to unfold in a two-state folding mechanism.[12; 13; 14; 15; 16] We will assume a two-state folding mechanism for the discussion here, however, three-state mechanisms will be discussed later. Assuming a two-state mechanism means that for the points on Figure 2 only folded and unfolded proteins are present ($f_F+f_U=1$), where f_F and f_U are the fraction protein present in the folded and unfolded $\Theta = y_F f_F + y_U f_U$, where y_F and y_U represent the spectroscopic values characteristic of the folded and unfolded states and f_F and f_U are the fractions of folded and unfolded molecules. Combining these equations gives:

$$f_U=(y_F-y)/(y_F-y_U) \qquad (2)$$

Using the data from Figure 2 and equation (2), the data were expressed as fraction unfolded in Figure 3. The equilibrium constant K_{eq}, and the free energy change ΔG can be calculated using equations (3 and 4).

$$K_{eq}=f_U/f_N=f_U/(1-f_U)=(y_F-y)/(y-y_U) \quad (3)$$

$$\Delta G=-RT\ln[K_{eq}] = -RT\ln[(y_F-y)/(y-y_U)] \quad (4)$$

where R is the gas constant (1.987 cal/mol K) and T is the absolute temperature in Kelvin. Values for y_F and y_U are obtained by extrapolating the pre- and post-transition baselines to 0 M denaturant (Figure 2).

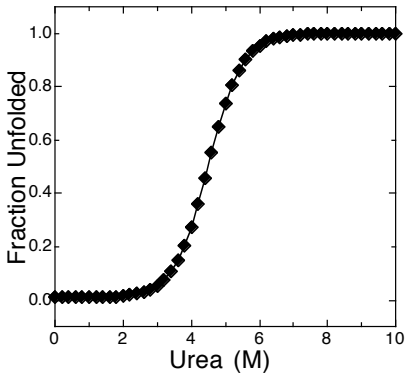

Figure 3. Fraction unfolded as a function of urea. This graph was constructed from the data shown in Figure 2 using the equation 2. The points represent the fraction of unfolded protein present at each urea concentration.

While being able to calculate ΔG in 4 to 5 M urea is helpful, what we really want is to measure the stability in water. Pace and Greene [17] noticed that ΔG varies linearly with the molar concentration of denaturant within the transition region (Figure 3). Using what has become know as the linear extrapolation method (LEM), they extrapolated the stability to 0 M denaturant (Figure 4).[17] We utilize equation (5) to determine ΔG:

$$\Delta G=\Delta G_{water}-m[D] \qquad (5)$$

where m is a measure of the linear dependence of ΔG on denaturant concentration [D].

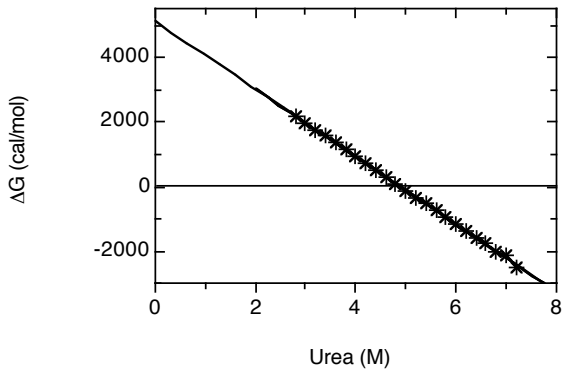

Figure 4. The linear extrapolation method. This figure shows the change in the Gibbs free energy as a function of urea. This graph illustrates the linear dependence of the transition region. The solid line represents the linear fit by equation (5). The stability of this protein is about 5.0 kcal/mol with a C_{mid} of 5 M and an m value of 1 kcal/mol M.

Typically the linear extrapolation method (LEM) is implemented through a more precise analysis as described by Santoro and Bolen. [18] They describe a method by which the transition region is characterized by two parameters m, the dependence of ΔG on denaturant, and [C_{mid}] the midpoint of the transition where ΔG=0. The pre- and post-transition regions are characterized by two parameters θ_N and θ_U , the intercepts of the unfolded and native baselines and a_N and a_U which measures the denaturant dependence of the pre- and post-transition baselines. The following equation represents the curve shown in Figure 3:

$$[\theta_{obs}] = \frac{[\theta_N] + a_N[D]) + ([\theta_U] + a_U[D]) * \exp[m * ([D] - C_{mid})/RT]}{1 + \exp[m * ([D] - C_{mid})/RT]} \qquad (6)$$

The best fit of the six parameters is preformed using a least-squares fit program.

Thermal denaturations

Thermal denaturations are most commonly monitored by CD or by the use of differential scanning calorimetry (DSC). We will first discuss the spectroscopic method of thermal denaturation (van't Hoff analysis). A useful measurement of ΔG requires extrapolating measurements from a narrow temperature range, where unfolding occurs, to a reference temperature such as 25°C. To obtain the enthalpy change (ΔH), the van't

Hoff equation is used:

$$d(\ln K_{eq})/d(1/T) = -\Delta H/R \qquad (7)$$

The van't Hoff plots (ln K_{obs} vs. 1/T) for protein unfolding transitions are usually non-linear, provided that the transition covers a wide temperature range. This indicates that ΔH varies with temperature, which is expected when the heat capacities of the native and unfolded protein differ.

$$d(\Delta H)/d(T) = C_p(U) - C_p(N) = \Delta C_p \qquad (8)$$

In equation (8), $C_p(U)$ and $C_p(N)$ are the heat capacities of the unfolded and native conformations, and ΔC_p is the change in heat capacity upon unfolding. With this in mind, ΔC_p and ΔH are both required to calculate ΔG as a function of temperature. Since ΔH is needed at only one temperature, the best temperature to use is T_m, the midpoint of the thermal unfolding curve where $\Delta G(T_m) = 0 = \Delta H_m - T_m \Delta S_m$. Now with these parameters and the modified Gibbs-Helmholtz equation (9) we can calculate ΔG(T).

$$\Delta G(T) = \Delta H_m (1 - T/T_m) - \Delta C_p [(T_m - T) + \ln(T/T_m)] \qquad (9)$$

We need T_m , ΔC_p and ΔH_m to calculate ΔG(T). The simplest method to determine ΔH is a plot of ΔG vs Temperature. From this plot we get T_m and ΔH_m where T_m is the

temperature at which ΔG is 0 and ΔH_m is the enthalpy at the T_m. Now all we need is to determine ΔC_p.

While there are many methods to determine ΔC_p, a useful technique to determine ΔC_p was described by Pace and Laurents [19]. In this method, ΔG calculated from urea denaturations performed at different temperatures are combined with ΔG values from the transition region of a thermal denaturation unfolding curve. A least squares fit to equation (9) yields ΔH, ΔC_p and T_m. In favorable conditions, where cold denaturation occurs and the maximum stability is greater than 5 °C, the ΔC_p values can be determined with an accuracy of 5 %. If the temperature of maximum stability is below 0 °C, there is greater uncertainty in the measurements.

Differential scanning calorimetry: DSC

The differential scanning calorimeter measures the heat absorption of a sample as a function of temperature. [20; 21] A pair of cells are placed in a thermostated chamber. The sample cell is filled with a protein solution and the reference cell is filled with an identical volume of solvent. The two cells are heated with a constant power input to their heaters during a scan. Any temperature difference between the two cells is monitored with a feedback system so as to increase (or decrease) the input power to the sample cell. Since the masses and volumes of the two cells are matched, the power added or removed by the cell feedback system is a direct measure of the difference between the heat capacity of the sample and reference solutions. The cell feedback power represents the

raw data, expressed in units of cal/min. By knowing the scan rate and the sample concentration, these units are converted to cal/mol-K.

In practice, the sample and reference cells can be slightly mismatched. The usual practice is to record a reference baseline for the experimental scan; this is subtracted from the experimental data to yield C_p vs. T. The peak occurs near T_m. The heat capacity (C_p) is the temperature derivative of the enthalpy function:

$$C_p = (dH/dT)_p \qquad (10)$$

The enthalpy is obtained from a DSC experiment by integration of the heat capacity curve between two temperatures (initial and final):

$$\Delta H_{cal} = \int C_p dT \qquad (11)$$

Multiple unfolding transitions

When the unfolding reaction shows more than one transition, unfolding is more complex than a two-state reaction.[22] This behavior is frequently observed for multi-domain proteins.[23; 24] Observing a single-state unfolding transition does not prove a two-state mechanism; it merely suggests that there are at least two-states. Insight into the folding mechanism can be gained by utilizing different techniques and probes to follow

the unfolding transition. Non-coincidence of plots of fraction unfolded as a function of temperature or denaturant concentration determined by different spectral probes indicates that an intermediate is present and hence a two-state mechanism cannot be used in analysis of the data. However, coincidence of the unfolding data is only support and again does not prove a two-state mechanism.

The best support for two-state thermal unfolding is to show that ΔH_{vH} determined by the van't Hoff relationship is identical to that determined by calorimetry ΔH_{cal} [25]. When $\Delta H_{cal} > \Delta H_{vH}$ it is clear evidence that significant concentrations of intermediates are present at equilibrium. If intermediates occur with similar T_m values, the separate equilibria will result in a broadening of the transition curve; a lower slope on the van't Hoff plot and an underestimate of ΔH for the process, *i.e* $\Delta H_{cal} > \Delta H_{vH}$. In some cases, two separate peaks are seen in the C_p vs. T plot. If the transitions correspond to independent folding of two protein domains, they can often be studied separately.

Measuring the stability of a protein by hydrogen exchange

If the conformational stability of a protein is 5.0 kcal/mol, there will be one unfolded protein molecule and ~ 5000 folded molecules at any one time. This makes measuring the equilibrium between native and unfolded proteins very difficult. In traditional methods, as described above, the conformational stability is extrapolated from stabilities determined under denaturing conditions, where both the folded and unfolded molecules are more evenly populated. The technique of amide hydrogen exchange,

however, allows us to measure this dynamic equilibrium under conditions that are non-denaturating where to folded protein dominates the population.[26; 27] This is a significant advantage over traditional methods such as solvent and thermal denaturations.

All proteins exist in a dynamic equilibrium between the native (closed) form and the unfolded (open) form. When a protein is placed in D_2O the amide hydrogens begin to exchange with deuterium. The rate of this exchange is governed by several factors. Most important to our study is the burial of the amide and/or hydrogen bonding which exists in the structured form. Exchange can occur by local fluctuations or by a global unfolding event.

The exchange process for an individual backbone amide can be described by a general two-step process:

$$\left(\mathrm{N-H\cdots O=C}\right)_{\mathrm{cl}} \underset{\mathrm{k_{cl}}}{\overset{\mathrm{k_{op}}}{\leftrightarrows}} \left(\mathrm{N-H}\right)_{\mathrm{op}} \xrightarrow{\mathrm{k_{rc}}} \mathrm{N-D} \qquad (12)$$

where, k_{op} and k_{cl} are the rate constants for structural opening and closing, and k_{rc} is the rate constant for the intrinsic rate constant of exchange for a random coil. The second step of this reaction is considered irreversible because the protein is in a vast excess of D_2O, thus back exchange is improbable. The measurable rate constant is $k_{ex}=k_{op}k_{rc}/(k_{op}+k_{cl}+k_{rc})$. If we adjust the conditions such that $k_{cl}<<k_{rc}$, then the $k_{ex}=k_{op}$ or we say that we are in EX1 conditions. However, if we adjust the conditions such that

k_{cl}>>k_{rc}, then the k_{ex} =$K_{eq}k_{rc}$, the so-called EX2 limit. Using this relationship we can determine the conformational stability of the U<->N equilibrium.

$$\Delta G_{HD}=-RT\ln(k_{ex}/k_{rc}) \quad (13)$$

It has been shown that the average of ΔG_{HD} for the three most stable residues provides a good estimate of the global conformational stability of a protein.[28] In order to determine the global stability using the hydrogen exchange method, and to compare ΔG_{HD} with ΔG values determined by more traditional methods, it is important to account for several issues: 1) confirm the protein is in EX2 exchange conditions 2) account for isotope effects and 3) account for proline bond isomerization.

The prime determinant for EX2 is that the rate of exchange is at least 10 times slower than that of the rate of closing. Exchange from the open conformation depends on the k_{rc}. The k_{rc} rate constants were determined from short peptides. These studies showed that the exchange rate depends on the pH, temperature, identity of the neighboring side chain residues in the sequence, solvent isotope effects and small ionic strength effects.[29; 30]

The measured hydrogen exchange stabilities of the most stable residues are generally higher than those measured by chemical or thermal denaturation. There are two main reasons for these discrepancies: solvent isotope effects and Xaa-proline bond isomerization. [8]

To achieve agreement between traditional stability measurements and hydrogen deuterium exchange results, corrections for isotope effects and proline isomerization need to be applied. D_2O affects the stability of a protein any where from 0 to 2 kcal/mol compared to normal water.[8; 31] With this in mind, all comparisons with traditional experiments should be performed in D_2O. The denatured states that are sampled under native state conditions in the hydrogen exchange (HX) experiments differ from those found in the equilibrium unfolded forms since in HX experiments the proline residues do not have time to isomerize before the protein refolds. This can be rationalized because the refolding time associated with protein folding is usually on the order of milliseconds to microseconds, whereas proline isomerization is on the order of seconds. Proline isomerization rates are affected by the amino acid in the Xaa-Pro position.[32] This difference in the unfolded states for the different techniques can be corrected for by applying the equilibrium constants for proline isomerization determined from peptide models to that of the protein. These corrections were shown to be adequate to achieve agreement between the conformational stability determined by hydrogen exchange and traditional methods for a variety of different proteins.[8]

Folding kinetics: introduction

In the 1960's, Christian Anfinsen showed that a denatured protein, which had essentially all of its native three dimensional structure disrupted and all the disulfide groups reduced, could refold into a well-defined unique structure in which the biological

activity was completely restored.[33] This study, along with others, lead to the "thermodynamic hypothesis"; that is, the hypothesis that proteins in a normal physiological environment (solvent, pH, ionic strength, presence of components such as metals, or prosthetic groups, temperature etc.) fold into a native conformation that depends only on intra-molecular interactions as designated by its amino acid sequence. This hypothesis assumes that the native state corresponds to the global free energy minimum.

In the late 1960's, Cyrus Levinthal performed a simple calculation to determine how long it might take for a protein to fold into its native structure. In a standard illustration of Levinthal's paradox, if each connection between amino acid residues can have three possible states, so a polypeptide chain of 101 amino acids could exist in 3^{100}= $5 * 10^{47}$ configurations. Even if a protein is capable of sampling 10^{13} configurations per second, it would take 10^{27} years to try them all. Levinthal concluded that proteins must fold by specific pathways.[1; 2] This introduced the hypothesis of "kinetic control" and lead to a search for folding pathways. Here we will limit our discussions to two-state (Model 1), three-state models including on-pathway (Model 2), off-pathway (Model 3) and prolyl peptide bond isomerization (Model 4) mechanisms. While prolyl isomerization is an example of a three state pathway, it is often over looked due to the disparity in rate constants (μs vs. ms).

Three-state (Model 1)

$$U \underset{k_{IU}}{\overset{k_{UI}}{\leftrightarrow}} I \underset{k_{NI}}{\overset{k_{IN}}{\leftrightarrow}} N \qquad (14)$$

On-pathway (Model 2)

$$U \underset{k_{IU}}{\overset{k_{UI}}{\leftrightarrow}} I \underset{k_{NI}}{\overset{k_{IN}}{\leftrightarrow}} N \qquad (15)$$

Off-pathway (Model 3)

$$I \underset{k_{IU}}{\overset{k_{UI}}{\leftrightarrow}} U \underset{k_{NU}}{\overset{k_{UN}}{\leftrightarrow}} N \qquad (16)$$

Prolyl peptide bond isomerization (Model 4)

$$U_{slow} \underset{k_{U_fU_s}}{\overset{k_{U_sU_f}}{\leftrightarrow}} U_{fast} \underset{k_{NU_f}}{\overset{k_{U_fN}}{\leftrightarrow}} N \qquad (17)$$

Protein folding kinetic experiments

This section will discuss the basic methods for measuring protein folding kinetics and analyzing these reactions. We will limit the discussion to reversible, monomeric, unimolecular protein folding reactions. Non-reversible protein folding reactions usually yield aggregated protein, which makes it difficult to determine a folding pathway. Protein folding reactions that are not unimolecular require special techniques and are beyond the scope of this review.

The classical method of sample mixing has been used in biochemistry for over sixty years. It involves the rapid mixing of two solutions (in our case a protein and denaturant) which streams into an optical cell. The most commonly used fast mixing technique is the "stopped flow" method in which data collection begins after the flow through the cell is abruptly stopped.[22] The protein starts in one set of conditions and is rapidly changed to another, shifting the equilibrium between Native (N) and Unfolded (U) and the change in a spectroscopic probe is monitored as a function of time. An example refolding experiment is given in Figure 5. Figure 5 shows a simulated stop flow experiment of protein that shows a two-state folding mechanism. The figure illustrates the "dead time", the denatured signal, and the native signal.

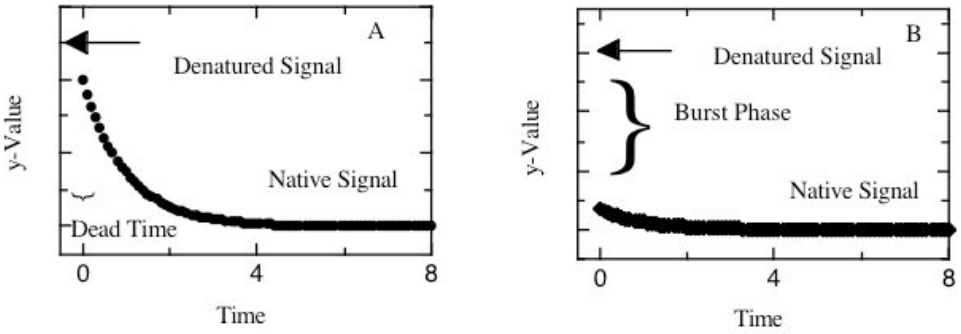

Figure 5. Simulated rapid mixing experiment. Protein under denaturing conditions is rapidly mixed to dilute the denaturant and folding is initiated. The reaction can be monitored by any spectroscopic method (e.g. NMR, Fluorescence, FTIR, CD, or absorbance.) Panel B shows an experiment with a dead time of ~10 ms and the presence of a Burst Phase signal, where the exponential fit of the data does not extrapolate to the denatured signal at time =0.

A protein folding kinetic experiment will tell: 1) the number of kinetic phases (1 to *n*) , 2) the amplitudes of each phase (A_i), 3) the time constants (τ_i) of each phase and 4) the signal at infinite time A(∞).

$$A(t) = \sum_{i-1}^{n} A_i \exp(-t/\tau_i) + A(\infty) \qquad (18)$$

A(t) is the signal followed by the spectroscopic probe as a function of time and A(∞) is the final signal when equilibrium is reached. All of these parameters are dependent on the probe, except τ or the relaxation time, which is probe independent because it is only a function of the rate constant.

$$\tau = \frac{1}{k_{obs}} = \frac{1}{(k_f + k_u)} \qquad (19)$$

However, τ will be affected by the solution conditions and we can manipulate this. A description of the kinetic mechanism begins with determination of the number of kinetic phases and the A_i or amplitude(s) of each phase. A folding reaction with multiple detectable phases will not fit to equation (18) with i=1. As an example, if there are two kinetic phases one would expect:

$$A(t) = A_1 \exp(-t/\tau_1) + A_2 \exp(-t/\tau_2) + A(\infty) \qquad (20)$$

The amplitudes are A_1 and A_2 and the relaxation times are τ_1 and τ_2. However, if your reaction fails to detect two phases, the use of multiple probes is useful in determining the mechanism. In sequential folding mechanisms, it is expected that the folding rates determined by two different probes will have similar relaxation rates, assuming that each probe can detect the same reaction. The use of multiple probes has been useful in determining folding pathways[34] for cytochrome c[35; 36; 37], ribonuclease A[38; 39] and lysozyme.[40; 41]

A common observation in refolding experiments is the presence of a signal in the dead time, a "burst phase intermediate", resulting from processes faster than the

detection limit of the instrument.[42; 43; 44] It is normally assumed that this fast phase represents the formation of a specific intermediate. Panel B in Figure 5 shows a simulated refolding experiment where the burst phase intermediate is present.

Amplitude analysis

The amplitude analysis relies on the fact that the relaxation times for the reactions depend only on the final solution conditions, whereas the amplitudes depend on both the final and initial conditions. To test the possibility of two folding phases, samples are equilibrated at different starting conditions where the protein is either in the native or unfolded form. The unfolding (refolding) reactions are initiated by rapidly changing the conditions by dilution with refolding or denaturant buffer. The resulting curves are fit to equations 18 or 20 and the relaxation times and amplitudes determined. Usually, the amplitudes are expressed as fractional amplitudes;

$$\alpha_i = A_i / [A(0) - A(\infty)] \qquad (21)$$

The denaturant dependence of the relaxation times and amplitudes are monitored by jumping to different final solution conditions. The amplitudes of the kinetic phases are proportional to the population of the species involved and the urea dependence can serve to discriminate between a two-state folding mechanism and a multiple-state mechanism. For example, a decrease in the amplitude for the initial, presumably faster event, and a

concomitant increase in the slower process would be consistent with a burst phase (Figure 6).

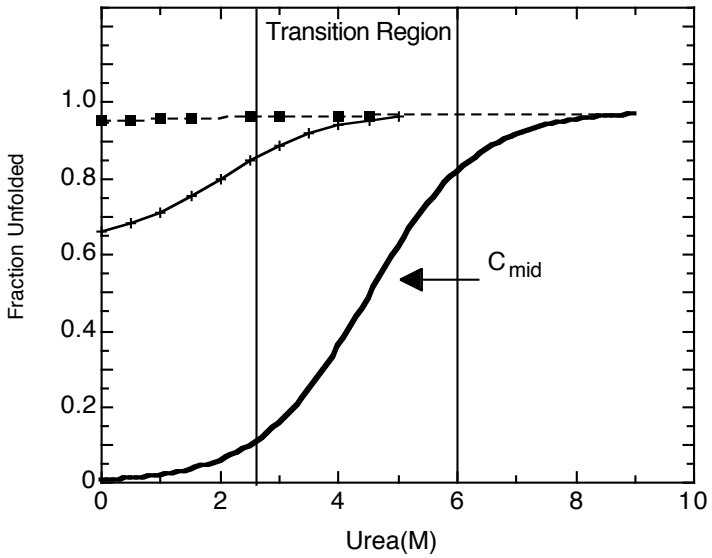

Figure 6. Simulated data of an amplitude analysis. The simulated data shows the urea dependence amplitude (α) of a refolding experiment associated with a two-state model (filled square dotted line) and a burst phase event (solid line and cross). This figure shows the equilibrium C_{mid} as well as points out the regions of the graph that are considered the transition region for the equilibrium transition.

Chevron analysis

The chevron analysis requires the complete analysis of the denaturant dependence of the refolding and unfolding rates as well as their amplitudes.[45; 46] The logarithms of the reciprocal of the relaxation time (the observed rate constants) are plotted as a function of the denaturant concentration. For a simple folding reaction, the

plot resembles a V shape and has therefore been called a chevron plot. This V shape reflects the acceleration of the unfolding reaction at high denaturant concentration and the acceleration of the refolding rate constant at low denaturant concentration.

The rate constants for the two-state mechanisms are k_f (refolding) and k_u (unfolding). The denaturant dependence to the observed kinetics can be described by equation (22):

$$\ln k_{obs} = \ln[k_f^{water} \exp(\frac{-m_f[D]}{RT}) + k_u^{water} \exp(\frac{m_u[D]}{RT})] \qquad (22)$$

where D is the denaturant concentration, m_f ,m_u describe the denaturant dependence of the refolding and unfolding reactions and k_f and k_u are the extrapolated rate constants in the absence of denaturant (Figure 7).

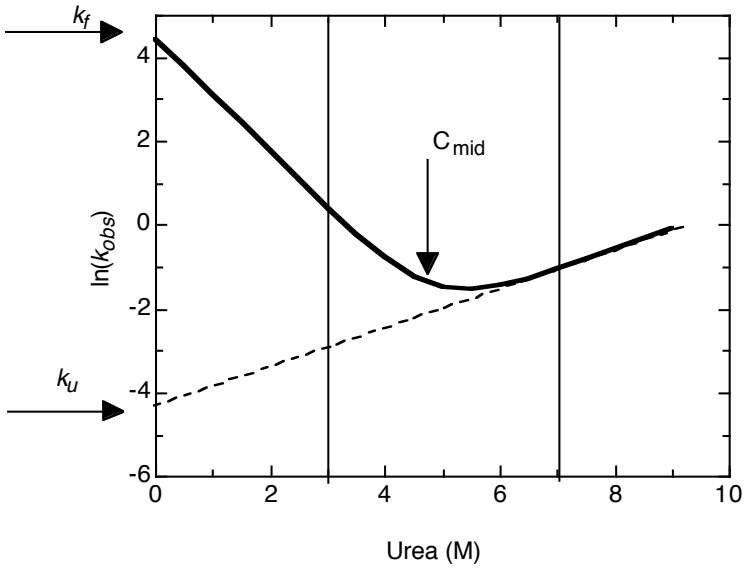

Figure 7. Simulated chevron plot showing the linear dependence of folding and refolding on denaturant concentration. The arrows show the k_f , k_u and C_{mid} for this reaction.

Thermodynamic parameters can be calculated from these kinetic parameters. ΔG_{water} can be calculated for a simple two-state mechanisms with the equilibrium constant for unfolding as discussed below:

$$K_U = \frac{[U]}{[N]} = \frac{k_f}{k_u} \qquad (23)$$

Similarly, the m values can be calculated from the summation of the absolute values of

$$m_{eq} = m_f + m_u \qquad (24)$$

The concentration of denaturant at which 50% of the protein is denatured, C_{mid} can be calculated using

$$C_{mid} = \frac{\ln(k_f^{water} / k_u^{water})}{(m_f + m_u)} \qquad (25)$$

The chevron analysis can provide several useful insights into the folding mechanism. First, if the unfolding and refolding reactions meet smoothly in the center of the chevron, the folding reaction is completely reversible. Second, the denaturant dependences of the folding and unfolding rate constants provide information about the structural changes associated with burial of surface area. Third, the values determined with the chevron can be compared with equilibrium values to determine if identity is achieved; any deviation suggests the presence of an intermediate.

Two-state mechanism

Strictly speaking, the two-state mechanism requires that the folding and unfolding events are discrete, first order reactions with k_{NU} and k_{UN} rate constants.

$$U \underset{k_{UN}}{\overset{k_{NU}}{\leftrightarrow}} N \qquad (26)$$

Several simple tests are used to test the two-state assumption. [47; 48; 49; 50] For example, within the transition region, both the folding and unfolding reactions can be followed under the same conditions and should yield the same τ. Folding and unfolding rates should also be independent of the probe used to monitor the reactions. These criteria are useful for validating the two-state assumption. There are three traditional kinetic tests to determine if a protein folds using a kinetic two-state mechanism. The first test would be to compare the ratio of the forward and reverse rate constant (equation 23). For a two-state mechanism, the ratio must give the same equilibrium constant as that measured by an equilibrium method under the same conditions.

Test two is to show that the sum of the forward and reverse m values equal the m value for equilibrium unfolding for a two-state transition, where all three are algebraically positive (equation 24).

The third test looks at deviations from the simple V shaped curves of rate constants versus urea that are shown in Figure 7. Often, three-state folding mechanisms as described by models 2 and 3 give rise to roll over. Roll over is the non-linear urea dependence in the refolding limb of the chevron plot. Several explanations for roll over have been used to explain two-state folding mechanisms including the Hammond effect[48], on-pathway[43;40;44;45], off-pathway[40;44;45] and parallel pathway folding[40] mechanisms. The lack of roll over does not disprove a three-state mechanism.

Three-state mechanism

Folding intermediates often accumulate rapidly when refolding is initiated. With the presence of apparent two-state folding proteins, one pressing issue is whether these folding intermediates are on-pathway or off-pathway.[48]

On-pathway

$$U \underset{k_{IU}}{\overset{k_{UI}}{\leftrightarrow}} I \underset{k_{NI}}{\overset{k_{IN}}{\leftrightarrow}} N \qquad (27)$$

Off-pathway

$$I \underset{k_{IU}}{\overset{k_{UI}}{\leftrightarrow}} U \underset{k_{NU}}{\overset{k_{UN}}{\leftrightarrow}} N \qquad (28)$$

Folding intermediates are most commonly identified as either multiple phases in the refolding reaction, a loss of amplitude during the dead time of the experiment (Burst Phase) and as roll over.

Standard kinetic analysis shows that when equilibrium between two-states proceeds the formation of the third-state, the rate of formation of the third state is the proportion of f_I to, the fractional concentration of I, for the on-pathway model.

$$k_f = f_I * k_{IN} \quad (29)$$

Conversely, k_f is proportional to (1-f_I) in the off-pathway model.

$$k_f = (1 - f_I) * k_{UN} \quad (30)$$

$$f_I = [I]/([I] + [U]) = K_{UI}/(1 + K_{UI}) \quad (31)$$

The accumulation of I has opposite effects in the two models. The accumulation of [I] in the on pathway model speeds up the formation of [U], while the accumulation of [I] in the off pathway model slows down the accumulation of [U]. The accumulation of [I] leads to an event called roll over.

Thermodynamic parameters can be calculated from these kinetic parameters. ΔG_{water} can be calculated from the equilibrium constant for unfolding:

$$K_{eq} = \frac{k_{UI}}{k_{IU}} * \frac{k_{IN}}{k_{NI}} \quad (32)$$

Similarly, the m values can be calculated from the summation of the absolute values of

m_{IN} , m_{NI} and m_I:

$$m_{eq} = m_{IN} + m_{NI} + m_I$$
$$m_I = m_{IU} + m_{UI} \qquad (33)$$

Roll over

Roll over is the apparent loss of linearity seen in the refolding limb of the chevron plot. Several explanations have been offered to explain nonlinear "curved" chevron plots. Intermediates can give rise to roll over by their effect on the fractional population of the intermediates, essentially slowing down the observed reaction rate (equations 29, 30). The Hammond postulate has also been offered as an explanation for curved chevron plots (discussed below).[51]

The formation of intermediates has been associated with nonlinear chevron plots.[48; 52] The k_f rate falls below the expected linear dependence on denaturant concentration. This is directly related to the rate at which N is formed from I (equation 31). The off pathway explanation is that the slowing down of the folding rate, compared to expected, and is caused by the accumulation of I (Figure 8). The lines in Figure 8 (dotted and dashed) intersect; and at this point, k_{NI} has a large effect on the available population of N, thus decreasing k_{obs} and creating roll over.

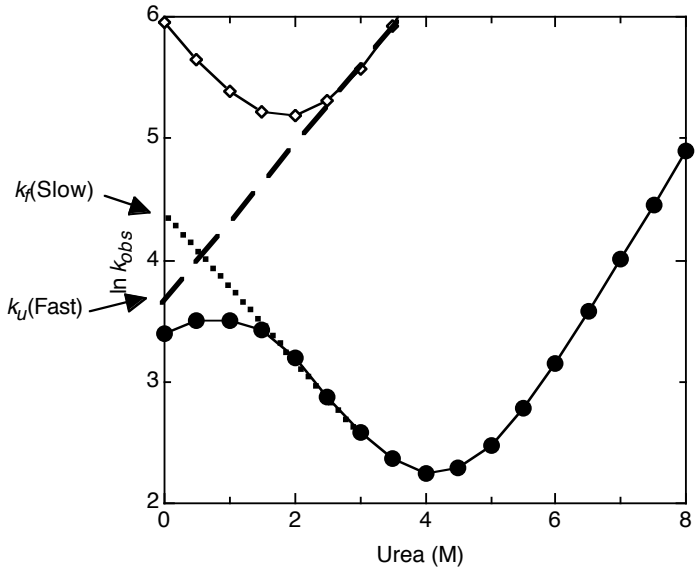

Figure 8. Simulated kinetic data showing roll over. The open diamonds represents the fast folding reaction I->N this reaction is not observed in the folding reaction. The dashed line shows the predicted k_u (fast) unfolding rate constants urea dependence. The dotted line shows predicted the k_f (slow) refolding rate constants urea dependence.

On-pathway intermediates can give rise to roll over by increased accumulation of I (equation 30) and this accumulation leads to an decrease in the observable rate constant. The determination of on-pathway vs. off-pathway intermediates requires the ability to selectively perturb these equilibriums and thus affect the observable rates.

The third explanation given for non linear chevron plots has to do with the Hammond postulate. [44; 53] The Hammond postulate states that the transition state of a protein folding reaction will be affected by the energetic changes applied by the denaturant used to initiate the folding or unfolding reaction. The reasoning is that the denaturant will affect the transition state structure thus changing its energy. A simple

example is illustrated in Figure 9. The folded and unfolded protein is represented by their respective energy wells. The transition state is represented by the intersection of these two wells. In the presence of denaturant the energy of one of the states is modified and the energy of the transition state is subsequently modified. The movement along the reaction coordinate is known as the Hammond effect and can give rise to a curved chevron plot.

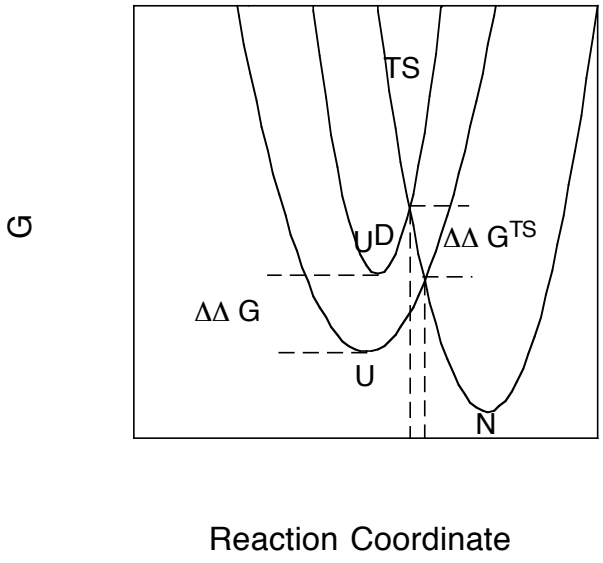

Figure 9. An illustration of the Hammond effect. The Gibbs free energy plots of the unfolded protein U, unfolded destabilized protein U^D, and native state N. The intersection of these energy wells represents the transition state for the folding reaction. The change of structure of the unfolded state (U to U^D) leads a change in the position of the transition state.

Prolyl peptide bonds isomerization in protein folding

Brandts first suggested that the slow refolding phases observed in many protein refolding experiments could be explained by incorrect isomers of Xaa-Pro peptide bond in the unfolded state.[54] Peptide bonds are planar and can be either in the *trans* or *cis* conformation. Most peptide bonds prefer the *trans* conformer in linear peptides and less than 0.1% assumed the *cis* conformer. However, the peptide bond between proline and its proceeding amino acid (Xaa-Pro) typically exists as a mixture of *cis* and *trans* isomers in solution, unless structural constraints prevent stabilization of one of the isomers. Figure 10 shows the alternative isomeric states of the Xaa-Pro peptide bond.

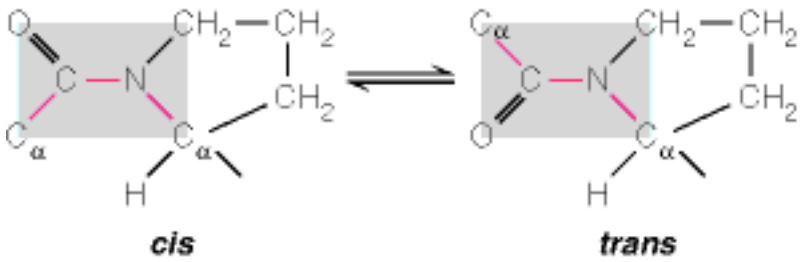

Figure 10. Alternative isomeric states of the Xaa-Pro peptide bond.

The *cis* content is typically in the range of 10% to 30% in unfolded polypeptides and the *cis* to *trans* isomerization rate is very slow (10 to 100 sec at 25°C). The equilibrium of this inter-conversion is strongly influenced by the identity of the residue preceding the Pro (Xaa-Pro).[32]

The role of prolyl peptide bond isomerization in protein folding can be elucidated through three experimental observations. First, the fraction of the slow refolding population is determined by the number of proline residues and on their isomeric states in the native protein. In particular, the presence of *cis* proline isomers in the folded state leads to a high population of slow folding species. This is because the *trans* population is favored in the absence of structural constraints. Second, prolyl peptide isomerization is independent of additives such as urea. Third, the refolding of the slow state molecules can be affected by the coupling of structural refolding and bond isomerization.[55]

Double jump

A commonly used technique to determine the presence and rate constants of an intermediate is the double jump technique. [56]As discussed earlier the Xaa-proline bond has an equilibrium of ~4:1 in the unfolded conformation since the two configurations are not very energetically different, whereas the Xaa-Proline bond in the folded conformation is fixed to either *trans* or *cis*. As a protein unfolds, the proline bond becomes free to isomerize. Initially all the prolines will remain in the native configurations, and the protein is in a "fast folding form". Upon refolding, the protein has no slow bond isomerization (as seen in native state hydrogen exchange). However, for long denaturation times, the isomerization equilibrium will be reached, and a slow refolding species will be populated. Refolding would then yield a "slow folding" species as the prolines must isomerize to achieve the native conformation.

Figure 11. An outline of a typical double jump experiment. Jump 1 starts the unfolding reaction followed by a delay time that is varied from experiment to experiment followed by a second jump with returns the protein to the native state.

One can unfold the native protein to populate the unfolded non-isomerized species. Over the delay time the prolines will isomerize according to their rate constant until equilibrium is achieved. Renaturation to the native state is then monitored in the second jump Figure 11. When the fast and slow refolding species are populated, the refolding becomes biphasic. The amplitudes of the two refolding reactions are directly related to the populations of the fast and slow reactions and thus a graph of delay time vs amplitude can be fit to determine the rate constant for Xaa-Pro isomerization.

Case studies

Good models for protein folding studies should have reversible unfolding at a wide variety of conditions. The amino acid sequence and three dimensional structure should be known, and the gene and protein should be available for mutational studies in order to determine residue-specific effects.

Cold-shock protein B (CspB) two-state folding

CspB is a 67 amino acid residue protein from *Bacillus subtilis*. Its tertiary structure consists of a single five-stranded β-barrel. CspB does not contain any ligands, disulfide bonds, or any cis-prolyl bonds. Its folding is reversible and both its folding and unfolding are rapid. Equilibration occurs in less than 100 ms at the transition midpoint, where folding is slowest. Schindler et al used a series of kinetic tests including amplitude analysis, chevron analysis, and double jumps in order to establish the two-state mechanism.[57] The most remarkable test was the use of a pressure jump technique. The use of the pressure jump technique allows for the measurement of relaxation times in the 100's of microseconds. With the use of this technique, they were able to show the rapid unfolding and refolding of a protein clearly follows a two-state mechanism. While this was not the first protein exhibiting a two-state folding mechanism, it certainly shows the strongest evidence to date.

The establishment of a two-state folding mechanism raised serious questions as to whether proteins with multiple folding states folded on an off-pathway or on-pathway mechanism.

Barnase and Im7 (on-pathway vs. off-pathway)

Some computational studies on small proteins have suggested that protein folding intermediates result from kinetic traps that are off-pathway and these species accumulate to detectable levels. However, other folding simulations suggest that folding occurs rapidly through on-pathway intermediates. This controversy is an old one that probably started when Levinthal proposed that the only way a protein could fold in a timely manner is through defined folding intermediates. One of the first proteins that revealed the possibility of off-pathway intermediates was Barnase. In this study, they noticed that Barnase does not follow the two-state chevron formalism but rather exhibited what would later become known as roll over.[44; 52] They concluded that the slowing down of the folding rate, compared to that expected, could have been caused by the accumulation of an intermediate (on-pathway/off-pathway equations 27, 28) or by a change in the reaction mechanism to give a different rate-limiting step (Hammond effect). It is impossible to determine by conventional kinetic methods, whether an intermediate is on-pathway or off-pathway.

The situation changes if the kinetics of folding and unfolding of intermediates are measurable. The introduction of sub-millisecond measuring mixing devices has made it possible to measure the folding and unfolding kinetics of sub-millisecond intermediates in some protein folding reactions.[58] The Roder laboratory has compared a traditional stop flow instrumentation (dead time ~ 5 ms) and a continuous flow instrumentation to measure folding rates in the microseconds (~ 100 μs). [53] These data revealed a distinct

kinetic phase on the 100 μs time scale that resembles the formation of the intermediate ensemble that was distinct from the folded and unfolded states. Importantly, the rate constants for the intermediates formation were only 10-fold greater than the subsequent phase. This allowed the authors to rule out the presence of a pre-equilibrium step, thus eliminating the possibility of an off-pathway intermediate. They followed this work with kinetic modeling that shows that the data are best described by a model that included an on-pathway intermediate. [53]

The older concepts of "folding pathway" and "productive folding intermediates" remain useful concepts today. Larger proteins often have competing parallel pathways much like that seen with RNase T1.

RNase T1 (prolyl isomerization)

This case study on RNase T1 is the paradigm of Model 4, the prolyl isomerization folding mechanism.[55; 59; 60; 61] Ribonuclease T1 (RNase T1) is a small single domain protein of 104 amino acids with an extended α-helix of 4.5 turns and two antiparallel β sheets. The protein has two disulfide bonds linking residues 2-10 and 6-103. The protein contains four proline peptide bonds; two are *trans* (Trp59-Pro60 and Ser72-Pro73) and two are *cis* (Try38-Pro39 and Ser54-Pro55) in the native protein. The unfolding is reversible under many conditions and is well described by a two-state mechanism under equilibrium conditions. RNase T1 has been shown to refold to a native-like catalytically active form in the absence of disulfide bonds.

The refolding kinetics are dominated by three slow folding processes. [50] Both phases originate from the unfolded protein and are catalyzed by PPI, a prolyl isomerase. RNase T1 has four parallel folding pathways. [55] About 3.5% of the unfolded population refolds in the dead time of the experiment. The slow refolding kinetics are composed of a sequential two step reaction (with time constants of 190 sec and 500 sec) and a very slow reaction with a time constant of 3000 sec. [56] Structural intermediates form very rapidly within the dead time of the experiment, as shown by CD of the 20 sec burst phase intermediate. The transient formation of these intermediates was determined by the double jump technique. [50]

The kinetic mechanism proposed by Schmid et al is represented in Figure 12.[59] About 3.5% of the isomers are in the correct conformation and fold in less than 20 sec (the dead time of manual mixing experiment). Three slow refolding phases were measured corresponding to two prolines in the incorrect conformation (39 *trans* 55 *trans*) and two prolines in the incorrect conformation (39 *cis* 55 *trans*, 39 *trans* 55 *cis*). In this model, it is suggested that all of these U states can regain some conformation rapidly U->I. The slow, rate-limiting steps are then caused by the isomerization of the proline peptide bond. The major unfolding state with two *trans* conformers can go into either refolding pathway. The refolding kinetics of RNase T1 are dominated by the proline isomerization seen in model 4, where folding and isomerization are interrelated. In all pathways, native-like conformations formed rapidly, followed by slow proline isomerization events.

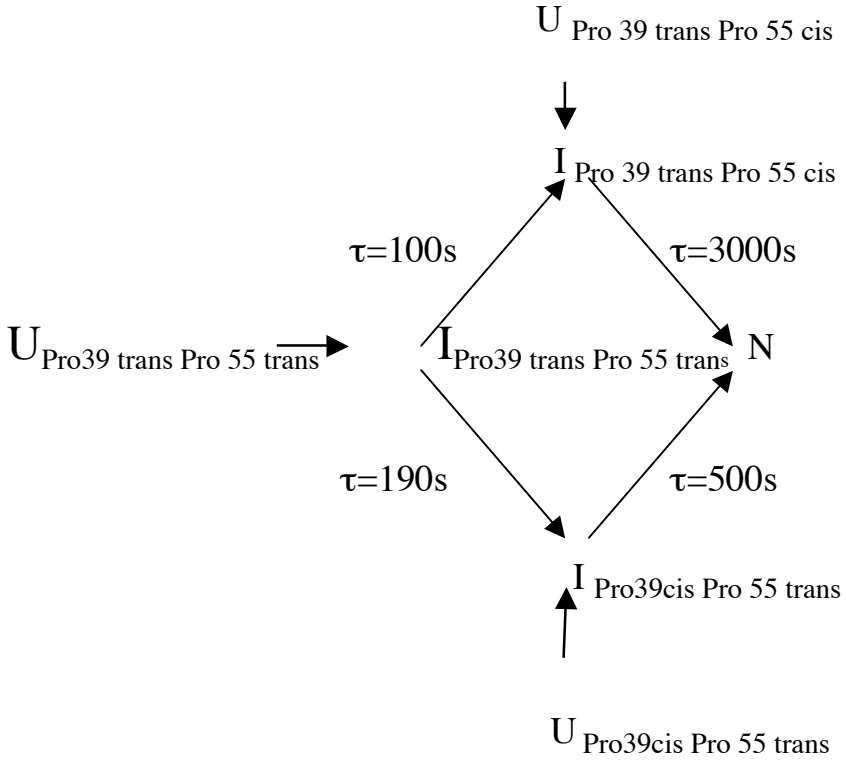

Figure 12. Kinetic model for the refolding reactions of RNase T1.[60] U stands for unfolded species, I stand for intermediates and N is the native protein. The proline designated in this figure has assumed alternate isomers in the unfolded protein. The time constants given in this figure represent the folding conditions of 0.15 M GdnHCl, pH 8.0 at 10°C. About 3.5% of all unfolded molecules have both correct isomers and refold rapidly (in 20 sec dead time for manual mixing).

U1A and S6 the Hammond postulate

The human spliceosomal protein U1A is a 102 residue protein containing no disulphide bonds and no *cis* prolyl peptide bonds. The chevron plot for U1A is symmetrically curved like a bell curve rather than a V shape like most two-state proteins.[53] To determine if curved chevron plots are general phenomena or if it is something specific to U1A and Chymotrypsin Inhibitor 2 (CI2)[48]. CI2 is a well-studied protein and exhibits two-state folding. Several CI2 mutants exhibit the bell shaped curvature similar to U1A. As a further test, the authors utilized a structural analog of U1A, S6 a ribosomal protein from *Thermus thermophilius*. [48] S6 is a 101 residue protein with no disulphide bonds nor any *cis* prolyl peptide bonds. The wt S6 protein exhibited a V shaped chevron. With the use of mutation studies, the S6 protein was able to generate curvature in the unfolding limb of the chevron plot. The authors extensively eliminated the possibilities of an unfolding intermediate. They then applied a Brönsted analysis where they calculated the change in accessible surface area of the unfolding reaction and showed that the transition states structure could be changing. [48]

Our model system HPr

The histidine containing protein (HPr) is a small globular protein that functions as a phosphocarrier protein of the bacterial phosphoenolpyruvate dependent sugar phosphotransferase system (PTS).[62] The purpose of the PTS system is to translocate sugar across the cell membrane and then phosphorylate the sugar. HPr is one of several proteins involved in this process, serving as a phosphocarrier between enzymes I and the sugar-specific permease enzyme II. The phosphryl transfer occurs at the active site His 15 of HPr by enzyme I. More detailed reviews of the PTS system are available.[62]

HPr is a monomer in solution and contains no disulphide bonds and no *cis* prolyl bonds. We work with two HPr proteins, one from *Escherichia coli* (*ec*HPr) and one from *Bacillus subtilis* (*bs*HPr). These proteins have less than 35% identity, yet have homologous folds. The NMR and crystal structures of these proteins have been determined. [63; 64; 65; 66; 67; 68; 69] The structures reveal a fold consisting of three α-helices, referred to as A, B, C and four β strands (Figure 13). This type of fold is commonly referred to as an open-faced β sandwich.

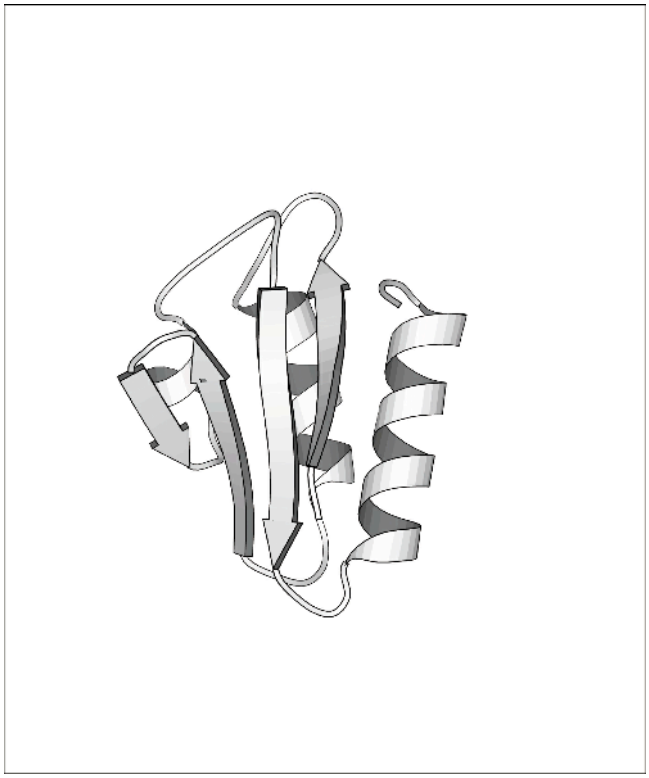

Figure 13. A ribbon diagram of the *bs*HPr protein.

The thermodynamics of *ec*HPr [13 ; 16 ; 70; 71; 72] and *bs*HPr [14; 71]have been extensively studied. The equilibrium stability has been characterized by CD, ANS fluorescence, DSC, and NMR hydrogen exchange. In equilibrium experiments, these proteins have been described as two-state. The folding kinetics of *ec*HPr was previously determined by ANS fluorescence.[16]

CHAPTER II

MATERIALS AND METHODS

MUTAGENESIS

The *E. coli* strains (ES7R or ESK108) were utilized for the expression and purification of the HPr proteins. These cell lines have the endogenous HPr protein removed and replaced by tetracycline resistance. The plasmid used for HPr expression is a pUC-based vector with the HPr gene and promoter inserted in the PUC 19 multi-cloning region.[73] The *B. subtilis* gene has been cloned into this vector behind the *E. coli* promoter. All variants used here were produced by site-directed mutagenesis using the quick-change mutagenesis kit from Stratagene.[74]

The mutagenesis technique that was utilized is a variation of the overlap extension method using polymerase chain reaction (PCR). For our purposes, two complementary primers, mutagenic primers containing at most three mismatched base pairs between primer and target DNA were designed such that the melting temperature is at or near 60°C when determined by T_m(°C)=4°C(GC)+2°C (AT), where GC is the number of G-C base pairs and A-T base pairs formed when the oligo-nucleotide anneals to the target DNA. These primers are used in a standard PCR reaction using PFU, a high

fidelity polymerase, to amplify the entire plasmid. We utilized a PCR reaction consisting of an initial denaturation of 5 min at 95°C, followed by a repeating cycles of denaturation annealing and extension for 20 cycles. The parameters associated with the amplification cycles are 30 sec denaturation at 95°C, then an annealing cycle for 30 sec at the determined T_m -5°C, followed by an extension at 68°C for 6 minutes (1 min for every 1000 bases of the plasmid). After 20 cycles, the reaction was completed with a 5 min extension at 68°C (Figure 14).

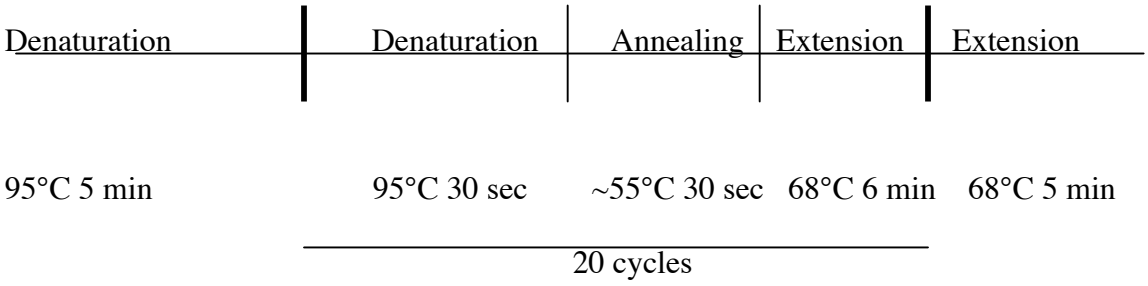

Figure 14. Outline of the amplification procedure. This diagram shows the three stages in the PCR reaction. Of note is the extension phase that contains a 6 min extension time. This is to accommodate the amplification of the entire plasmid which is 6 kb.

A sample of the reaction was then treated with DPN1, an enzyme which digests methylated DNA. This is critical; PCR does not generate methylated DNA. This allows for a selection of mutant plasmid. Following the treatment of the PCR reaction with DPN1, the reaction is transformed into XL1Blue cells and plated on LB amp plates. The *ec*HPr plasmid contains ampicillin resistance. The selection eliminates any *E. coli* that have not taken up the plasmid and successfully repaired the unligated ends. Plasmid

from *E. coli* that survived the selection are then purified and sequenced to confirm the mutation has been successfully incorporated into the plasmid.

Protein expression and purification

The *E. coli* strain ESK108 and ES7R lack endogenous HPr. Thus, those cells transformed with a plasmid expressing an *ec*HPr variant will only produce that form of HPr eliminating the possibility of wild-type contamination. The cells are grown to late log phase in terrific broth and then harvested.[75] For the *ec*HPr protein, the purification is a modification of that described by Anderson et al.[76] Following the cell harvest by centrifugation, the cells are disrupted in a French press cell at 12,000 psi. The disrupted cells are then centrifuged at 30,000 x g for 20 minutes to remove cell membranes and other insoluble material. The supernatant is then dialyzed against 10 mM Tris, 0.1 mM EDTA, pH 8.0. The supernatant is then loaded on a Q-Sepharose column (100 mL bed volume) equilibrated in the same buffer and eluted with a linear salt gradient from 0-0.25 M NaCl with a total volume of 1 L. *ec*HPr containing fractions are identified by SDS PAGE, pooled and lyophilized. The protein is then desalted and further purified on a G-25 sephadex column in 50 mM ammonium bicarbonate buffer. Fractions containing pure *ec*HPr, as determined by SDS PAGE, or mass spectroscopy are lyophilized and then stored at -20°C for later use.

The *B. subtilis* HPr (*bs*HPr) protein purification procedure differs somewhat from that of the *ec*HPr protein. The buffer used is 10 mM Tris, 0.1 mM EDTA, pH 7.6 and

the ion exchange column is DEAE Sephacel. The column is eluted with a 0-0.25 M NaCl gradient with a total 1 L volume. *bs*HPr containing samples are identified as before and loaded on a G-25 column in 50 mM ammonium bicarbonate buffer. At this point the fractions are treated as described for the *ec*HPr purification.

The mutants containing the variant D69X, when expressed, are insoluble and are packed in inclusion bodies. Cells were grown and harvested in the same manner as for soluble proteins with the exception that the large scale growth was done at 42°C. This is done to help partition the protein into inclusion bodies. The cells were harvested and French pressed as previously described. The cells were then centrifuged at 10,000 x g for 15 min. The supernatant was discarded and the pellet was solublized by adding ~100 mL of 2 M guanidine hydrochloride to a 6 L terrific broth preparation. The sample was then dialyzed against 100 mM ammonium bicarbonate. The solution was then centrifuged at 30,000 x g for 30 min to remove an insoluble material. The solution was then loaded on a G-25 column and fractions are collected and analyzed for purity by SDS PAGE. The procedure generated highly pure protein.

Equilibrium unfolding curves

Circular dichroism at 222 nm was used to monitor the equilibrium unfolding transition using either an Aviv 62DS or Aviv 202SF spectropolarimeter equipped with a temperature control and stirring unit. Thermal unfolding curves were performed in 10 mM potassium phosphate at pH 7.0 with heating rates from 30 to 90 °C per hour in

cuvettes with path lengths of 1 cm. The reversibility of the thermal unfolding transition was determined by monitoring the return of CD signal at 222 nm upon cooling from 90 to 10 °C immediately following the thermal transition. In all cases, the transitions were independent of heating rate and independent of protein concentration. Some aggregation occurred upon prolonged exposure to high temperatures.

For urea denaturation curves, the urea solutions were prepared fresh daily in buffered solutions containing 10 mM potassium phosphate at pH 7.0. When the effects of pH were evaluated the buffers used were a combination of sodium phosphate, citrate, and borate at a concentration of 10 mM. When the effects of salts were evaluated the buffer system was MOPS at 30 mM, pH 7.0. The concentration of the urea stock solution was determined by refractive index measurements.[77] Since the folding reaction is rapid (less than 2 min), the urea denaturations were preformed using a titration system. In this case, the urea and buffer solutions contain the same amount of protein. A small aliquot is removed from the cuvette and replaced with a high concentration of denaturant effectively changing the denaturant concentration. The sample is allowed to achieve equilibrium and the CD signal is determined.

For the analysis of unfolding curves, a modification of the linear extrapolation method of Santoro and Bolen [18] was used employing the curve fitting method in KaleidaGraph software.[78] Urea denaturation, with an observed signal θ_{obs} can be analyzed using equation:

$$[\theta_{obs}] = \frac{[\theta_N] + a_N[D]) + ([\theta_D] + a_D[D]) * \exp[m * ([D] - C_{mid})/RT]}{1 + \exp[m * ([D] - C_{mid})/RT]} \qquad (34)$$

θ_{obs} is the spectroscopic signal which was used to follow the unfolding reaction, D is the denaturant concentration using either urea or GdnHCl, a_N and a_D represent the denaturant dependence of the pre- and post-transition base lines, *m* is a measure of the dependence of the ΔG_{UN} on denaturant concentration, and C_{mid} is the midpoint of the denaturation curve. An estimate of the conformational stability in water (ΔG_{UN}) can be determined by multiplying the C_{mid} times the m value. This assumes a two-state relationship where the transition region can be extrapolated to 0 M urea.

The analyses of the thermal denaturation are based on another linear relationship, a variation of the van't Hoff equation.

$$\theta_{obs} = ((N_o + a_N(T)) + (D_o + a_D(T)) * \exp(\frac{\Delta H_{VH}}{RT}(T - T_m)))/(1 + \exp(\frac{\Delta H_{VH}}{RT}(T - T_m)) \qquad (35)$$

Equation (35) is a variation of the van't Hoff equation that we use to describe the two-state, thermally induced unfolding reaction.[79] θ_{obs} is the spectroscopic signal that was used to follow the unfolding reaction, a_N and a_D represent the unfolding dependence of the pre- and post-transition baselines, and N_o and D_o are the intercepts of the native and unfolded baselines. We also obtain the thermal midpoint (T_m) of the unfolding curve and the van't Hoff enthalpy change (ΔH_{vH}). This linear relationship can be applied to the

two-state approximation and linear pre- and post- transition baselines in a manner analogous to that employed for urea unfolding.

NMR hydrogen exchange

Samples were exchanged using a spin column method[80], rather than dissolving lyophilized protein directly into D_2O. This is necessary because it is uncertain if the protein, in the lyophilized state, was in a native or unfolded form and direct addition of D_2O could cause the measurements to be incorrect. The procedure of preparing the spin column begins with presoaking 1 g of G-25 resin in 10 mL of D_2O for at least three hours. The D_2O is drawn off and replaced with the buffer in which you are going to perform the experiment. The resin should equilibrate for at least 20 min. The resin is then washed a second time in the experimental buffer. This mixture is then poured into a spin column and set in a 15 mL conical tube and spun at 1000 x g (≈ 3000 rpm) for one minute in a clinical centrifuge. The best results were obtained with a swinging bucket rotor. The choice of time and velocity vary from column to column. The conditions were selected such that the volume of the solution added was the same as the volume displaced from the spin column. After the column had been packed, it was washed three times. The protein sample, which consisted of 10-25 mg of protein in ~400 μL of buffer in H_2O, at the experimental pH, was added to the resin and spun. The resulting flow through was collected and used for the NMR experiments. This is considered to be time 0 when the sample it initially added to the column. This procedure allowed for

approximately 80 percent of the initial protein stock to be recovered and used for the experiment.

Monitoring the exchange process was done by collection of two-dimensional sensitivity enhanced (^{1}H, ^{15}N) HSQC spectra at regular intervals over the course of the exchange process.[71] These were acquired on a Varian Unity, Varian Unity Plus or Varian Inova-class NMR spectrometer operating at 500 or 600 MHz proton frequencies. Spectra were obtained using a ^{1}H sweep width of 6410 Hz, ^{15}N sweep width was 100 Hz, with a decoupler offset of 1034 Hz. The length of the experiment was 10 min for the initial experiments and 30 min for the latter ones. This was achieved by increasing the number of transients averaged per free induction decay.

All files were processed and peak picking was performed using the nmrPipe processing software. [81] The residue specific decay curves were determined using the methods and programs written by Ron Peterson. [82]

Folding kinetics

The stopped flow CD measurements at 222 nm were performed on an Aviv 202 SF CD. HPr (~3 mg/mL) in denaturant was rapidly mixed to give an 10 fold dilution in 10 mM potassium phosphate at pH 7.0. 10 to 20 repeats were averaged under identical conditions to improve signal to noise. All kinetics were performed at 25 °C. Unfolding kinetics were fit to either a single or double exponential using KaleidaGraph software.[78] Folding reactions were performed as well; HPr (~3 mg/mL) in buffer was rapidly mixed to give an 11-fold dilution in denaturant 10 mM K/Pi pH 7.0 to initiate refolding. Kinetic

measurements were performed 10 to 20 times under identical conditions. These data were then averaged and fitted to either a single or double exponential using software.[78] The initial and final CD signals were determined from the fit to the kinetic experiments.

Data analysis

It has been established for a number of proteins that fold by a two-state mechanism that the denaturant dependence follows a linear dependence [83] as defined by:

$$\ln k_{NU} = \ln k_{NU}^{H2O} - ((m_{N-TS})/RT)[urea] \quad (36)$$

where k_{NU} and k_{NU}^{H2O} refer to a given rate constant for inter-conversion between states U and N, in the presence and absence of denaturant, respectively. N-TS refers to the denaturant dependence of the rate k_{UN} and reflects the change in solvent exposure between the N and the transition state (TS) to state U.

In a two-state reaction, the observed rate can be described by:

$$k_{obs} = k_{UN} + k_{NU} \quad (37)$$

Therefore, using equation 36 to describe the microscopic rate constant for folding and unfolding, and substituting these into equation 37, the denaturant dependencies for the observed rate in a two-state reaction can be fit using equation 38.

$$\ln k_{obs} = \ln[k_{UN}^{H_2O} \exp(\frac{-m_{UN}[D]}{RT}) + k_{NU}^{H_2O} \exp(\frac{m_{NU}[D]}{RT}) \quad (38)$$

where m_{UN} and m_{NU} refer to the change in solvent accessibility (to denaturant) between the unfolded and native states, respectively, relative to the rate-limiting step. Data were fitted using the non-linear regression software. [78]

ΔG_{UN} for a two-state reaction can be calculated from the kinetic rate constants using the equation 39.

$$\Delta G_{UN} = -RT \ln(\frac{k_{UN}}{k_{NU}}) \quad (39)$$

Three state folding

For the three-state folding model we assumed a pre-equilibrium whereby the amplitudes of the refolding phase represent the population of the intermediate in the unfolding mechanism. Thus the U to I transition is represented by the amplitudes and can be fit using equation 34. The I to N transition is represented by the k_{obs} generated from the chevron and is treated as a two-state analysis.

Differential scanning calorimetry (DSC)

All measurements were performed on a Microcal VP-DSC model calorimeter. Reported results were carried out at a scan rate of 60 °C/hr over a temperature range of 5 to 100°C. No scan rate dependence was observed from 30 to 90°C/hr. Two consecutive scans were performed to assess reversibility. HPr solutions ranging from 25 to 150 μM were dialyzed overnight at 4°C against buffer containing 10 mM Potassium Phosphate pH 7.0 or the specific concentration of salt. Dialyzed protein and buffer were used to load the sample and reference cell respectively. Experimental data were corrected for small mismatches between the two cells by subtracting a buffer *vs*. buffer baseline scan prior to data analysis. After normalizing for protein concentration a chemical baseline was calculated from the progress curves and applied. Then a two-state fit was applied using the Origin software package.[84]

Amyloid formation conditions

For some experiments, the protein of interest was dissolved from lyophilized powder in water and diluted from stock solutions to begin the reactions. Protein samples were incubated for at least eighteen hours in solutions containing various amounts of TFE (0-60%) within a buffer system of sodium phosphate, citrate, and borate. This buffer system allows for the effective use of pH values ranging from 2 to 10. Incubation occurred at room temperature in a sealed container to prevent evaporation.

Congo red and thioflavin T measurements

Congo red (CR) analysis was performed as described previously.[85; 86] A 10 μL sample of an aggregation reaction was added to 240 μL of 25 μM CR in a 0.1 M potassium phosphate buffer with 150 mM NaCl at pH 7.4. CR was prepared in 100 mM potassium phosphate buffer with 150 mM NaCl (pH 6.0) After 30 minutes absorbance was read at 540 nm and 480 nm and the amount of CR bound was calculated as follows:

$$Cb[M] = (A_{540}/25{,}295) - (A_{480}/46{,}306) \quad (40)$$

For Thioflavin-T (ThT) analysis, 10 μL of an aggregation reaction was added to 990 μL of 3 μM ThT in 50 mM potassium phosphate buffer at pH 6.0. Fluorescence was read immediately with $\lambda_{ex} = 450$ and $\lambda_{em} = 480$ nm.

Native Gel for Amyloid

A 8% acrylamide gel was poured with 25 mM Tris, 250 mM glycine (electrophoresis grade), pH 8.0 with 10% glycerol. The gels were stained with saturated Coomassie (R-250) in water to visualize the proteins.

CHAPTER III

ROLE OF SUPERPROTECTION AND THE ASP69 HYDROGEN BOND NETWORK IN *bs*HPr

INTRODUCTION

The determination of conformational stability is important to our understanding of protein folding. Understanding the forces that fold a protein is of paramount importance to the field of protein chemistry. One of the critical events in understanding the conformational stability of a protein is determination of its folding mechanism. Usually a protein of less than 100 amino acids will fold and unfold by an equilibrium two-state mechanism.

In this chapter, we study the *bs*HPr protein, a small 88 amino acid protein that was previously determined to fold by a two-state folding mechanism.[14] However, new evidence shows the protein to fold by a three-state mechanism (Chapter I discusses how to differentiate between a two-state and three-state folding mechanism). Furthermore, we found two ways to switch the folding mechanism to two-state; the D69A* *bs*HPr variant and the addition of salt.

Both the conditions that induce two-state mechanisms are linked to the destabilization of a hydrogen bond network. The hydrogen bond network consists of

hydrogen bonds from a side chain to two amide backbones, involving the side chain Asp 69 and the amides of residues 30 and 31 (Figure 15).[87] This hydrogen bond network is conserved in all HPr proteins and is important to the overall stability of the protein.

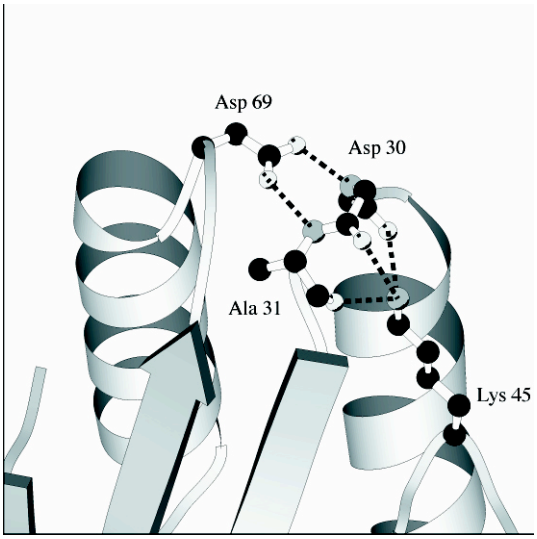

Figure 15. Diagram of the native Asp 69 interaction in wt *bs*HPr. Asp 69 makes two side-chain to main-chain hydrogen bonds with the backbone amides of residues 30 and 31.

RESULTS

Structural changes upon denaturation

The circular dichroism (CD) spectra of native and unfolded wt *bs*HPr are shown in Figure 16. The spectrum of the native form is a typical alpha+beta protein having minima at 222 and 206 nm. The variants G49E and D69A* all have identical

spectra. The greatest difference between native and unfolded CD spectra is at 222 nm Figure 16, and this wavelength was utilized to monitor unfolding.

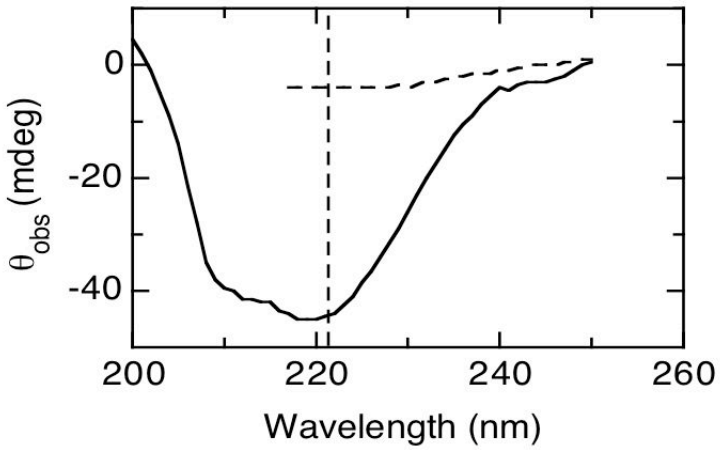

Figure 16. The circular dichroism spectrum of the native and unfolded *bs*HPr. The solid line represents folded protein. The unfolded spectrum (dashed line) was collected in 10 M Urea. The vertical line represents the greatest change in signal between native and unfolded states.

Chemical denaturation

To assess the conformational stability and unfolding mechanism for wt *bs*HPr, wt* and D69A*, chemical denaturation (urea and GdnHCl) was followed by circular dichroism to monitor changes in secondary structure. The wt* protein is a stabilized variant (G49E) of *bs*HPr which behaves identically to wt [32]. The *bs*HPr protein gives good signal change when followed by circular dichroism between native and unfolded

states (see Figure 16). These curves were fit to the linear extrapolation method (LEM) to estimate the free energy of unfolding in the absence of denaturant $\Delta G_{UN.}$[17; 88] (Figure 17)

$$[\theta_{obs}] = \frac{[\theta_N] + a_N[D]) + ([\theta_D] + a_D[D]) * \exp[m * ([D] - C_{mid})/RT]}{1 + \exp[m * ([D] - C_{mid})/RT]} \qquad (41)$$

θ_{obs} is the spectroscopic signal which was used to follow the unfolding reaction, D is the denaturant concentration using either urea or GdnHCl, a_N and a_D represent the denaturant dependence of the pre- and post-transition base lines, *m* is a measure of the dependence of the ΔG_{UN} on denaturant concentration, and C_{mid} is the midpoint of the denaturation curve. An estimate of the conformational stability in water (ΔG_{UN}) can be determined by multiplying the C_{mid} times the m value. This assumes a two-state relationship where the transition region can be extrapolated to 0 M urea.

The conformational stability determined by urea denaturation for wt *bs*HPr (ΔG_{UN} = 4.9 kcal/mol) fits well to a two-state mechanism and has a conformational stability of (ΔG_{UN} = 5.1 kcal/mol). However, stabilities determined from GdnHCl denaturations (ΔG_{UN} = 3.2 kcal/mol) do not agree with urea denaturations (Table1). This is not unexpected because GdnHCl is a more effective denaturant as well as a salt and it can screen coulombic interactions. [89; 90] The effects of GdnHCl will be explored further later in this section. Interestingly, the ΔG_{UN} determined by urea and GdnHCl for D69A* all agree well with each other (ΔG_{UN} = 4. kcal/mol for urea and GdnHCl).

Table 1 Thermodynamic parameters for the stability of the *bs*HPr proteins

	wt	D69A*	wt + salt	wt (D_2O)	D69A* (D_2O)	wt + salt (D_2O)
TDC[a]						
T_g (°C)	74.2	69.7	68.5			
ΔH_{vH} (kcal mol^{-1})	73.5	71.2	62.0			
ΔC_P (kcal mol^{-1} K^{-1})	1.45	1.59	ND			
$\Delta G(25°C)$ (kcal mol^{-1})	5.1	4.4	ND			
UDC[b]						
m (kcal mol^{-1} M^{-1})	1.0	1.1	1.1	0.99	1.1	1.1
C_m (M)	4.9	3.9	3.4	4.9	3.9	3.4
ΔG_{UDC} (kcal mol^{-1})	4.9	4.2	3.8	4.8	4.2	3.7
GDC[c]						
m (kcal mol^{-1} $M^{-1)}$)	2.4	2.4	ND			
C_m (M)	1.3	1.8	ND			
ΔG_{GDC} (kcal mol^{-1})	3.2	4.3	ND			
DSC[d]						
T_g (°C)	74.9	70.0	68.5			
ΔH_{cal} (kcal mol^{-1})	52.7[g]	71.5	63.3			
$\Delta H_{vH}/\Delta H_{cal}$	1.39	1.00	0.98			
HD[e]						
ΔG_{HD} (kcal mol^{-1})	—	—	—	6.8	4.9	3.7
Kinetic Data[f]						
ΔG_{UI} (kcal mol^{-1})	2.0	—	—	2.0	—	—
ΔG_{IN} (kcal mol^{-1})	4.9	—	—	5.0	—	—
ΔG_{UN} (kcal mol^{-1})	6.9	4.6	3.9	7.0	4.7	ND

[a]From the analysis of Thermal Denaturation Curves (Figure 18), as monitored by CD at pH 7.0. The ΔC_p and $\Delta G(25°C)$ values are from the fit of equation 42 to the combined data from thermal and urea unfolding studies as shown in Figure 19.

[b]From the analysis of Urea Denaturation Curves, as monitored by CD at 25°C, pH 7.0. The values in D_2O were collected at pH* 5.5 to match the conditions used in the HD experiment.

[c]From the analysis of GdnHCl Denaturation Curves, as monitored by CD at 25°C, pH 7.0.

[d]Results from the Differential Scanning Calorimetry studies on the HPr proteins at pH 7.0 (Figure 30).

[e]Results from the HD data collected at 25°C, pH* 5.5 in D_2O. The values were obtained by averaging the three largest residue-specific ΔG_{HD} values.

[f]Derived from the kinetic data shown in Table 2 at 25°C, pH 7.0.

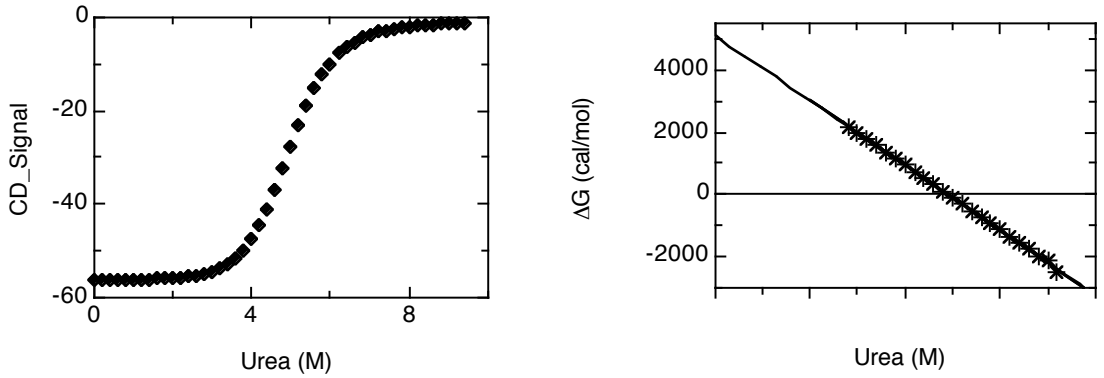

Figure 17. Typical urea denaturation of wt *bs*HPr monitored by CD at 222 nm in 10 mM potassium phosphate pH 7.0 298 K. The right panel shows the linear extrapolation of the data.

Chemical denaturation were performed in deuterated buffer to assess isotope effects on stability and for comparison with ΔG_{HD} values determined for slow exchanging amide protons (later in this chapter). The stability for wt *bs*HPr increased by 0.5 kcal/mol in D_2O relative to water whereas D69A* *bs*HPr stability did not vary.

Thermal denaturation

Typical thermal denaturation curves followed by CD and DSC for wt *bs*HPr are shown in Figure 18. A two-state relationship was assumed, where the signal at -35 mdeg represents the native form of the protein and -12 mdeg represents the unfolded form. A cooperative transition between the two-states is observed as a function of temperature.

$$\theta_{obs} = ((N_o + a_N(T)) + (D_o + a_D(T)) * \exp(\frac{\Delta H_{VH}}{RT}(T - T_m)))/(1 + \exp(\frac{\Delta H_{VH}}{RT}(T - T_m)) \quad (42)$$

Equation 42 is a variation of the van't Hoff equation that we use to describe the two-state, thermally induced unfolding reaction.[79] θ_{obs} is the CD signal that was used to follow the unfolding reaction, a_N and a_D represent the unfolding dependence of the pre- and post-transition base lines, and N_o and D_o are the intercepts of the native and unfolded baselines. We also obtain the thermal midpoint (T_m) of the unfolding curve and the van't Hoff enthalpy change (ΔH_{vH}).

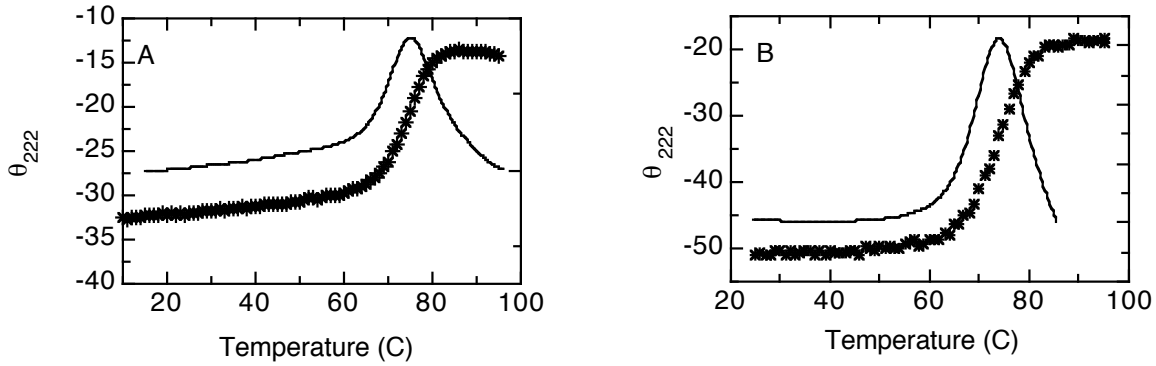

Figure 18. Thermal unfolding of the HPr variants as monitored by CD and DSC. A) The stars represent typical van't Hoff thermal denaturations for wt *bs*HPr. The solid line represents a typical thermal denaturation for wt *bs*HPr followed by calorimetry. 2B) The stars represent a typical van't Hoff thermal denaturation for D69A* *bs*HPr. The solid line represents a typical thermal denaturation for D69A* *bs*HPr followed by calorimetry. A point of notice is that the pre-transition baselines for the wt protein seem to have some slope to them whereas the pre-transition baselines for D69A* are flat.

Differential scanning calorimetry (DSC) was also used to analyze the thermal unfolding of wt *bs*HPr and D69A* (Figure 18). These experiments were analyzed using the software package that comes with the instrument[84] to determine the midpoint of

thermal denaturation T_m, and the calorimetric enthalpy of change ΔH_{cal}.The two-state fit to the DSC data for the wt *bs*HPr protein yielded a T_m =75.0 °C and ΔH_{cal}= 53.0 kcal/mol and for D69A* T_m =70.0 °C and ΔH_{cal}= 71.5 kcal/mol. However, the unfolding of these proteins is problematic in that they aggregate upon unfolding at DSC concentrations (60% reversibility wt *bs*HPr, 35% reversibility D69A* *bs*HPr). With this in mind, we varied the protein concentrations to assess the effects of aggregation on the thermodynamic parameters determined from DSC. The thermodynamic parameters associated with the various protein concentrations do not differ. Therefore, protein aggregation does not affect this analysis. The ΔH_{vH} from thermal denaturations followed by CD were compared with the ΔH_{cal} ($\Delta H_{vH}/\Delta H_{cal}$) to assess the use of the two-state analyses for these proteins.[21] For wt *bs*HPr ($\Delta H_{vH}/\Delta H_{cal}$)= 1.4 and for D69A* *bs*HPr ($\Delta H_{vH}/\Delta H_{cal}$)= 1.0. The results of the DSC experiments are in poor agreement with the thermal denaturation curves analyzed using a two-state fit for the wt protein, but are in excellent agreement for the variant. Thus, the thermal denaturation of the wt protein does not appear to follow a two-state folding mechanism, However, the D69A* protein does seem to follow a two-state folding mechanism.[21]

It was necessary to determine the change in heat capacity (ΔC_p) in order to calculate ΔG_{UN} for comparison to other techniques. We utilized the method established by Pace & Laurents that allows for an excellent estimate of (ΔC_p) due to the cold denaturation phenomena.[19] Equilibrium urea denaturations followed by CD (results discussed later) over a wide temperature range were performed. These data were

combined with the data in the transition of the thermal unfolding curve in the absence of urea and analyzed with a modified version of the Gibbs-Helmholtz equation (Figure 19).

$$\Delta G(T) = \Delta H_g(1 - T/T_g) - \Delta C_p[(T_g - T) + T\ln(T/T_g)] \qquad (43)$$

The results were fit to equation 43. All the variants showed little difference in the ΔC_p values suggesting that the overall change in accessible surface area between folded and unfolded changes little.

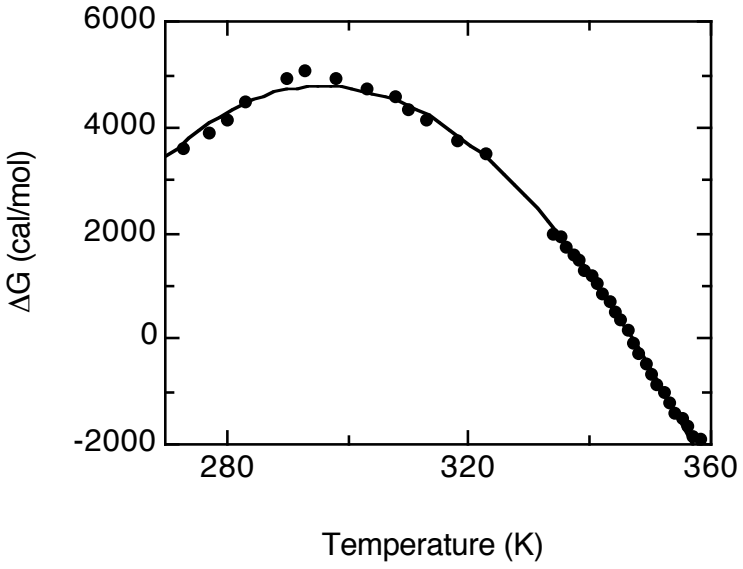

Figure 19. Conformational stability as a function of temperature. The solid line represents the fit using equation 43.

Folding kinetics

Stopped flow CD was used to follow the refolding of wt *bs*HPr, wt*, and D69A* at 222 nm. The refolding for wt *bs*HPr shows a large decrease (60% of the total change) in CD signal within the dead time of the instrument (~5 ms) (Figure 20). This burst phase is followed by a slower reaction that fits well to a single exponential (k_{app} = 15/s at 0.84 M urea), indicating that wt *bs*HPr refolding is at least a biphasic process.

$$U \underset{k_{IU}}{\overset{k_{UI}}{\leftrightarrow}} I \underset{k_{NI}}{\overset{k_{IN}}{\leftrightarrow}} N \qquad (44)$$

Such biphasic kinetics suggests a population of a kinetic folding intermediate forms very early in folding. The burst phase intermediate is also observed for the wt* variant. In contrast, the D69A* protein refolds and fits well to a single exponential process. The refolding phase of D69A* accounts for 80% of the total change, suggesting that it follows a two-state folding mechanism.

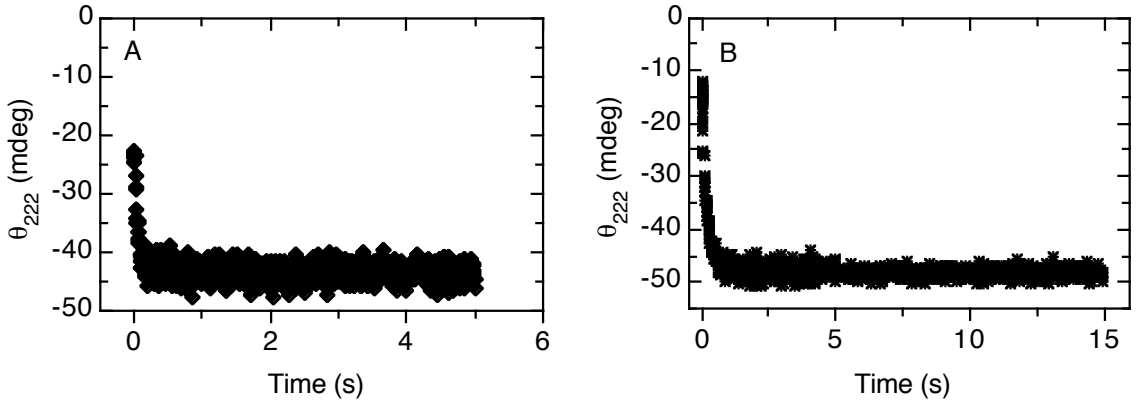

Figure 20. Progress curves for the refolding of wt *bs*HPr and B) D69A*.

Wt bsHPr and wt fold by a kinetic three-state mechanism*

Because the formation of the intermediate occurs within the dead time of the instrument (τ = 5 ms), the $k_{burst\ phase}$ of the burst phase must be (> 200/s). Since the $k_{burst\ phase}$ is ~5 times the magnitude of the slow, observable phase (43/s), we can assume that the burst phase represents a pre-equilibrium between the unfolded and intermediate states. Therefore, the stability of the transient intermediate can be determined by analyzing the amplitude of the burst phase as a function of final urea concentration (Figure 21). [42] Assuming a two-state transition between the U state and I state, equation 22 can be applied to determine the thermodynamic stability of the folding intermediate. The conformational stability of the U to I is ΔG_{UI}= 2.0 kcal/mol (C_{mid} =2.0 M , m = 1.0 kcal/mol M) (Table 2)**.**

Table 2 Kinetic parameters for the HPr proteins

	Wt	D69A*	wt + salt	wt (D_2O)	D69A* (D_2O)
k_f (s^{-1})	42	140	90	79	84
k_u (s^{-1})	0.0099	0.055	0.12	0.016	0.031
m_f (kcal mol^{-1} M^{-1})	0.79	0.96	0.85	0.76	0.85
m_u (kcal mol^{-1} M^{-1})	0.31	0.24	0.20	0.25	0.32
m (kcal mol^{-1} M^{-1})	1.10	1.20	1.05	1.01	1.17
C_m (M)	4.4	3.8	3.8	4.9	4.0
ΔG_{UI} (kcal mol^{-1})	2.0	—	—	2.0	—
ΔG_{IN} (kcal mol^{-1})	4.9	—	—	5.0	—
ΔG_{UN} (kcal mol^{-1})	6.9	4.6	3.9	7.0	4.7

[a] The kinetic rate constants and the associated m_u and m_f values are from a fit of equation 45 to the urea dependence of the observed rate constants as shown in Figure 22. The m value is the sum of the resolved kinetic m values and the C_m represents the urea concentration where $k_f = k_u$. The ΔG_{UI} values for wt was determined from a fit of equation 41 to the amplitudes of the burst phase (Figure 21) and the ΔG_{IN} values are determined from the observed rate constants for folding and unfolding. The ΔG_{UN} values for wt in H_2O and D_2O are the sums of the ΔG_{UI} and ΔG_{IN} values while the ΔG_{UN} for D69A* is calculated from the observed rate constants.

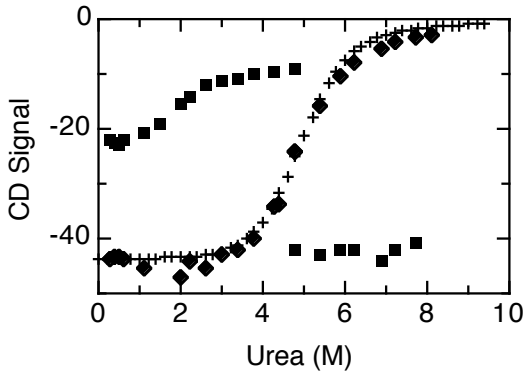

Figure 21. Amplitude analysis of wt *bs*HPr. The squares represent the burst phase amplitudes calculated from the refolding and unfolding curves. The diamonds are the CD signal at completion of the refolding and unfolding experiment and the crosses represent the urea-induced equilibrium unfolding reaction followed by CD at 222 nm.

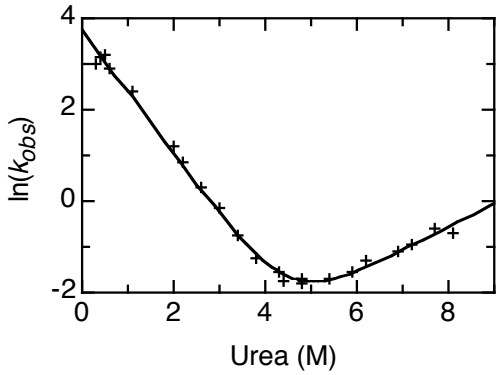

Figure 22. Chevron plot of wt *bs*HPr. The observable rates of refolding and unfolding are plotted against urea concentration. The curve represents the best-fit using equation 46.

The kinetics for *bs*HPr and G49E folding were examined over a wide range of urea concentrations. All the refolding experiments show the presence of a burst phase. All the unfolding experiments fit well to a mono exponential, indicating a lack of intermediates in the unfolding reaction. Figure 22 shows the combined kinetic data for all the experiments, ignoring the burst phase. The dependence of ln k_{obs} on urea has the typical V shaped curve, with the left arm of the curve dominated by the refolding rate and the right arm of the curve dominated by the unfolding rate. The k_{obs} data were fit using a simple two state model (this describes the transition from I to N) in which the logarithms of the folding and unfolding rate constants depend linearly on denaturant with the slopes of m_{IN} and m_{NI} respectively.

$$\ln k_{obs} = \ln[k_{IN}^{water} \exp(\frac{-m_{IN}[urea]}{RT}) + k_{NI}^{water} \exp(\frac{m_{NI}[urea]}{RT})] \qquad (45)$$

The refolding rate constants (k_{IN}) and unfolding rate constants (k_{NI}) are extrapolations of the urea dependences of the k_{obs}. From these data, the thermodynamic stability for the I to N transition can be determined from the kinetic data. The wt *bs*HPr protein folds from the I state to the native state with a folding rate constant of k_{IN}= 43/s and unfolds from the native state with a rate constant of k_{NI}= 0.0099/s.

There are three traditional kinetic tests to determine if a protein folds using an equilibrium two-state mechanism.[46] The first test would be to compare the ratio of the forward and reverse rate constants.

$$K_{eq} = \frac{k_{UI}}{k_{IU}} * \frac{k_{IN}}{k_{NI}} \qquad (46)$$

For a two-state mechanism, the ratio must give the same equilibrium constant as that measured by an equilibrium method under the same conditions. When we evaluate the kinetic data using equation 46, we do not find agreement with the equilibrium data. However, if we ignore the burst phase (U to I transition) the kinetic data agrees better with the equilibrium data.

Test two is to show that the sum of the forward and reverse m values equals the m value for equilibrium unfolding for a two-state transition.

$$m_{eq} = m_{IN} + m_{NI} + m_I$$
$$m_I = m_{IU} + m_{UI} \qquad (47)$$

where all three are positive. The summed m values for the I to N transition, not including the U to I transition, are equal to that of the equilibrium m value.

The third test looks at deviations from the simple V shaped curves of rate constants versus urea that are shown in Figure 22. Often, three-state folding mechanisms show deviations from linearity known as roll over (See Chapter I). While we do not observe roll-over, we believe the wt *bs*HPr does not adhere to a two-state mechanism.

The unity between the U to I transition and the equilibrium denaturation data suggest that the burst phase may be an artifact of the stopped flow experiment most often attributed to aggregation or solvent reorganization. [43] Previous studies have shown that transient protein aggregation can cause deviations form linearity in the chevron plot and the presence of a burst phase. An important consideration in our analysis therefore was to determine weather wt *bs*HPr folding was concentration dependent. We investigated this by varying the protein concentration (20, 30, and 40 μM). Figure 23 shows the effect of protein concentration on the burst phase amplitude over this concentration range. We did not observe a change in the observable rate constants, suggesting that the observed folding kinetics of wt *bs*HPr are independent of protein concentration.

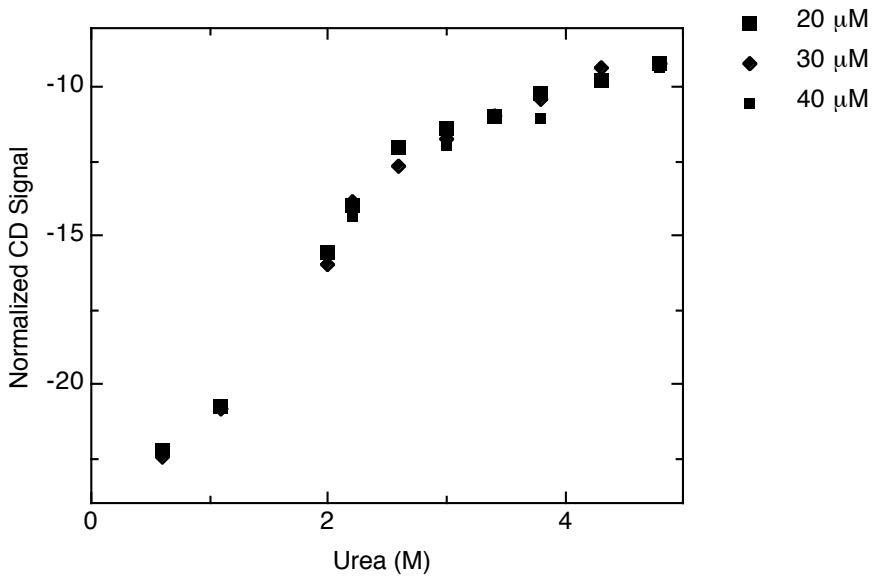

Figure 23. Concentration dependence of the *bs*HPr refolding reaction. The refolding amplitudes observed from refolding jumps from 5 M urea to 0.6 M urea at 20, 30, and 40 μM protein concentrations.

To further investigate the possibility of aggregation, we investigated the structure of the burst phase intermediate (Figure 24). The CD spectrum of the burst phase intermediate was reconstructed from the burst phase amplitudes measured at different wavelengths. [42] The large CD signal and spectral properties indicate that this early intermediate has a significant amount of secondary structure. The spectrum of this intermediate appears to be very similar to that of the native, fully-folded *bs*HPr protein.

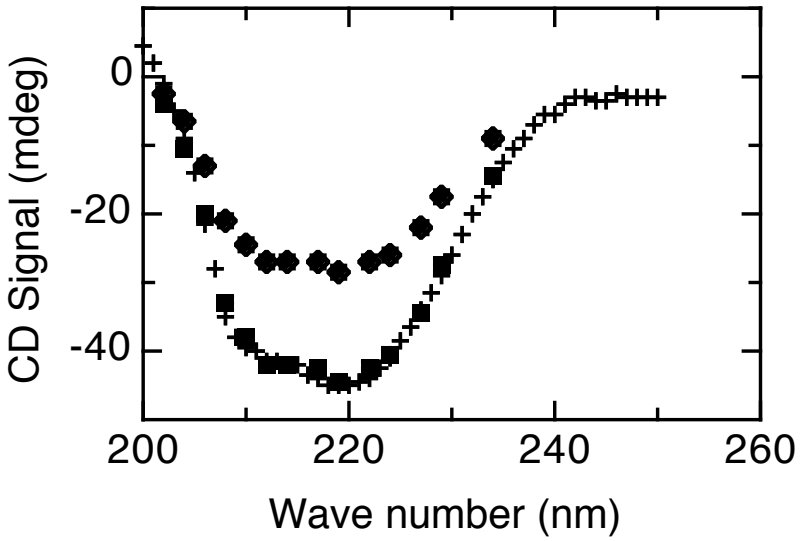

Figure 24. CD spectrum of the burst phase intermediate. CD spectrum of wt *bs*HPr in 0.6 M Urea (plus sign). Filled dots represent the CD spectrum of the burst phase. The filled squares represent the final CD signal from the refolding experiments.

D69A folds by a two-state mechanism*

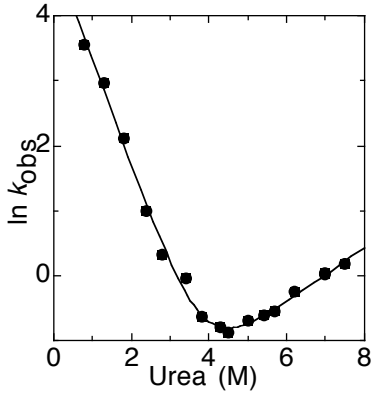

Figure 25. Chevron analysis of D69A*. The observable rates of refolding and unfolding are plotted against urea concentration. The curve represents the best fit to equation 46.

Figure 20 shows representative refolding traces for D69A* measured on a time scale which monitors fast refolding process. The data show that the protein folds by a very rapid, apparently monophasic process consistent with a single rate-limiting activation barrier. We can monitor 80% of the refolding process. Similarly, the unfolding data for D69A* fit well to a single exponential function at all urea concentrations studied (data not shown). The rates of refolding and unfolding of D69A* measured at a range of final urea concentrations are plotted in Figure 25. The natural logarithm of both the observed folding and unfolding rates of D69A* display a linear dependence on denaturant[49] (Figure 25). This behavior is commonly found in proteins with less than 100 amino acids and can be described by a two-state mechanism.

$$U \underset{k_{UN}}{\overset{k_{NU}}{\leftrightarrow}} N \qquad (48)$$

The data in Figure 24 can be fit to equation 38 providing the k_{UN}, k_{NU}, m_{UN}, and m_{NU} values. D69A* folds to the native state with a rate constant k_{UN} of 141/s and unfolds with a rate constant k_{NU} of 0.055/s. There is a strong denaturant dependence to the refolding rate constant (m_{UN}= 960 cal/mol M) while the unfolding rate constant changes only slightly with denaturant concentration (m_{NU}= 250 kcal/mol M). Since the m value depends upon the difference in the surface area exposed between the two-states, the compactness of the transition state, relative to the folded state, can be represented by the

β_T value ($\beta_T = m_{U\text{-}TS}/m_{eq}$).[83] This value reflects the placement of the transition state on the folding reaction coordinate based upon compactness. When $\beta_T = 0$ the transition state has the same solvent accessible surface area as the unfolded state, whereas a $\beta_T = 1$ represents a transition state which is as compact as the native state. The $\beta_T$value of wt *bs*HPr is 0.7 and of D69A* is 0.8, both showing that the transition state is highly compact.

As discussed earlier, if D69* folds in a reaction that can be described by equation 48, a two-state mechanism, then the ΔG_{UN} and m_{UN} determined by equilibrium and kinetic analyses must agree. The values for ΔG_{UN} and m_{UN} calculated from kinetic data using equation 45 are 4.6 kcal/mol and 1.2 kcal/mol M. These values compare well with ΔG_{UN} of 4.2 kcal/mol and m_{UN} of 1.2 kcal/mol M determined from equilibrium measurements. This suggests that a two-state mechanism is sufficient to describe the folding and unfolding of D69A*.

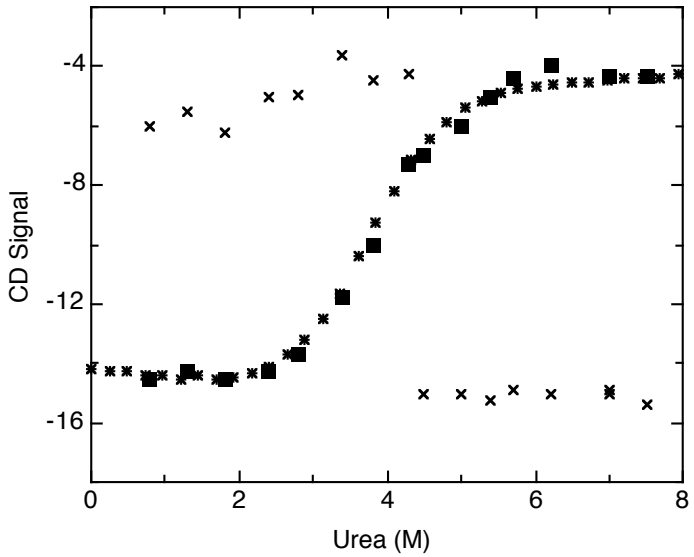

Figure 26. Amplitude analysis of D69A*. The X's represent the refolding and unfolding amplitudes. The filled squares represent the signal upon completion of the kinetics experiments. The stars represent the equilibrium urea denaturation followed by CD 222 nm.

The validity of the two-state fit for D69A* is confirmed by the refolding and unfolding amplitudes (Figure 26). These experiments determine whether the observed refolding amplitude is consistent with the expected change in signal between the unfolded and native states at each urea concentration. Figure 26 shows that at all urea concentrations studied there is no dead time phase detected by CD in the refolding of D69A*. Together with the fact that the mono exponential folding and unfolding reactions describe the equilibrium properties of the protein, these data confirm that D69A* refolds in a two-state manner.

The folding intermediate of wt bsHPr is destabilized in salt.

The equilibrium ΔG_{UN} calculated from GdnHCl denaturations of wt *bs*HPr are not equivalent to ΔG_{UN} calculated from urea denaturations (ΔG_{UN} GdnHCl = 3.2 kcal/mol; ΔG_{UN}= 4.9 kcal/mol for urea). Interestingly, ΔG_{UN} calculated for D69A* from urea and GdnHCl are equivalent. With this in mind the effects of salt on equilibrium stability were evaluated (Figure 27). Salt does not destabilize D69A* whereas wt *bs*HPr is destabilized at low salt concentrations. The wt *bs*HPr protein is destabilized by salt presumably due to general coulombic screening of favorable interactions. Above 0.25 M NaCl, the wt and D69A* proteins both show similar increases in stability.

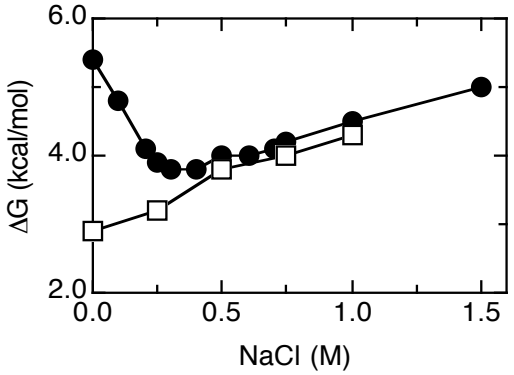

Figure 27. The effects of NaCl on stability for wt *bs*HPr and D69A*. Stabilities were determined from urea denaturation experiments. The filled circles represent wt *bs*HPr and the closed squares D69A*.

We also evaluated the effects of salt on the refolding kinetics of wt *bs*HPr. Figure 28 shows the refolding progress curves of wt *bs*HPr in 100 mM NaCl at pH 7.0 . The

refolding of wt *bs*HPr appears to be well described by a monophasic process fitting well to a single exponential relaxation. The progress curve shows 80% of the total refolding process, suggesting that we have destabilized the burst phase intermediate.

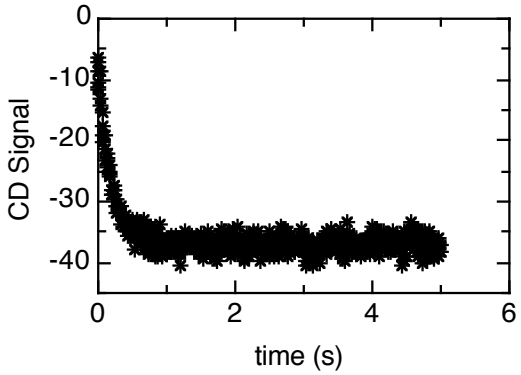

Figure 28. Progress curve for refolding of wt *bs*HPr in 100 mM NaCl .

DSC curves were also performed in 100 mM NaCl. Figure 29 shows the DSC raw data and the curve suggests the possibility of a third state with a T_{m1} ~ 45 °C and T_{m2} ~ 72 °C. In comparing the two curves, with and without salt, it appears that an early transition may be present in the unfolding reaction; it is especially evident in 100 mM NaCl and the addition of 250 mM NaCl eliminates the early transition. This would explain why the two-state analysis did not agree with the CD thermal denaturation.

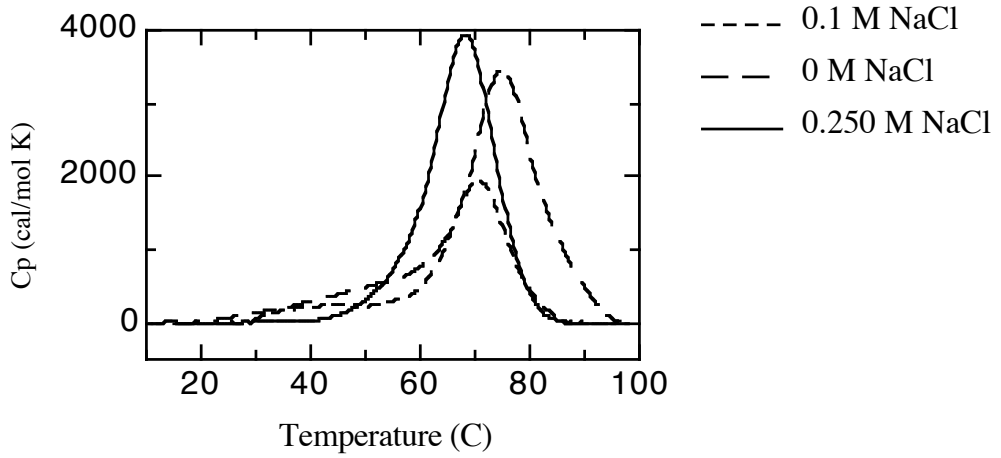

Figure 29. Thermal denaturation followed by DSC for wt *bs*HPr in various concentrations of salt. The 0.1 M sample represented by short dashed lines suggests the presence of a intermediate in the pre-transition baseline, whereas at 0.25 M NaCl (solid line) the pre-transition is flat.

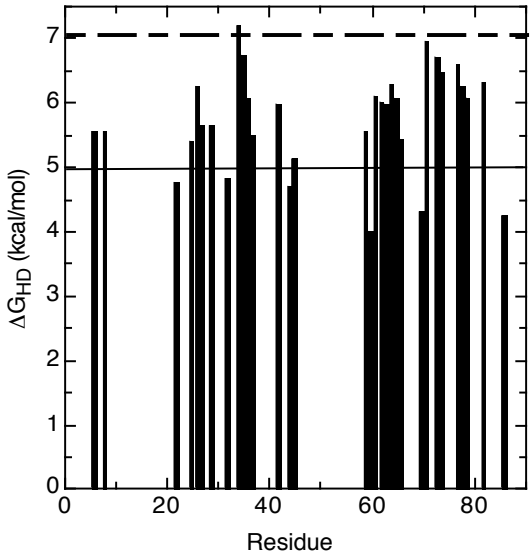

Figure 30. Residue specific conformational stabilities calculated from a hydrogen deuterium exchange. Experiments performed at pH 5.5 298 K and followed by NMR. The solid line represents the stability calculated from a urea denaturation in identical conditions as the H/D exchange followed by CD and analyzed using a two-state assumption. The dotted line represents the stability calculated using the kinetic folding data using a three state model.

It would appear from our data that the structure of the intermediate is similar to that of the folded native structure. Suspecting that this makes detection of a equilibrium intermediate difficult, we decided to utilize another technique to determine the stability of the wt *bs*HPr protein. Equilibrium HD exchange of amide protons with solvent is a probe for local structural stability. Under EX2 conditions ($k_f >> k_{rc}$), the free energy for the opening reaction that leads to exchange of individual protons can be obtained from the measured rates of exchange (k_{obs}) and an estimate of the random coil exchange rate (k_{rc}).

$$\Delta G_{HD} = -RT\ln(k_{obs}/k_{rc}) \qquad (49)$$

We were able to measured 32 exchange rates for non-overlapping amide protons by following changes in their peak intensities over time in ^{1}H-^{15}N HSQC spectra. The observed rates, k_{ex}, were obtained from single exponential fits to the data. Random coil (k_{rc}) exchange rates were calculated as described in Materials and Methods. Even though the EX2 exchange mechanism was not tested directly, the assumption is very likely to be correct for both experimental conditions, on the basis of our determination of k_f in 0 M urea and estimates of the fastest k_{rc} (in 0 M urea k_f= 79/s and k_{rc}=2.5/s). Since the measure of k_f reflects global folding, protons exchanging through local fluctuations should have even higher closing rates. Figure 30 shows the ΔG_{HD} values for the amide protons calculated from equation 49. In Figure 30, the cutoff line at 4.9 kcal/mol is the ΔG_{D2O} value from urea denaturation in D_2O and represents the expected ΔG_{HD} for protons if they exchange through a global unfolding pathway only. What is apparent from these data is that a number of protons have stabilities greater than the equilibrium ΔG_U values. Allowing for a 0.5 kcal/mol variance and correcting for proline isomerization does not account for these discrepancies.[91] However, when we compare the thermodynamic stabilities calculated using a three-state analysis of the kinetic data, we get excellent agreement with the HD exchange data (dotted line in Figure 30).

We have been able to remove the kinetic intermediate in two ways: one by the addition of salt and secondly, by the D69A* variant. With this in mind, hydrogen

exchange was performed on D69A* and wt *bs*HPr in 100 mM NaCl. In both cases proteins show good agreement between HD exchange and equilibrium urea denaturations followed by CD (Figure 31). Even though we did not test if the exchange was in EX2, the assumption is likely correct given that the (k_f = 83/s D69A* and k_f = 98/s wt *bs*HPr in 100 mM NaCl compared with k_{rc} is 2.4/s which is ~ 40 to 50 times slower. In Figure 31 the line at 4.2 kcal/mol is the ΔG_{D2O} for D69A* and ΔG_{D20} =3.7 kcal/mol for wt *bs*HPr in 100 mM NaCl. These values were determined from equilibrium urea denaturation experiments and represent the expected ΔG_{HD} for protons if they exchange through global unfolding only. It is apparent that $\Delta G_U = \Delta G_{HD}$. The implications of this discrepancy in free energy for the urea denaturation and hydrogen exchange data will be discussed below.

DISCUSSION

Some of the data presented here have been previously analyzed using a two-state analysis (urea and thermal denaturations followed by CD).[14] However, recent data suggests that a three-state analysis better describes the unfolding of wt *bs*HPr. We have measured and calculated the equilibrium conformational stability ΔG_{UN} for wt *bs*HPr by 5 equilibrium methods and 1 kinetic method. We used two methods of thermal denaturation (DSC and van't Hoff analysis) two methods of solvent denaturation (urea

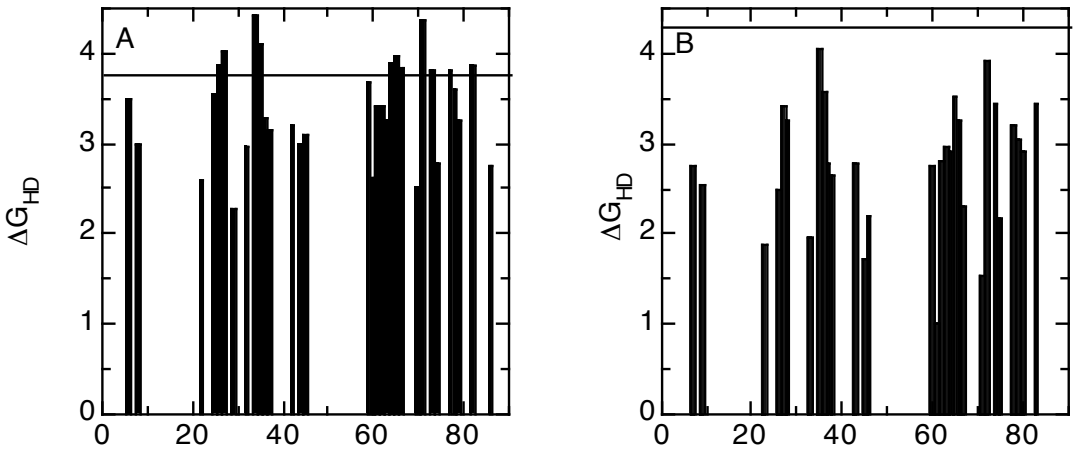

Figure 31. Residue specific conformational stabilities for D69A* and wt *bs*HPr in 100 mM NaCl. The ΔG_{HD} values were calculated from a hydrogen deuterium exchange experiment performed at pH 5.5 followed by NMR. The solid lines represent the stability calculated from urea denaturations in identical conditions as the H/D exchange followed by CD and analyzed using a two-state assumption: A) The exchange behavior of wt *bs*HPr in 100 mM NaCl, B) The exchange behavior of D69A*. As one can see, agreement between the traditional method and H/D exchange has been achieved.

and GdnHCl) the equilibrium method of hydrogen exchange and a kinetic method. We know that the measurements are correct and reproducible.

The data that most supports a two-state unfolding mechanism are the agreement of equilibrium thermal and urea denaturations as followed by CD. Below we discuss the various methods and show how the denaturations followed by CD miss the first transition in the three-state unfolding mechanism.

Thermal denaturations: DSC

Differential scanning calorimetry (DSC) measures heat released upon folding/unfolding. [92] In a DSC experiment, a sample is heated over a range of temperature and at some point, the protein starts to unfold and releases heat. If one plots heat released as a function of temperature, then the area under the curve represents the total heat or enthalpy change (ΔH_{cal}) for the entire process. The characteristic peak of such a plot provides the temperature (T_m) at which the reaction is half-complete. Plots of heat released as a function of temperature are fit using a variety of models to provide the total heat or enthalpy change, heat capacity change, and T_m of the unfolding reaction. The unfolding followed by DSC shown in Figure 29 for wt *bs*HPr is reproducible, independent of concentration and scan rate, but the base lines are not well-defined. This is problematic, since well defined baselines are needed for accurate fitting. It appears that the pre-transition base line contains an unfolding transition. In fact, the addition of salt (100 mM NaCl) resolves the first transition and allows for it to be visualized during

the unfolding. Interestingly, the addition of 250 mM NaCl eliminates the sloping baselines and the possible intermediate (Figure 29).

van't Hoff analysis

The van't Hoff analysis measures thermodynamic parameters indirectly by measuring the equilibrium constant which is related to ΔG by $\Delta G=-RT \ln K_{eq}$ where R is the gas constant and T is the absolute temperature.[93] In the simplest case, the protein unfolds in a two-state manner which is described by $K_{eq}=[F]/[U]$, where the brackets indicate concentrations of the two species. We monitor the change in circular dichroism as a function of temperature to obtain $K_{eq}(T)$. This allows us to calculate ΔH_{vH} as the change in the K_{eq} with temperature. In our experiments, the initial CD signal is operationally defined as folded and the signal at 100 °C is defined as unfolded. The unfolding curves have well defined pre- and post-transition baselines. The pre-transition baselines are slightly sloped for the wt *bs*HPr (Figure 18). Ultimately, we are indirectly measuring these thermodynamic parameters based on the assumption that our spectroscopic signal detects all the states involved in this equilibrium process.

One of the hallmarks of a two-state unfolding analysis is that the enthalpies of unfolding determined by van't Hoff analysis and differential scanning calorimetry are equal.[93] The ΔH_{vH} does not equal ΔH_{cal} for the wt *bs*HPr protein. Determination of the enthalpy of unfolding by van't Hoff analysis yields an enthalpy change which is greater than that determined by DSC: 85 vs. 53 kcal/mol. The pre- and post-transition baselines

are very important for an accurate determination of ΔH_{vH}. As you can see from the unfolding curves, (Figure 18) we get excellent unfolding data and ΔH_{vH} never varies more than 3%. However, examination of the DSC curves suggests that the difference in ΔH_{vH} and ΔH_{cal} may be due to baseline discrepancies which are revealed in the DSC unfolding data (Figure 18). The baselines for unfolding do not appear flat (Figure 18) and this would provide an increased error in the analysis. The lack of agreement between these two thermal denaturations suggests that a two-state analysis does not accurately describe the thermally induced unfolding, but rather a three-state analysis is needed to accurately describe the unfolding of wt *bs*HPr.

Solvent denaturation

Our probe for monitoring the solvent-induced unfolding of wt *bs*HPr is the change in the far-UV CD Signal. We revisited urea-induced unfolding to ascertain if a three-state unfolding mechanism may apply to wt *bs*HPr. Solvent unfolding data were analyzed with the linear extrapolation method. Key to this analysis is that the curves are reversible, independent of protein concentration and the transition induced by urea is two-state. The agreement between thermal and urea unfolding followed by CD suggests that we are measuring the differences between the same two spectroscopic states. The coincidence of calculated ΔG_{UN} supports the use of a two-state assumption. Interestingly, the thermally unfolded and the urea unfolded protein do not have the same signals. There are several reports where thermally and solvent induced unfolded proteins give different

signals.[94] It was shown that in many cases the two different unfolded signals can be adequately described as energetically similar. While we are probably measuring the difference between the same two states, folded (25°C) and unfolded (urea or temperature), the question still remains: are the operationally defined folded and unfolded states energetically the same as those we measure by DSC or H/D exchange monitored by NMR?

Equilibrium hydrogen exchange

It has long been recognized that protein equilibrium conformational stabilities can be estimated from the rate constants of exchange of the most slowly exchanging amide hydrogen. It was shown that if the effects of D_2O and proline isomerization are considered, the conformational stabilities calculated from hydrogen exchange rate constants are in excellent agreement with those determined by traditional methods. [91] The advantage of H/D exchange is that we will determine the absolute stability of the unfolding process regardless of intermediates (the technique is model independent).[8] The stabilities of wt *bs*HPr calculated from equilibrium hydrogen exchange and urea induced denaturations followed by CD do not agree. In contrast, D69A* follows an apparent two-state unfolding mechanism and, stabilities determined from urea denaturation, thermal denaturation and hydrogen exchange of D69A* show agreement.

Kinetics

Initially, kinetic folding experiments were utilized to show that the H/D exchange reactions were in EX2. Interestingly, the kinetic data suggests that the equilibrium urea denaturation followed by CD is not a two-state process and is not reporting on the total conformational stability ΔG_{UN}. It is clear from the data presented here that the kinetic folding of the wt *bs*HPr protein is three-state process. The U to I transition is exhibited by a burst phase and the intermediate appears to have a structure similar to that of the wt *bs*HPr protein. When conformational stability is calculated from the kinetic data, the stability is ΔG_{UN} = 6.6 kcal/mol. The equilibrium data calculated from equilibrium urea denaturations followed by CD is ΔG_{UN} = 4.9 kcal/mol. These clearly do not agree. However, the hydrogen exchange stabilities are in agreement with data calculated from kinetic data. I propose that the U to I transition is silent in the equilibrium experiments followed by CD, most likely because the intermediate is structurally similar to that of the wt *bs*HPr protein.

The equilibrium unfolding of wt *bs*HPr consists of three-states: the native state, an intermediate state, and an unfolded state. The native state is composed of three helices (A,B,C) and four beta strands (A,B,C,D) as defined by X ray and NMR structures. The intermediate state consists of an alpha+beta mix, mostly likely the C Helix and D strand, as the C helix and D strand are the folding core as defined by the H/D experiments. The CD spectra of the I and N states are nearly identical and the

transition between them is only weakly cooperative. The unfolded state is devoid of structure.

This model suggests that the removal of the intermediate state should bring identity to conformational stabilities calculated from equilibrium urea denaturations, kinetic experiments and hydrogen exchange experiments. We were successful in removing the burst phase by two methods 1) the addition of 100 mM salt and 2) mutation of Asp 69.

100 mM NaCl removed the intermediate state and subsequently brought identity to the determination of the conformational stability. The NaCl dependence to stability for wt *bs*HPr shows an initial sharp decrease in stability followed by an increase in stability. We attribute this initial decrease due to Debye Hückel screening of favorable coulombic interactions. The increase in stability beyond 0.25 M NaCl are probably due to salt stabilization. The interactions screened at low NaCl are probably interactions involved in the stabilization of the intermediate which includes the interactions made by Asp69.

Aspartic acid 69 is critically important to the stability of the wt *bs*HPr protein. Aspartic acid 69 is involved in a hydrogen bond network with the amides of residues 30 and 31. The removal of the side chain of this residue destabilizes the protein by 2.0 kcal/mol. In order to study this interaction, the D69A mutant was made in the G49E stabilizing background. The G49E *bs*HPr (wt*) behaves similarly to the wt protein. To support the idea that the removal of Asp 69 and the salt screening of the wt protein are achieving the same goal, the effects of salt on stability of the D69A wt* protein were investigated. D69A* shows no Debye Hückel screening, suggesting, that Asp 69 is

involved in the favorable coulombic interactions which were screened by salt in wt *bs*HPr.

CHAPTER IV

THE ROLE OF ASP 69 IN SUPER PROTECTION OF *ec*HPr

INTRODUCTION

Properly defining the folding mechanism allows one to determine the thermodynamic stability of a protein (Chapters II, III). Here, we perform a comprehensive kinetic analysis of the HPr protein from *E. coli* (*ec*HPr) in order to test the two-state folding mechanism, the EX2 assumption for the hydrogen exchange experiment, and to validate the correction factor for proline isomerization (Chapters I and III address these concerns in detail). Previously a kinetic analysis helped in defining a three-state folding mechanism of *bs*HPr (Chapter III).

The histidine containing protein (HPr) is a small globular protein that functions as a phosphocarrier protein of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS).[62] The HPr protein contains no cysteine, tryptophan, or tyrosine residues and has two prolines that are in the *trans* conformation. When the conformational stability is determined by solvent denaturation and compared with hydrogen exchange super protection is observed. Super protection is when ΔG_{UDC}

determined by traditional methods does not agree with ΔG_{HD} determined by hydrogen exchange. Under the conditions of the HD experiment reported by Huyghues-Despointes, the super protection for *ec*HPr is minimal and falls within the 1 kcal/mol allowance specified by Huyghues-Despointes et al.[8] However, under many conditions, super protection exists and ΔG_{HX} for wild type *E coli* HPr are on the order of 3 kcal/mol greater than ΔG unfolding values as measured by urea denaturation (Figure 32). (For more information regarding HPr see chapter I)

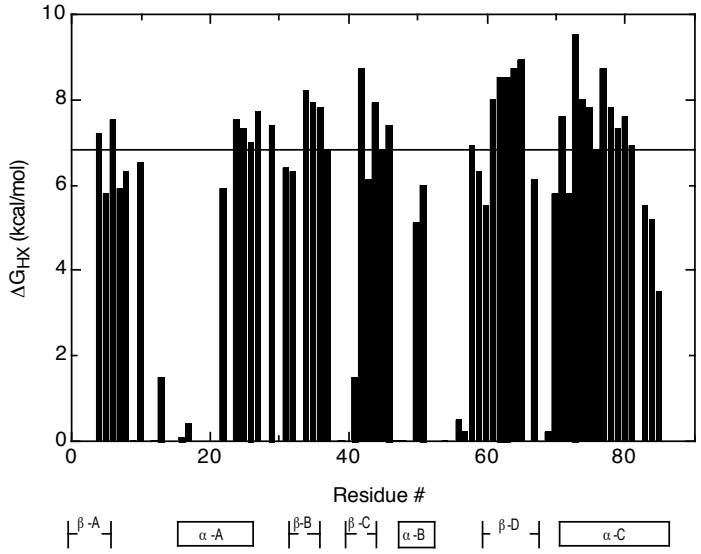

Figure 32. Equilibrium hydrogen exchange of K49E *ec*HPr. The residue specific ΔG values are reported for the K49E HPr protein as represented by the bars. The solid line represents the urea denaturation as monitored by circular dichroism. Secondary structural elements have been indicated below the y-axis. $\Delta G_{HD} = 9.0$ kcal/mol

In this chapter we show that *ec*HPr folds by a non two-state mechanism, which accounts for super protection. While D69A* folds by a two-state folding mechanism.

RESULTS

In this study, we utilized the stabilized variant of *ec*HPr, K49E.[66] This variant is 2.0 kcal/mol more stable than wt *ec*HPr but shows all the features of the wt protein. Here we will call the K49E variant wt*.

Folding kinetics

The folding kinetics for wt* were followed in D_2O and water. It was necessary to investigate the effects of D_2O on the folding pathway because the hydrogen exchange experiments are performed in D_2O and D_2O has been shown to both stabilize and destabilize proteins.[24]

Figure 33 shows the refolding reaction monitored by circular dichroism at 222 nm. In the refolding progression, the unfolded protein (–10 mdeg) refolds and achieves the folded signal (–60 mdeg) in less than 1s. This reaction was fit to a mono exponential function (equation 18) giving us information about the number of phases in the folding reaction, the amplitude, the relaxation time, and the signal at equilibrium. This reaction fits well to mono exponential as evident by the residuals. We see no evidence of proline bond isomerization, which usually causes a biphasic characteristic to the refolding reaction. The amplitudes show that we capture 71% of the reaction.

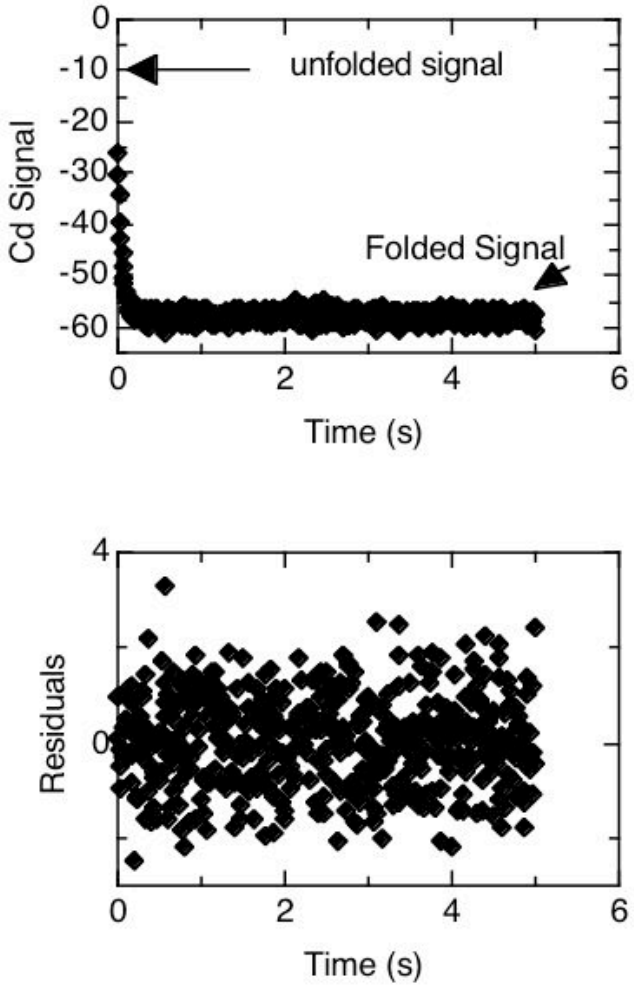

Figure 33. Refolding progress curve for wt*. The progress curve was fit to a mono exponential function. The residuals shown in the lower panel are sufficiently small and random suggesting a good fit.

Unfolding kinetics

Figure 34 shows the unfolding reaction for wt* as followed by circular dichroism at 222 nm. This reaction was performed by doing a rapid dilution of 1 to 10 yielding a final urea concentration 9 M urea and we can follow the progress of unfolding from folded (-60 mdeg) to unfolded (–10 mdeg) protein. The residuals of the monophasic exponential fit were poor (Figure 34), suggesting the presence of multiple kinetic events. When a bi-exponential function (equation 20) was applied, the fit improved, as seen by the residuals. The amplitudes of the reaction show that we capture 90% (the summation of the two amplitudes) of the reaction. The unfolding rate constants in water are 0.002 and 0.0008 per second, only a 2-fold difference. However, when the analogous reaction was performed in D_2O we see a 17-fold difference in the two rate constants (0.008 and 0.00047 per second).

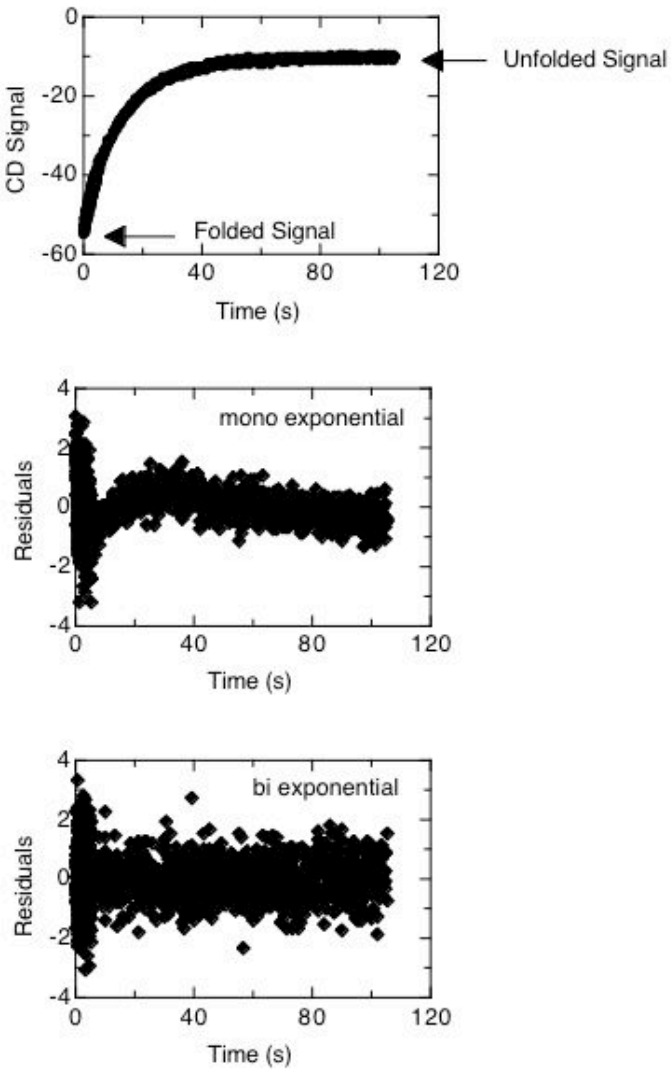

Figure 34. Unfolding progress curve for wt*. The middle graph is the residuals of mono-exponential fit and the fit is poor. However, when a bi-exponential fit was applied, the residuals improved (bottom panel).

Chevron

Figure 35 shows a traditional chevron plot for wt*. The refolding limb shows a linear urea dependence of the observable rate constant. The unfolding limb also shows a linear dependence on urea concentration. The extrapolated folding rate in H_2O is 50/s for wt* and increases to 115/s in D_2O. Since the random coil hydrogen exchange rate constants are on the order of minutes, the hydrogen exchange process adheres to an EX2 mechanism for HPr. For the observable folding rates, a Tanford β value can be determined to be 0.69 suggesting that 69 percent of the accessible surface area was buried prior to the transition state. For comparison the β value for the folding reaction in D_2O was 0.71 suggesting that the transition state is nearly identical in D_2O and H_2O. The ΔG calculated with the two folding phases are 6.0 kcal/mol and 6.5 kcal/mol in water and increases to 7.5 kcal/mol and 8.5 kcal/mol in D_2O (Table 3).

Table 3. Kinetic parameters for wt and wt* variants.

		wt	wt	wt*	wt*	S31A wt*	S31A wt*	wt* D_2O	wt* D_20
k_f	(/s)	11.01		50.64		42.69		115.75	
k_u	(/s)	0.0027	0.0013	0.0020	0.0008	0.0147	0.0053	0.0005	0.00089
m_f	(cal/mol M)	-0.59		-0.74		-0.65		-0.82	
m_u	(cal/mol M)	0.58	0.52	0.37	0.33	0.29	0.26	0.27	0.14
M	(kcal/mol M)	1.17	1.11	1.11	1.07	0.93	0.91	1.09	1.00
C_{mid}	(M)	4.19	4.83	5.39	6.08	5.04	5.83	6.63	5.58
ΔG_k	(kcal/mol)	4.91	5.35	5.99	6.51	4.70	5.31	7.50	8.50
β	m_f/m_{eq}	0.50		0.69		0.71		0.75	
α_f		0.75		0.71		0.67		0.71	
α_{uf}		0.07	0.92	0.09	0.89	0.85	0.15	0.08	0.92

Kinetic parameters were determined from stop flow experiments. The last column contains data collected in D_2O under identical conditions. The thermodynamic parameters calculated from the kinetic data agree well with equilibrium data. The α_f or α_u amplitude was calculated from a 0-9 M jump for unfolding and a 9 M to 0.9 M refolding jump. β_T is the ratio of (m_f/m_{eq}), ΔG was calculated using $-RT \ln(k_f/k_u)$.

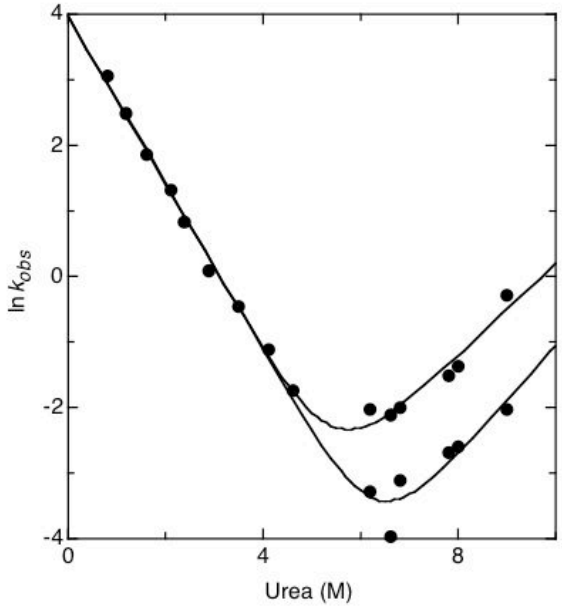

Figure 35. Representative data from the chevron analysis of wt*. The observed rates of refolding and unfolding are plotted against urea concentration. The curve represents the best fit to equation 46.

Equilibrium analysis of kinetic data

Figure 36 shows the equilibrium denaturation of K49E followed by circular dichroism at 222 nm. Overlaid on this plot are the [A_{inf}] or final amplitudes from the kinetic analysis. These values [A_{inf}] agree with the equilibrium values showing that the kinetic reaction has come to equilibrium. This plot also shows us the amplitudes of each kinetic reaction as a function of urea concentration. As one can see, the amplitudes for the reaction are linear with urea concentration and agree with the equilibrium baselines.

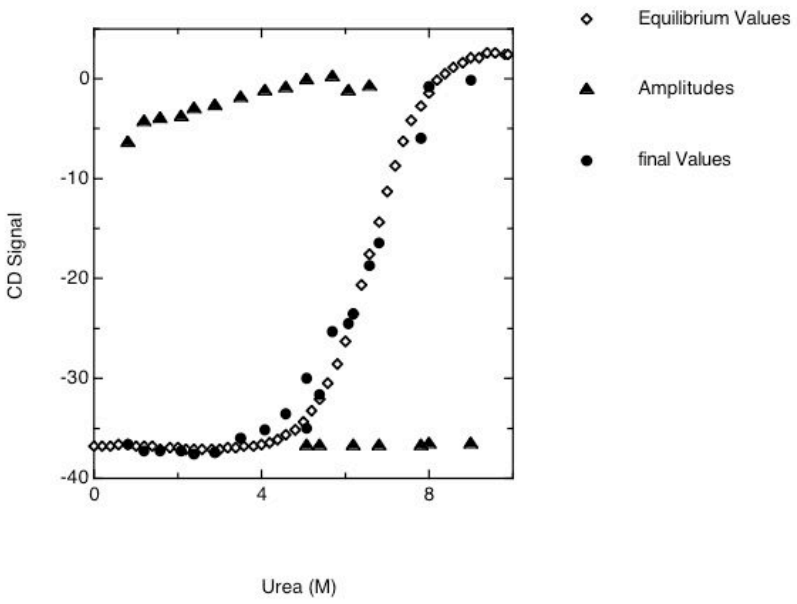

Figure 36. Representative equilibrium analysis of wt*. The open triangles are the equilibrium values determined by circular dichroism. The triangles represent the amplitudes for the folding and unfolding reactions. The unfolding reactions are determined from the summation of the two phases. The closed circles represent the final value as determined from a mono exponential fit of the kinetic data.

Investigation of proline and peptide bond isomerization

Double jump experiments were performed to investigate the possibility of proline or peptide bond-isomerization. The strategy of this type of experiment is to perform an unfolding reaction, wait a delay time, and then refold the protein. This essentially allows the protein to sample the unfolded state for a specified amount of time. If a peptide bond were to isomerize within this delay time, we would see biphasic refolding kinetics and

two amplitudes that reflect the two populations of unfolded molecules. Figure 37 clearly shows that the refolding experiments were all monophasic, for delay times varied from 100 milliseconds to twenty-five seconds. Experiments were also performed with cyclophilin, a prolyl isomerase, and no effects were seen. Therefore, we conclude that no peptide bond isomerization occurs in unfolded HPr.

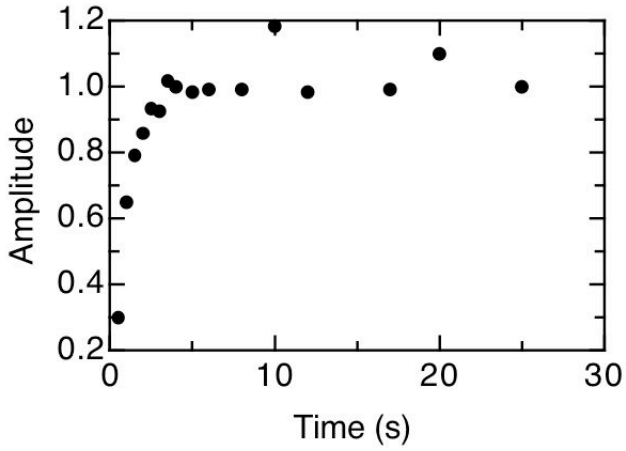

Figure 37. Double jump experiments to determine the presence of multiple unfolded forms. The amplitudes versus delay time are graphed to show the progress of a possible isomerization. Bond isomerization was not detected.

A double jump experiment was performed in order to further investigate the biphasic nature of the unfolding reaction. The biphasic unfolding kinetics suggest either the presence of an unfolding intermediate or two folded states. In this experiment, refolding is initiated for a variable a delay time, followed by an unfolding reaction. If an

intermediate is present, biphasic unfolding kinetics will be seen. As you can see from Figure 38, at no time did the unfolding reaction lose its biphasic character. From this we can determine that if we have an unfolding intermediate it forms in less than 350 ms (dead time of the double jump mixer). However, we cannot conclusively determine if there is an on pathway intermediate, off pathway intermediate, or parallel folding pathway (two folded forms).

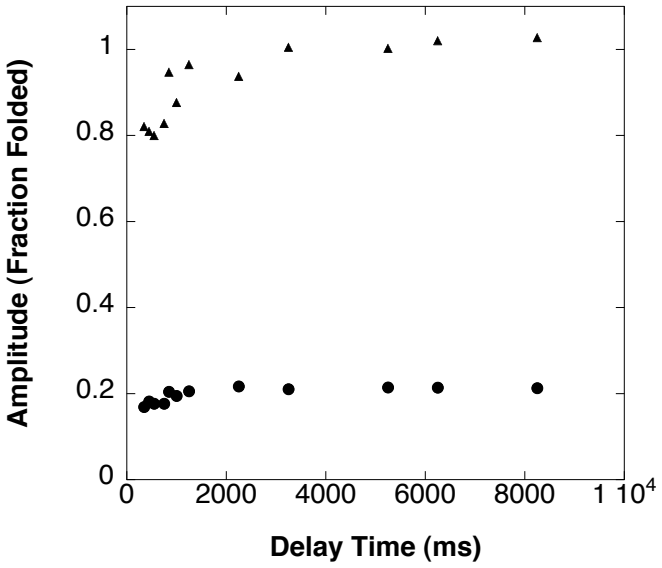

Figure 38. Double jump of wt* U->F->U. This experiment was designed to understand the presence of an unfolding intermediate. The reaction was performed by starting a refolding reaction followed by a delay time then the initiation of an unfolding reaction. Throughout these experiments, the unfolding reaction was biphasic.

Our analysis of the folding mechanism for wt* shows 3 main features: 1) There is no detectable *cis*/*trans* proline or peptide bond isomerization. 2) The hydrogen exchange reaction adheres to the EX2 mechanism. 3) The HPr protein contains an unfolding intermediate that may play a role in super-protection. With this in mind, we took a closer look at the residues which show super protection in the hydrogen exchange experiment. Figure 39 depicts the residues that are super protected as spheres. These residues seem to be grouped in the C helix and around the hydrogen bond network located at the N terminus of the C helix. In an effort to understand why the biphasic unfolding and super-protection exist, we characterized the folding kinetics and hydrogen exchange behavior of the Asp69 mutations.

Figure 40 shows a depiction of the hydrogen bond network generated from the crystal structure of wt *ec*HPr 1poh.pdb. This network contains three hydrogen bonds as designated by the dashed lines. Asp 69 forms hydrogen bonds with the amides of 30 and 31 and the hydroxyl of the serine at position 31. The residues associated with this hydrogen bond network are strongly conserved. Previously, Ser 31 was investigated due to the unique chemical shift of the side chain hydroxyl proton (5.75 ppm). [95] The chemical shift of Ser 31 hydroxyl is perturbed out of the water peak, an unusual event for hydroxyl protons. When Ser 31 was mutated to alanine (S31A) the protein was found to be only 0.5 kcal/mol less stable. Upon further investigation, Asp 69 was found to be no longer forming hydrogen bonds to amides 30 and 31, and Asp 69 was freely rotating and no longer involved in the hydrogen bond network.

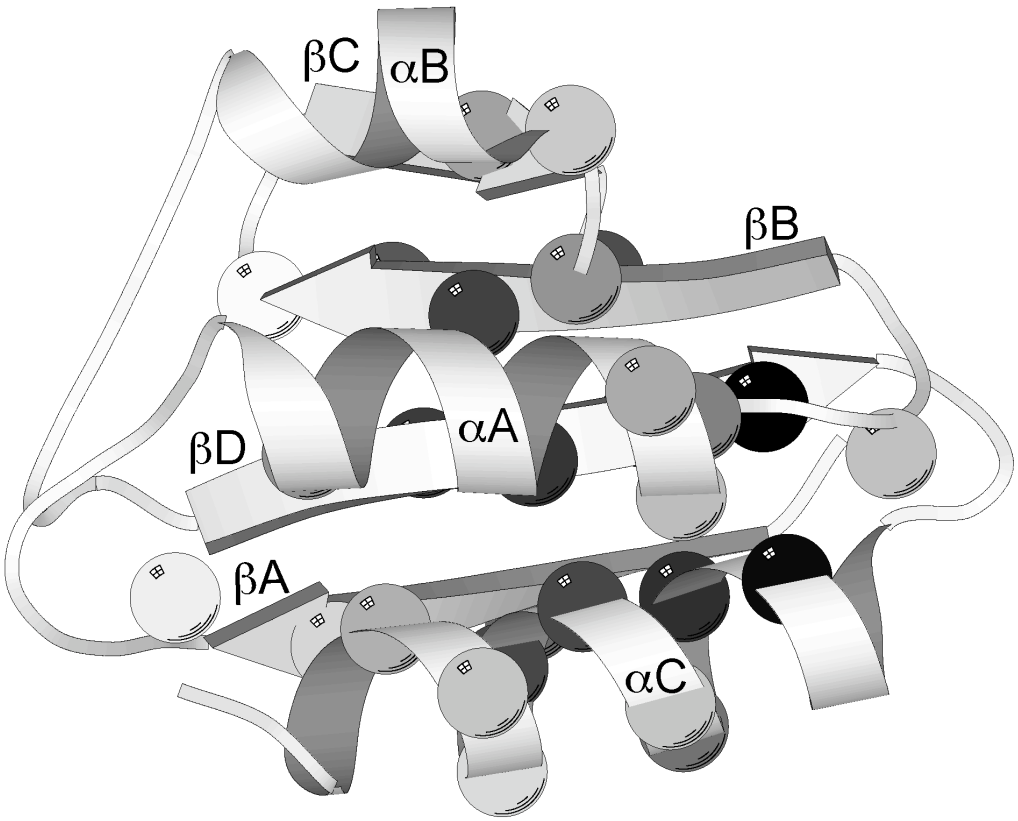

Figure 39. HPr structure with residues that are protected from exchange represented by spheres. The black spheres representing the most stable residues with gradations to white becoming progressively less stable. The C helix contains the greatest number of protected residues per unit of structure.

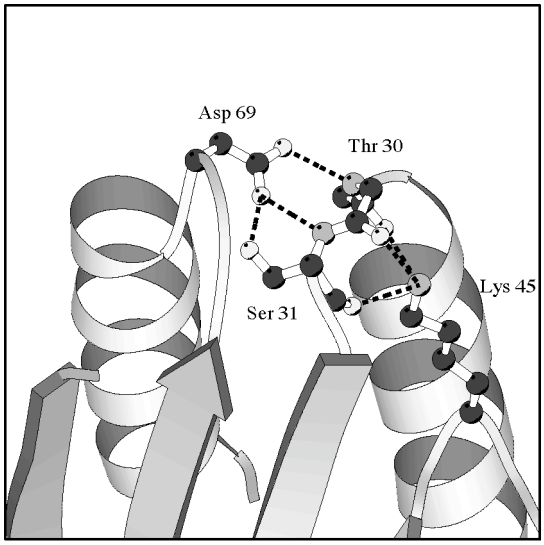

Figure 40. Diagram of the native Asp69 interaction. It makes two side-chain to main-chain hydrogen bonds with the backbone amides of residues 30 and 31 as well as a side-chain to side-chain interaction with Ser 31.

With the understanding that S31A* removes the Asp 69 hydrogen bond network and is only minimally destabilized (0.5 kcal/mol), we characterized the hydrogen exchange and folding kinetics for S31A*. We found that the protein fold by a non-two-state mechanism (Table 3) and super-protection persisted. With this in mind, we investigated the effects of the D69N and D69A mutations using hydrogen exchange and folding kinetics (Table 4). D69A* shows no super protection in the hydrogen exchange experiment. Figure 41 shows the residue specific stabilities determined by H/D exchange in 0.4M $(NH_4)_2\ SO_4$ at pH 7, 17 °C. The solid line represents the stability calculated by urea denaturation under identical conditions. As you can see, the agreement between

stabilities calculated by these two methods is excellent. This mutation also leads to the elimination of the biphasic-unfolding characteristic observed in wt*. The effects of D_2O were also investigated and we found that D_2O increases the folding rate by 1.5 fold and decreases the unfolding rate by fourteen fold just as seen with wt*. There were no significant changes to the *m* value or β value suggesting no significant change in the transition state. The elimination of super protection and the biphasic unfolding suggest that the non-native interaction model is correct.

Table 4 Kinetic parameters for the D69X wt* mutants.

		D69N wt*	**D69A wt***	**D69N S31A wt***	**D69N S31A wt* D_2O**
k_f	(/s)	19.6	39.1	34.0	63.0
k_u	(/s)	0.076	0.36	0.017	0.014
m_f	(kcal/mol M)	-0.68	-0.94	-0.75	-0.8
m_u	(kcal/mol M)	0.38	0.30	0.35	0.3
m	(kcal/mol M)	1.07	1.24	1.1	1.1
C_{mid}	(M)	3.1	2.2	4.1	4.5
ΔG	(kcal/mol)	3.3	2.8	4.5	5.0
β	mf/meq	0.64	0.76	0.68	0.71
α_f		0.69	0.63	0.66	0.66
α_{uf}		0.70	0.72	0.68	0.68

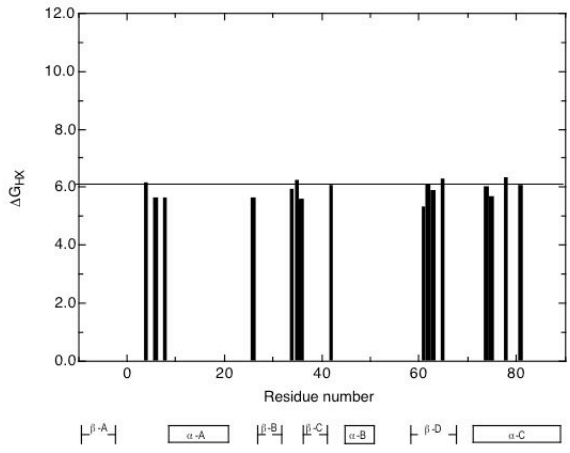

Figure 41. ΔG_{HX} of the D69A wt* in 0.4 M $(NH_4)_2SO_4$. The black solid line represents the urea denaturation performed under identical conditions by traditional urea denaturation.

DISCUSSION

It has long been suggested that the HX technique can be used to determine the conformational stability of a protein [93; 94; 95], and recently it has been shown that conformational stability determined by HD (ΔG_{HD}) is identical to the values measured by more conventional equilibrium unfolding studies (ΔG_U) using either thermal or solvent denaturation, provided the differences in Xaa-Pro isomerization and solvent (H_2O *vs.* D_2O) are included [8; 96]. This conclusion supports the notion that for some residues, the protein must globally unfold for exchange to occur.

In a previous analysis, Huyghues-Despointes, *et al.* [8] concluded that, with one exception, the corrected $\Delta G_{HD} = \Delta G_U$ are within the limits of the error on the two measurements (defined such that $\Delta G_{HD} - \Delta G_U < 1$ kcal mol^{-1}). There was one protein where ΔG_{HD} was substantially larger than ΔG_U; the src SH3 domain [97] showed a 1.4 kcal mol^{-1} difference (6.1 *vs.* 4.7 kcal mol^{-1}, respectively for ΔG_{HD} and ΔG_U). Here, we have shown that ΔG_{HD} for *ec*HPr and *bs*HPr are substantially larger than ΔG_U; we went on to show that *bs*HPr unfolds via a three-state mechanism (Chapter III).

We hypothesized that there are several reasons why ΔG_{HD} might be larger than ΔG_U, even after the correction for the differences in the *cis*/*trans* isomerization of Xaa-Pro bonds has been applied and the measurements of ΔG_{HD} and ΔG_U have been obtained in identical solvents (H_2O *vs.* D_2O) as described previously [8]. In the determination of ΔG_U, several assumptions have to be made. 1) The equilibrium unfolding/refolding must be reversible and exhibit two-state (or a finite number of states) behavior. 2) In the analysis of solvent denaturation studies, one assumes that the linear extrapolation method (LEM) applies and likewise for thermal unfolding experiments, one must have accurate values for ΔC_p in order to extrapolate stability measurements from higher to ambient temperatures. In the analysis of HX measurements, there are additional considerations. 1) Does the exchange process adhere to an EX2 mechanism such that equation (49) can be used to determine ΔG_{HD}? 2) Do the peptide-based values for k_{rc} apply to the unfolded (i.e. exchange competent) form of the protein?

The present study was motivated by our observation that ΔG_{HD} was substantially larger than ΔG_U, as determined by either solvent or thermal denaturation, for the wt HPr

protein from *E. coli*. To discover the origin of this super protection, we performed a detailed thermodynamic and kinetic comparison of the wt HPr protein and a key variant, D69A*. The variant removes a side chain that is involved in a key interaction found in all the HPr proteins studied to date [98]. The D69A mutation is very destabilizing, so we constructed the D69A variant in the background of the stabilizing variant K49E (wt*) [70].

The principal result in this study is that the folding of the wt HPr protein does not adhere to a two-state mechanism. The equilibrium thermal or solvent denaturation curves underestimate the true conformational stability. The HD data, on the other hand, do report on the overall conformational stability because the analysis of the data does not require a model for the overall conformational transition of the protein.

The kinetic data provide convincing evidence for non two-state behavior. The unfolding data for the wt* HP shows two phases. The rate constants for the observable phases change in the typical fashion providing the chevron behavior for the folding and unfolding reactions (Figure 35). In fact, refolding experiments at the C_{mid} show two phases. The ΔG calculated from the extrapolated rate constants agrees with that determined from hydrogen exchange. In addition, the kinetic m values also agree with the equilibrium m value determined from the urea denaturation measurements. Together, these results suggest that the same conformational transition is being monitored by the hydrogen exchange experiments and the observable unfolding reactions in the kinetic experiments.

The unfolding reaction of the wt HPr protein also shows two phases. We have elected to model the three-state character with a model that places the intermediate “on

pathway", but the alternative "off pathway" or "parallel pathway" mechanisms could also work [48]. Since we cannot measure the rate constants associated with this first transition in unfolding, we are not able to distinguish between the potential models. If we assume the slowest unfolding reaction is representative of the total unfolding rate constant and the refolding is representative of the total refolding rate constant we see that $\Delta G_k = 8.5$ kcal/mol and $\Delta G_{HD} = 9.0$ kcal/mol.

In contrast, all the available evidence suggests that the D69A* variant adheres to a two-state folding mechanism. There is no super protection in the HX experiment, and there is only one detectable unfolding phase. Furthermore, all the estimates of ΔG and the measurements of the m-values agree, as expected for a two-state folding mechanism. The HPr protein has been shown to follow a two-state kinetic folding mechanism using GdnHCl as the denaturant [16], although recent studies on Trp variants of the *E. coli* HPr protein suggest deviations from a strict two-state mechanism[126;128]. Further studies will also be necessary to determine if the D69A* mutation itself is responsible for the change in folding mechanism or if any destabilized variant can cause a switch from three- to two-state folding, as has been found for other proteins including ubiquitin [99] and protein G [100].

CHAPTER V

SOLUBILITY NOT STABILITY IS A KEY FACTOR IN AMYLOID FORMATION

INTRODUCTION

The formation of amyloid fibrils is of medical interest because it is associated with diseases such as Alzheimer's disease, Parkinson's disease and cystic fibrosis. [101] Fibrils are formed from the extracellular deposits of aggregated protein. [102; 103; 104; 105] These insoluble aggregates originate from a number of proteins that are structurally diverse; some twenty disease-causing proteins have been identified in humans. Amyloid formation is also of interest to the protein chemist because it has been shown that many proteins can be induced to form amyloid fibrils *in vitro*, suggesting that amyloid fibril formation is not limited to a select few proteins or by specific physiological factors, but rather it may be a general feature of all polypeptides. [106; 107]

Recently proteins such as human muscle acylphosphatase, hen egg white lysozyme, the B1 binding domain of protein G, the SH3 domain of PI3, and myoglobin have been shown to form fibrils *in vitro*. [108; 109; 110; 111] These fibrils show ultra-structure and dye binding characteristics identical to fibrils formed from the disease causing proteins. [112] In most cases, fibrils were obtained from solution conditions where the native conformation was destabilized by pH, temperature, and/or the addition of co-solvents like alcohols,

salts or metal ions. These results suggest that the ability of a protein to form fibrils *in vitro* depends upon the conformational stability of the protein, defined as the difference in free energy (ΔG) between the folded and unfolded conformations. [107; 110; 113; 114] It further implies that as proteins are destabilized, either through alterations in amino acid sequence or changes in the solution conditions, the formation of amyloid fibril is enhanced. Here, we wish to expand these conclusions to include the idea that the solubility of the folded (or partially folded) conformation(s) of a protein is important, and perhaps the major factor in amyloid fibril formation.

A significant population of partially unfolded protein is likely necessary for fibril formation to occur in many of the proteins studied to date. [101; 107; 110; 114; 115] For example, fibrils of acylphosphatase are formed in mixtures of water and 2,2,2-trifluoroethanol (TFE), an organic solvent that has been shown to promote secondary structure formation at low concentrations. Since TFE is known to increase the solubility of hydrophobic groups while decreasing the solubility of peptide groups [116; 117], structures that have exposed hydrophobic groups and buried peptide groups are favored. Other studies of fibril formation with transthyretin (TTR), [118] human lysozyme, [119] the light chain of IgG, [120] methionine aminopeptidase [121] and fibronectin [122] emphasize that conditions that favor partial unfolding of the protein can significantly enhance fibril formation. Together, these results also support the hypothesis that any protein can form amyloid under certain conditions, and fibril formation is accelerated when the conditions favor partial unfolding of the native conformation. [123]

While protein stability is certainly important for fibril formation, the association between protein solubility and fibril formation has not been directly correlated. The solubility of globular proteins is determined by the amino acid content, the pK values of the ionizable residues, and environmental factors such as temperature, pH and the presence of co-solvents. [124] Proteins and polypeptides are generally least soluble at pH values near their isoelectric point (pI), where the overall net charge is zero. [124]

The purpose of this study is to investigate the relationship between stability, solubility, and fibril formation using ribonuclease Sa (RNase Sa) from *Streptomyces aureofaciens* as a model system. RNase Sa is a small acidic protein (96 residues, pI = 3.5) with a mixed α+β structure. There are no lysine residues, and the protein contains a single disulfide bond. [125] We have recently made variants that replaced solvent-exposed acidic residues with lysine residues. [125] The mutant with three such charge reversals is denoted 3K and that with five charge-reversals is 5K. By reversing the charge at these sites, we have changed the pI of RNase Sa from 3.5 to 6.4 (3K) or 10.2 (5K).[125] Thus we are able, here, to study the relationship between stability, solubility, and fibril formation in a single protein over an exceptionally wide pH range.

RESULTS

We have investigated the propensity of RNase Sa and two variants (3K and 5K) to form fibrils in the presence of TFE and as a function of pH. Figure 42 shows the CD spectra for RNase Sa in water and 30% TFE. The presence of TFE induces significant β-structure in the protein at both pH 7 and 3.5. As shown previously, the formation of β-

structure appears to be a hallmark of amyloid formation and it is clear that 30% TFE induces substantial β-structure in RNase Sa.

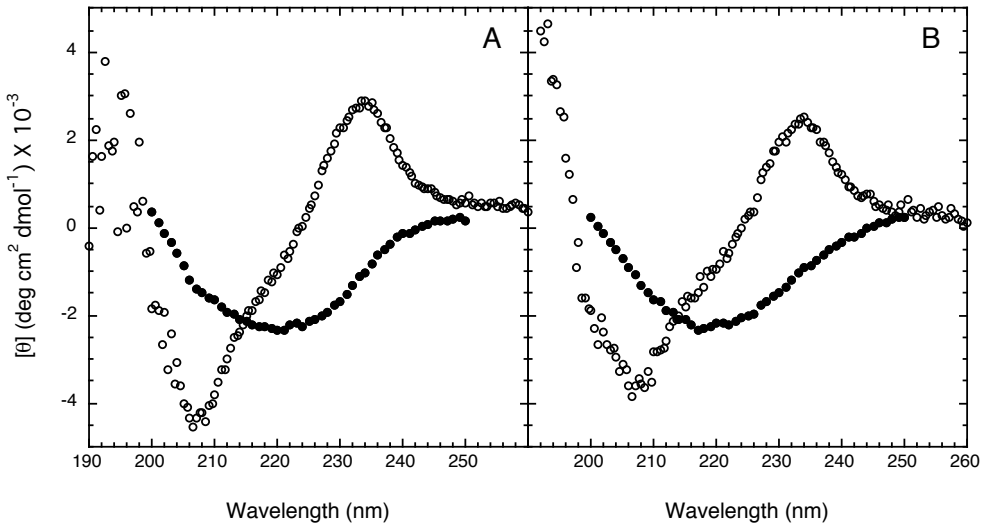

Figure 42. Far-UV CD spectra for RNase Sa. Representative spectra are shown for the protein in 0% TFE (open circles) or 30% TFE (filled circles) at pH 7 (A) and pH 3.5 (B). The spectra were obtained after 5 hr incubation at 25°C in 10 mM (citrate, phosphate, and borate) buffer. The presence of TFE clearly induces a substantial amount of β-structure in RNase Sa at both pH values.

The enhancement in the fluorescence intensity of Thioflavin T (ThT) upon binding to ordered protein aggregates is one method to show the presence of fibril formation. [114] Figure on page 118 shows ThT fluorescence as a function of TFE concentration for wild-type RNase Sa. At pH 5.5, no fluorescence is observed regardless of TFE

concentration. At pH 3.5, the fluorescence increases by ≈ 15 to 20-fold from 10 % TFE to 35 % TFE. Above 35% TFE, ThT fluorescence drops markedly, and CD spectra of the samples show a change from predominantly β-structure to substantial α-helical structure (Figure 42). This change in secondary structure at high TFE concentrations has also been observed with other proteins. [114]

When RNase Sa in placed in 30 % TFE at pH 3.5, ThT fluorescence was not observed for the first 2-3 hr (Figure 43). After 6 hrs, the fluorescence intensity reached a maximum and did not change over a period of several days, at which time fibril formation was clearly evident. The isolated fibrils have the same morphology, dye-binding properties, nucleation-dependent kinetics, and ultra-structure as those found for a number of other proteins. Figure 44 shows electron micrographs of the fibrils formed from RNase Sa. In contrast, at pH 7.0 in 30% TFE, no ThT fluorescence was observed and no fibrils formed, even after several days of incubation.

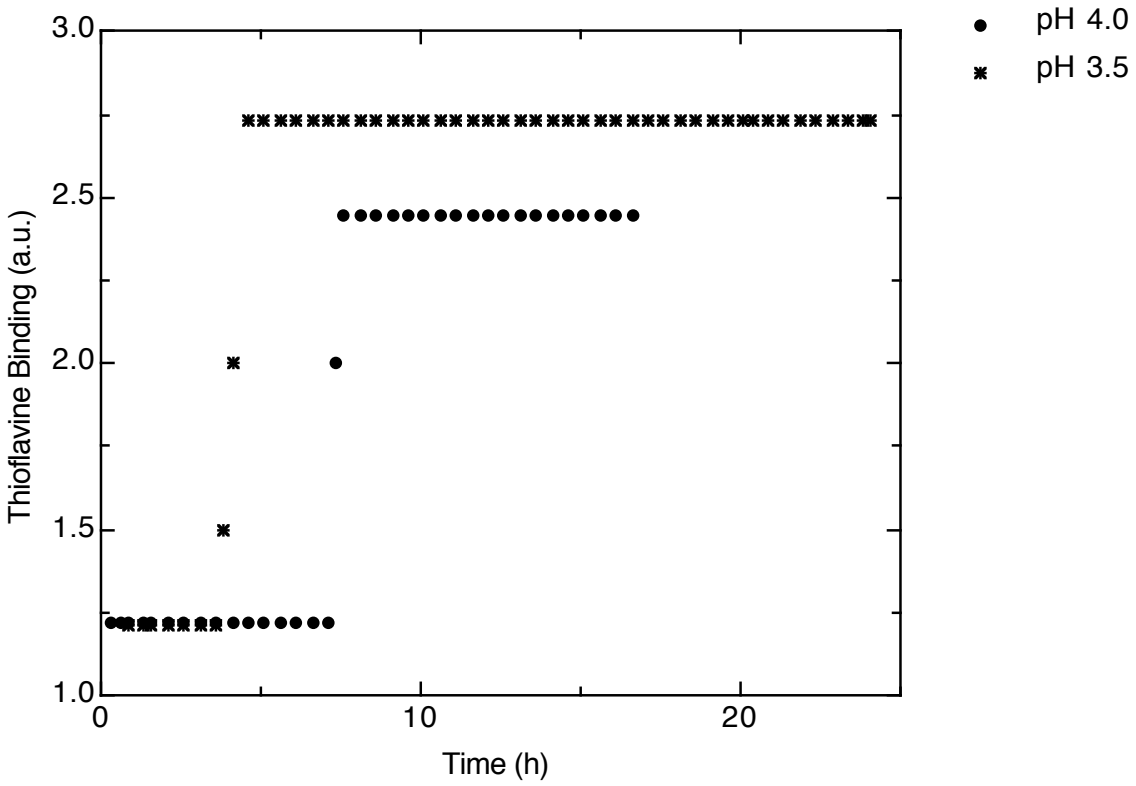

Figure 43. The kinetics of amyloid fiber formation. Fiber formation was monitored by thioflavine T binding. Samples were incubated in 30% TFE at two pH values (3.5 and 4.0). Thioflavine T binding was determined by removal of sample from the incubation at time intervals of 30 min and binding was determined as stated in the materials and methods.

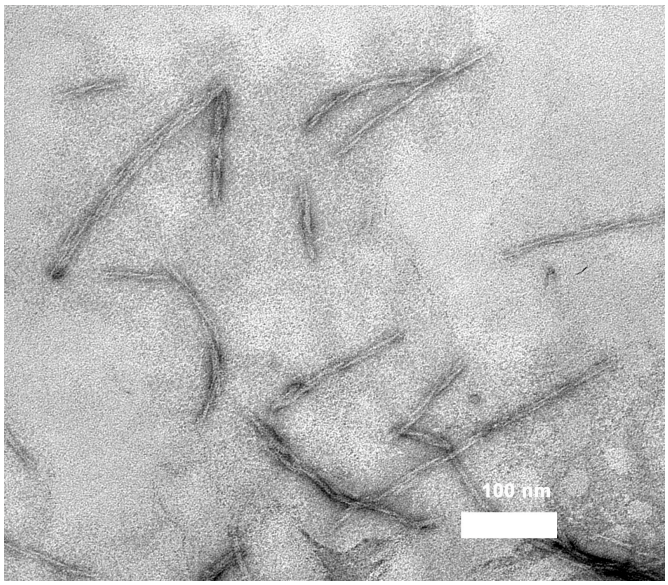

Figure 44. Electron micrographs of amyloid fibers of RNase Sa. The images were acquired using a Zeiss EM 10C TEM transmission electron microscope at 80 kV excitation voltage. Samples were incubated for 28 days at room temperature in 30% TFE, 50 mM acetate, phosphate, borate buffer pH 3.5 at a protein concentration of 0.375 mg/mL. A 3 μL sample was placed on a formvar and carbon coated grid. The sample was then negatively stained with 30 μL of 1% uranyl acetate and observed at a magnification of 50,000 X. The scale bar represents 100 nm making the approximate width of these fibers 12 nm and length approximately 100 to 200 nm.

The results shown in Figure 42 led us to study the effects of pH on fibril formation with our charge-reversal variants. RNase Sa, 3K and 5K were incubated in solutions of 30% TFE ranging in pH from 2 to 9.6. Figure 46 shows the correlation between fibril formation, conformational stability, and solubility for these proteins as a function of pH. All three variants are maximally stable near pH 5, but the minimum solubility of the proteins shifts with their pI, from near pH 3.5 for RNase Sa to pH > 9 for the 5K variant. [125] Fibril formation is most prominent at the pH where solubilty is minimal, i.e., near the

pI of the protein. These results show that fibril formation correlates with the pH-dependence of the protein solubility and not with conformational stability for all these RNase Sa variants. These results very strongly show that protein solubility is a major factor in fibril formation.

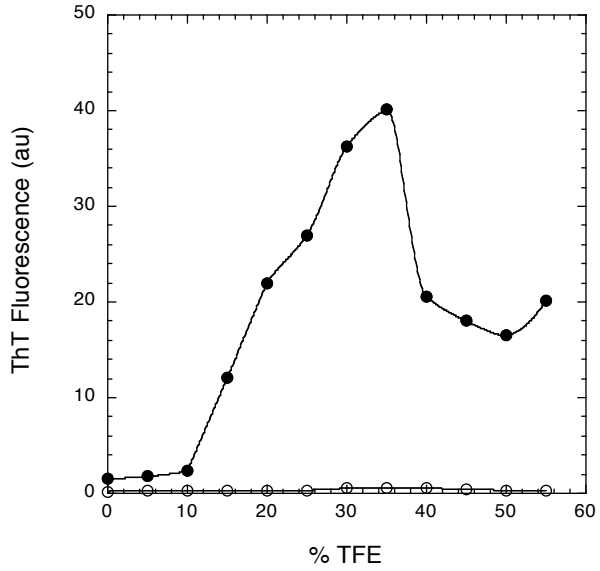

Figure 45. Fibril formation (monitored by ThT fluorescence) as a function of TFE concentration. The curves represent the changes of fluorescence with TFE concentrations at pH 3.5 (filled circles) and pH 5.5 (open circles) after 18 hr of incubation. ThT binding was monitored after removing an aliquot of the aggregation mixture (10 μL) and diluting it with 990 μL of 50 mM sodium phosphate, pH 6.0, containing 3 μM ThT. Fluorescence intensity was determined at 485 nm after excitation at 440 nm using a 96 well plate reader equipped with cutoff filters as previously described. [126]

We have measured the pH-dependence of fibril formation for a number of other proteins and summarized data from other studies. Table 5 shows the pH values and the

maximum conformational stability, pI, and the pH of maximum fibril formation for these proteins. Figure 47 shows the correlation between the pH of maximal amyloid formation and the pI of the protein. The agreement between pI and pH of maximal fibril formation (pH_{max}) is excellent, whereas there is little correlation between the pH-dependence of fibril formation and conformational stability. The data for α-synuclein, transthyretin, and the Aβ peptide are especially significant, since these proteins are known to form amyloid fibrils *in vivo*. [118; 126; 127]

In conclusion, our results suggest that the solubility of a polypeptide chain is a major factor that determines the *in vitro* conversion of globular proteins into amyloid fibrils under non-pathological conditions and may contribute to fibril formation *in vivo*. The results also suggest that increasing the solubility of a protein should prevent the formation of fibrils. We are in the process of testing this hypothesis with RNase Sa and other proteins.

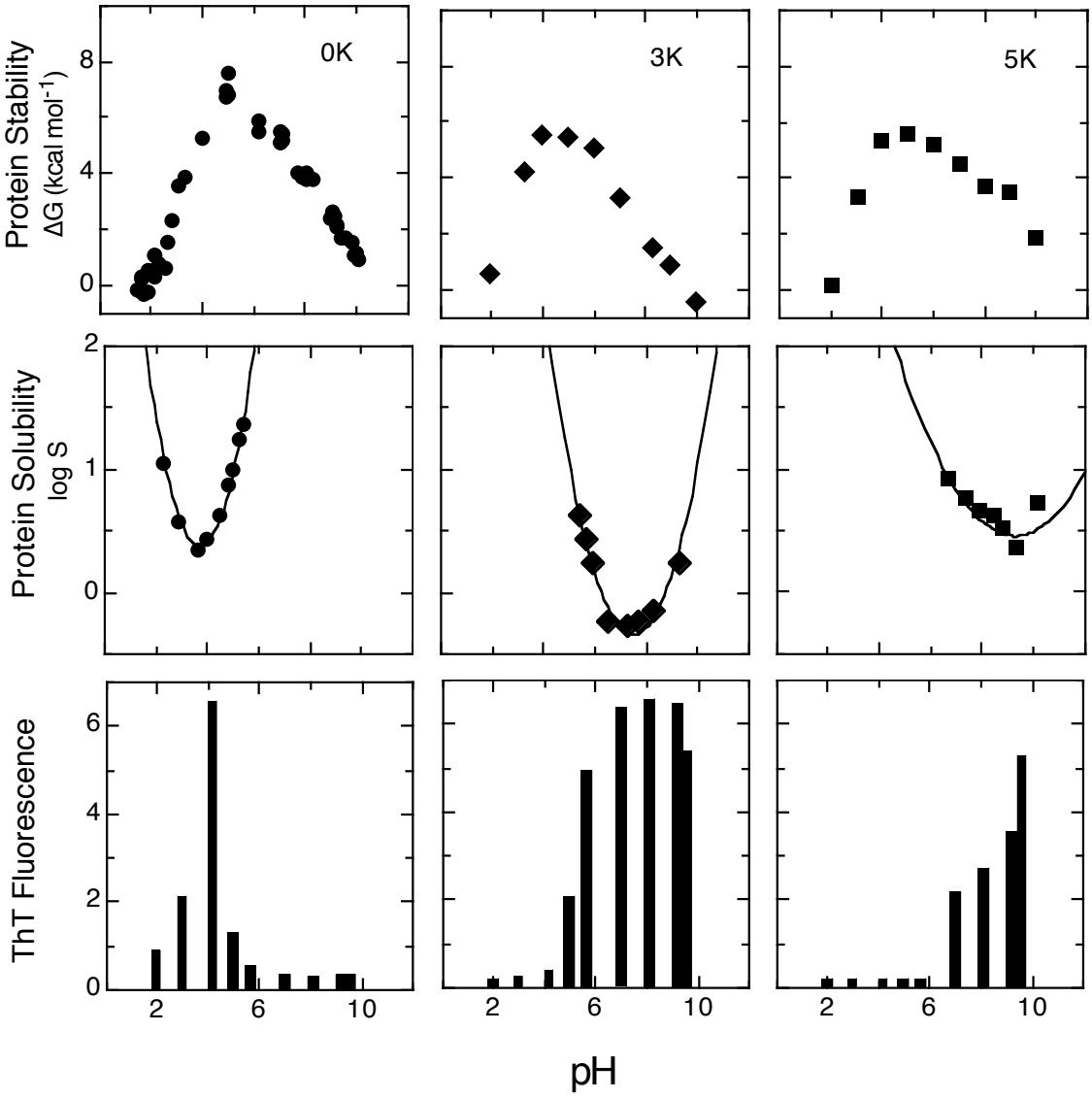

Figure 46. The conformational stability, solubility, and fibril formation are shown as a function of pH for wild-type RNase Sa, and the 3K and 5K variants. The proteins were purified as described previously, [128] and the pH dependence of stability, the pI determination and solubility have been described in more detail in an earlier report. [125] The solubility data are expressed as log S, where S is the solubility in mg mL^{-1}. The lines through the data are meant only to guide the eye. The amyloid formation data were obtained from the ThT fluorescence assay as described in Figure 45. The samples were incubated at room temperature in 30 % TFE at the indicated pH before measuring ThT fluorescence.

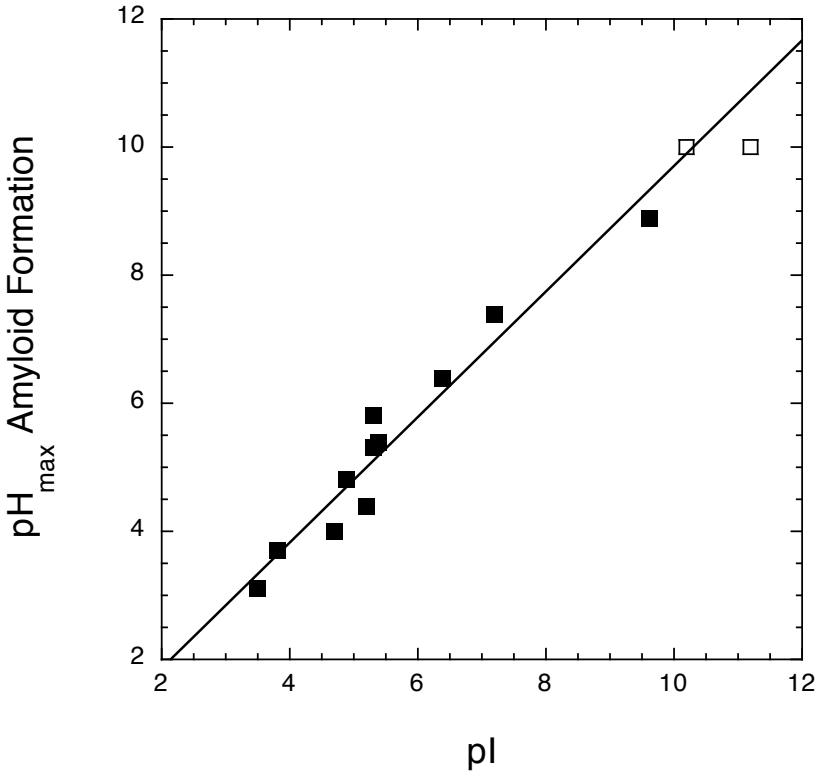

Figure 47. The correlation between the pH of maximal amyloid formation and the pI of the proteins listed in Table 5. The open squares represent the 5K variant of RNase Sa and hen egg white lysozyme (HEWL). These two proteins have pH_{max} values greater than 10 and could not be determined. The correlation coefficient for the 11 proteins shown as solid squares is 0.98.

Table 5. The relationship between the pH-dependence of stability, pI and the pH of maximum fibril formation for several proteins.

Protein	ΔG_{max} (pH)[i] (kcal mol^{-1})	ΔG_{pI}[j] (kcal mol^{-1})	pI[k]	pH_{max}[l]
RNase Sa[a g]	7.0 (5.0)	4.0	3.5	3.1
RNase T1[b]	8.8 (4.5)	8.0	3.8	3.7
α-synuclein[c]	NA	NA	4.7	4.0
*bs*HPr[d]	4.5 (7.0)	3.5	4.9	4.8
TTR[e]	NA	NA	5.2	4.4
RNase Sa2[a]	4.7 (4.5)	4.2	5.3	5.8
Aβ peptide[f]	—	—	5.3	5.3
*ec*HPr [d]	4.5 (7.0)	3.5	5.4	5.4
RNase Sa 3K[g]	5.5 (5.0)	4.2	6.4	6.4
RNase Sa3[a]	8.9 (5.5)	4.0	7.2	7.4
RNase A[b]	9.2 (7.0)	9.0	9.6	8.9
RNase Sa 5K[g]	5.5 (5.0)	2.0	10.2	>10
HEWL[h]	NA	NA	11.2	>10

[a] Stability and pI values are from Pace, et al. [129] and the pH of maximum fibril formation was determined as described in the text.

[b] Stability and pI values are from Pace et al. [130] and the pH of maximum fibril formation was determined as described in the text.

[c] The pI was calculated as described in Shaw, et al. [125] and the pH of maximum fibril formation is from Hoyer, et al. [127]. There are no entries for stability, since this is not a globular protein.

[d] Unpublished results from this laboratory (JPS and JMS).

[e] The pI was calculated as described in Shaw, et al. [125] and pH of maximum fibril formation is from Jiang, et al. [118]

[f] The pI and pH of maximum amyloid formation are from Wood, et al. [126]

[g] Stability and pI values are from Shaw, et al. [125] and the pH of maximum fibril formation are from Figure 46.

[h] Stability and pI are from Pfeil and Privalov [131] and the pH of maximum stability was determined as described in the text.

[i] ΔG_{max}(pH) is the maximum conformational stability, which occurs at the indicated pH, expressed in kcal mol^{-1}.

[j] ΔG_{pI} is the conformational stability at pH=pI for the indicated protein, expressed in kcal mol^{-1}.

[k] The pI value for the indicated protein, either measured or calculated as indicated.

[l] pH_{max} is the pH where fibril formation is maximal, based on the ThT assay described in the text.

CHAPTER VI

PURIFICATION OF AN AMYLOID INTERMEDIATE

INTRODUCTION

Formation of amyloid is of medical interest because it is associated with diseases such as Alzheimer's, Parkinson's, and Cystic Fibrosis. [123; 132] Recently, it has been shown that amyloid fibrils that have similar structural[109; 111] and toxicological properties [133] can be formed *in vitro* from a structurally diverse group of proteins that are not associated with any known disease. This suggests that fiber formation may be general feature of all polypeptides. [108] A general correlation between the ability to form amyloid and protein stability has been proposed. [101; 110; 114; 118; 120; 134; 135; 136] We recently showed that while stability is important, protein solubility also plays a key role in fiber formation. We showed that the fibril formation is most prominent at the pH where solubility is minimal, i.e. near the pI of the protein. The kinetics of fibril formation also showed that at the non-permissive pH, amyloid formation was extremely slow.

While it was thought that the aggregates may play an important role in the disease through recruitment of soluble proteins [137; 138] recent results suggest that the intermediates are the toxic species. It is thought that these intermediates form a pore that causes disruption of the cell membrane leading to cell death. [139] Solublization and

purification of any intermediates leading to fibril formation will be of paramount importance to understanding the toxicity as well as the oligomerization process.

While the induction of amyloid is well documented, the isolation of lower order oligomers is rare. To date, I only know of three such isolations; a transthyretin dimer [140], RNase A dimer [141; 142] and a PrP dimer[143]. Unfortunately, only the transthyretin intermediates were tested to see if they could nucleate further fiber formation. Our goal is to isolate an on-pathway stable and soluble intermediate in the oligomerization process.

In this chapter, we show how to purify the amyloid intermediates of ribonuclease Sa (RNase Sa). We discuss purification and solublization strategies that could be applied to other proteins. We show that these intermediates are capable of binding amyloid-specific dyes (Congo Red (CR) and Thioflavine T (ThT)), have a β-rich structure and are capable of seeding the formation of amyloid.

RESULTS

Figure 48 shows a gradient SDS gel (6-15%) in Tris glycine of samples that have been induced to form amyloid (*ec*HPr, 5K, Sa3, *bs*HPr, 3K, Sa and T1). Amyloid was induced by exposure to 30% TFE for twenty eight days at pH = pI of the protein. Amyloid formation was confirmed by CR binding, ThT binding, and electron microscopy. Samples were then boiled for 5 min in 2X sample buffer that contains SDS and β−mercaptoethanol, loaded on the gel and later stained with a coomassie stain.

Interestingly, the gel resolved the higher molecular weight oligomers (Figure 48). Table 6 shows the molecular weights of the aggregates as determined from their mobility through the gel. [75] As one can see from the gel as well as the calculated molecular weights, the bands appear to be increasing by ~10 kDa , as one would expect from oligomerization. Several proteins show a high molecular weight band that was unable to migrate into the gel; we suspect this is the full length amyloid fiber.

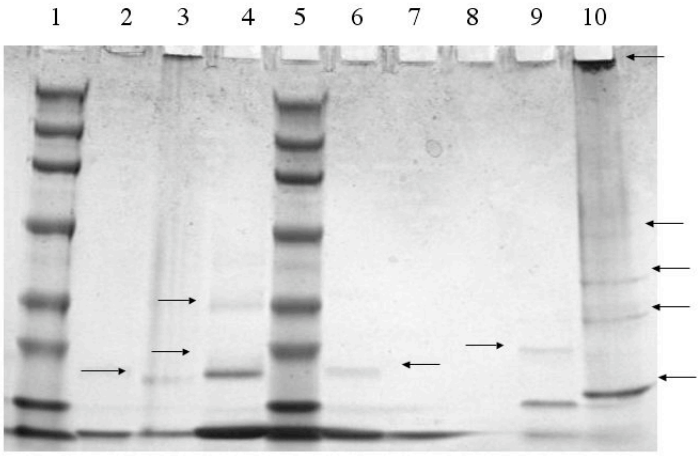

Figure 48. 6-15% SDS-PAGE of amyloid intermediates for a number of proteins. Samples were induced to form amyloid in 30% TFE at the pH where pH = pI for 28 days then pelleted and suspended in 2X sample buffer and boiled for 5 minutes. The arrows indicate bands of interest. Calibrated molecular weight standards were loaded in lanes 1 and 5 and used for molecular weight determination: myosin 209,000, β-Galactosidase 124,000, BSA 80,000, Ovalbumin 49,100, Carbonic Anhydrase 34,800, Soybean trypsin inhibitor 28,900, lysozyme 20,600, and Aprotinin 7,100. The other lanes are as follows : Lane 2 = *ec*HPr, Lane 3 = 5K Sa, Lane 4 = Sa3, Lane 5 = MWM, Lane 6 = *bs*HPr, Lane 7 = *ec*HPr, Lane 8 = 3K, Lane 9 = wt Sa, Lane 10 = T1. The molecular weights were calculated and are presented in

Table 6**.**

Table 6. Molecular weights determined from SDS gel mobility.

Lane	Protein	MW (kDa)	Oligomerization state	Molecular Weight (kDa)
1	MWM			
2	*ec*HPr			9.2
		10.6	1	
3	5K			10.6
		18.2	2	
		11.4	1	
4	Sa3			11.0
		30.1	3	
		18.2	2	
		11.0	1	
5	MWM			
6	*bs*HPr			9.2
		20.8	2	
		11.4	1	
7	*ec*HPr			9.1
		10.6	1	
8	3K			10.6
		8.7	1	
9	Sa			10.6
		31.2	3	
		20.8	2	
		10.6	1	
10	T1			11.1
		57.1	6	
		38.1	4	
		28.2	3	
		10.6	1	

The molecular weights were calculated from a linear regression of the calibrated molecular weight standards vs. the migration distances. [75] Of special note is that the oligomers were resistant to the denaturation process and the molecular weights seem to double, suggesting oligomerization. The molecular weights and tyrosine content were calculated from sequence data**.**

We found that oligomer visualization was hampered by the traditional SDS-PAGE gel system with any alcohol-based coomassie brilliant blue staining. Therefore, we sought to find a gentler method of separation and visualization. Aggregates were suspended in 10% glycerol and loaded on a 8% acrylamide native gel. [75] The gel was stained with a water-based coomassie brilliant blue stain. Figure 49 shows a comparison of the SDS-PAGE and the native gel system. The native gel shows enhancement for higher oligomeric states compared to the SDS-PAGE gel system (Figure 49). At least 5 oligomers can be identified from the native gel system whereas only 2 can be identified from the SDS PAGE system for wt Sa and its aggregates.

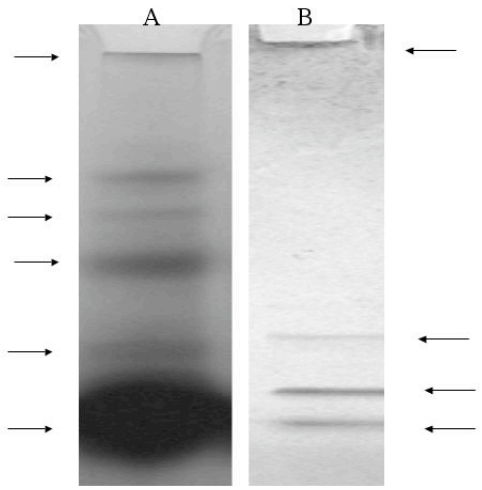

Figure 49. Comparison of native gel and SDS-PAGE of amyloid from wt Sa. A represents the native gel system developed to visualize the oligomers. At least 5 independent bands can be identified (arrows) from the native gel with water base staining as compared with the SDS-PAGE system (B) where only three bands can be identified. The residual higher order oligomers can be seen in the wells as denoted by the arrows.

Scaling up the reaction

All the previous amyloid induction reactions were performed in a volume of no more than 100 μL. With this in mind we investigated the concentration dependence of amyloid formation to determine the optimal protein concentration for at amyloid formation. Figure 50 shows the concentration dependence of amyloid formation. The data show that at 30 μM the amount of amyloid produced from the reaction is maximal. This maximal fiber formation is most likely becuase we have reached the maximal solbulity limit of RNAse Sa in 30% TFE at pH 3.5. Given the maximal solubility of RNAse Sa is ~2 mg/mL at pH 3.5 in water it seems likely that we have reached a solubility limit not an amyloid formation limit. We need to measure the solubility of native RNase Sa and of the amyloid fibrils in 30 % TFE to verify this hypothesis.

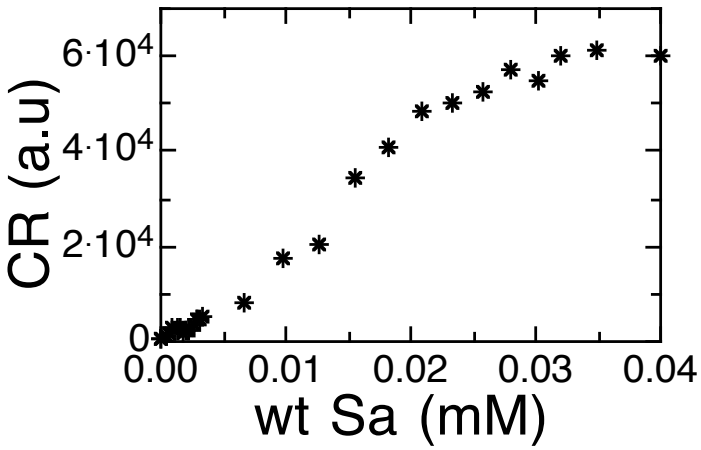

Figure 50. The concentration dependence of amyloid formation of wt Sa. These experiments were carried out at pH 3.5 30% TFE in 100 μL volumes to determine the optimal protein concentration for amyloid induction.

Aggregates were pelleted and 12 mg were suspended in a 10% Glycerol, Tris-Glycine pH 8.0 and separated on a mono Q anion exchange column using a linear NaCl gradient.

Figure 51 shows the elution gradient and absorbance profile. The peaks are labeled according to their mobility on a native gel as well as their dye binding properties (Figure 50). Good resolution of the monomer, dimer and trimer occurred. From the absorbance and volume of each peak, we calculated the amount of each pure oligomer recovered, and estimated the popultation of intermdiates Table 7. If we take these values as representations of the soluble concentrations of intermediates and higher order oligomers, we have approximately 30% conversion to oligomers. This is in fair agreement with 25% conversion calculated based on CR binding assays.

Table 7. Distribution of oligomers of RNase Sa recovered from the ion exchange column.

	Concentration of Sa recovered From Mono Q (mg)	**Population of Intermediates recovered from Mono Q (%)**	**Population of Intermediates from Gel (%)**	**Recovery (%)**
Monomer	**8**	**66.7**	**66.4**	**100**
Dimer	**0.27**	**2.3**	**3.7**	**62**
Trimer	**0.71**	**5.9**	**16.3**	**36**
Tetramer	**0.27**	**2.3**	**4.3**	**53**
Other		**22.9**	**10**	

The concentration of Sa recovered was determined from OD_{278}. The % population of intermediates recovered was determined from the 12 mg of solublized aggregates that were loaded onto the mono Q column. Population of intermediates from the native gel was determined using densitometry (

Figure **51**). Percent recovery was determined from the native gel data and population of intermediates recovered from the mono Q. These percentages are relative measures of the effectiveness of the mono Q column.

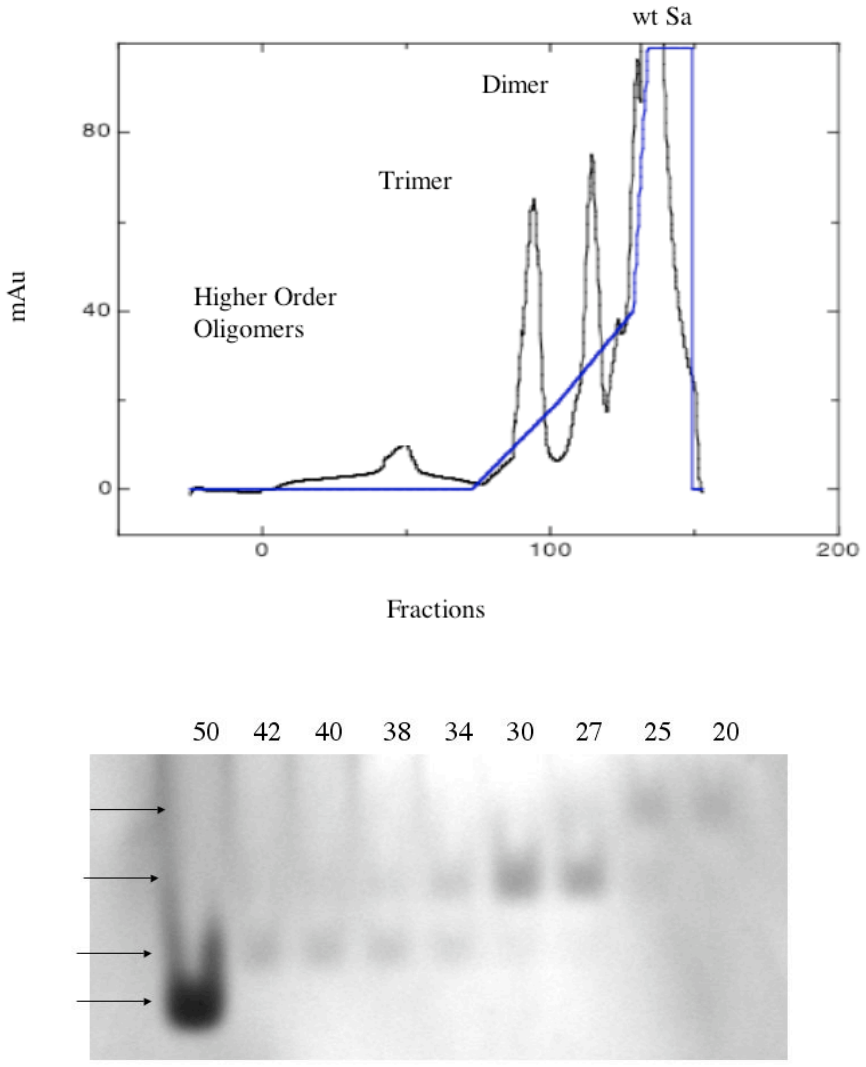

Figure 51. Elution profile and native gel of the amyloid purified from a mono Q column. The upper graph shows the elution profile from the mono Q column followed by absorbance at 254 nm. The solid line represents the salt gradient from 0 to 20% of a 1 M NaCl 10% glycerol Tris/Glycine elution buffer (%B). The lower gel shows the native gel

of the elutions from the mono Q column. Gel lanes are labeled to indicate the fractions (3 mL) collected from the column. The arrows indicate the different molecular weight species i.e. monomer, dimer, trimer, and tetramer.

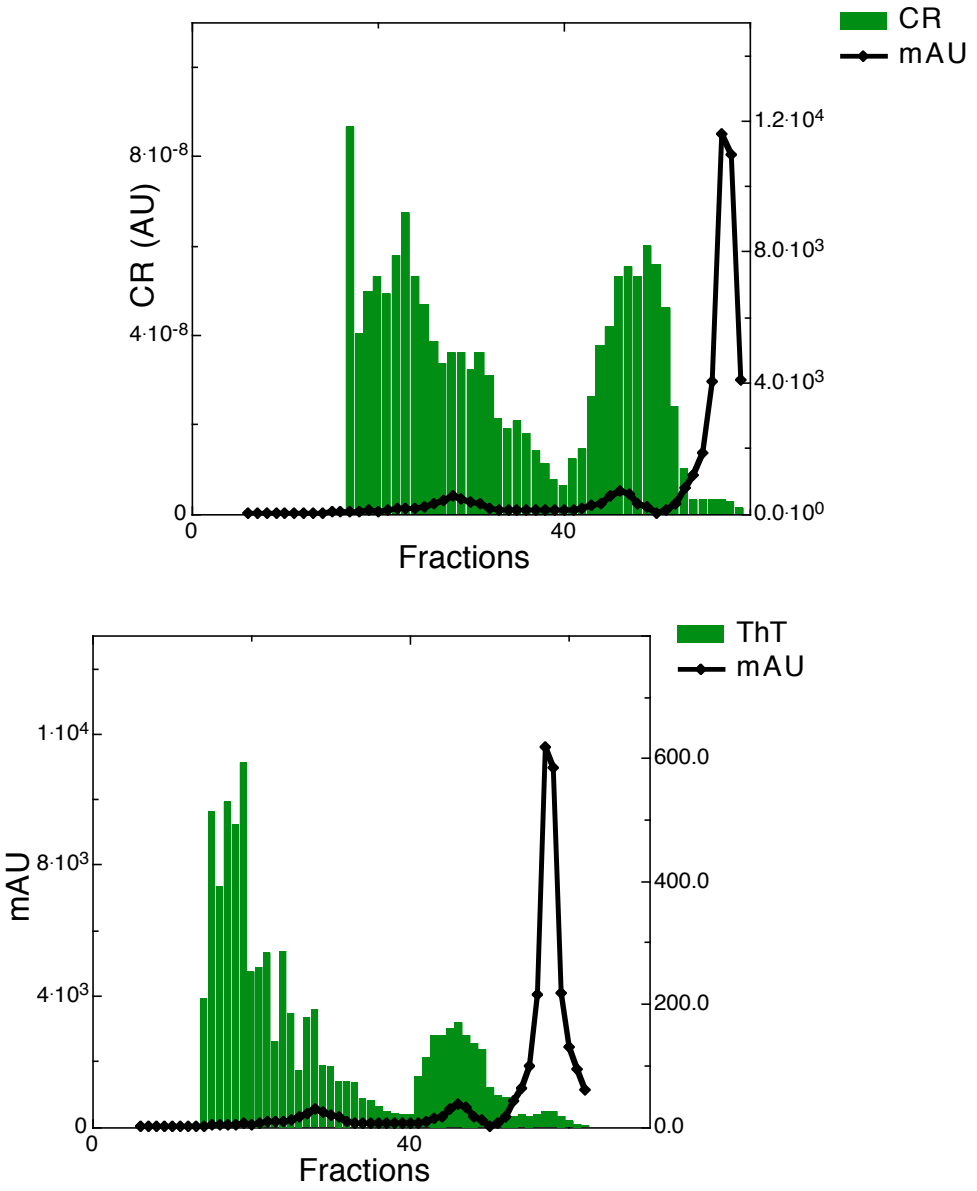

Figure 52. Dye binding properties of the amyloidgentic intermediates. Both graphs have the elution profile from the mono Q represented by the solid black line. The bars represent the dye binding abilities of the elution fractions. Results were normalized for protein concentration. The oligomeric states are represented as 1, 2, 3, 4 as determined from their mobility on a native gel.

Some of the hallmarks of amyloid formation are that fibers bind CR and ThT. Fractions collected from the mono Q elution were tested for CR and ThT binding. Figure 52 shows CR and ThT binding and compares this with the absorbance reading from a mono Q elution. The CR and ThT binding were normalized for protein concentration. The elution can be divided up into four major sections. Monomer (1), Dimer (2), Trimer (3), and Higher Order Oligomers (4). All of the oligomeric species isolated bind both Congo Red and Thioflavin T.

Spectral characterization of the dimer

Figure 53 shows the far-UV CD spectra of the monomeric Sa and dimeric Sa recovered from the mono Q column in 10% glycerol pH 8.0 Tris/Glycine. In the far-UV CD spectrum, the monomeric wt Sa recovered from the mono Q has a typical spectrum of the folded protein with a peak at 234 nm, whereas the dimer has a minimum at 217 nm indicative of a β-rich structure. [11]

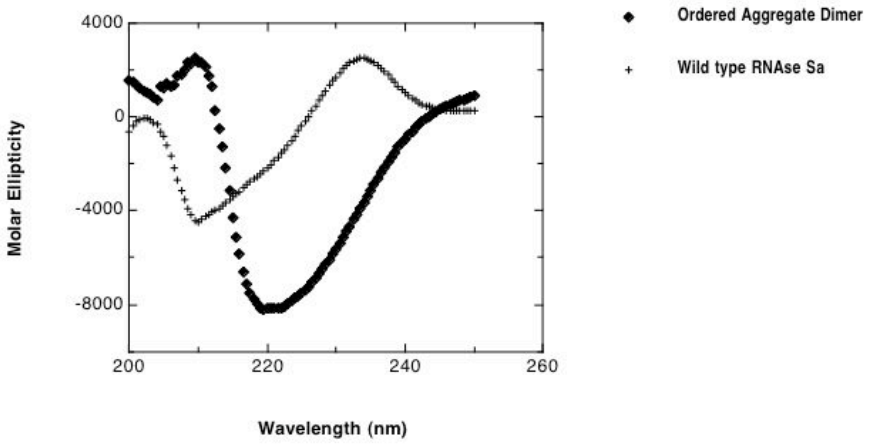

Figure 53. Far-UV circular dichroism spectra of the monomeric and dimeric wt Sa in 10% Glycerol Tris/Glycine pH 8.0. Spectra were blank subtracted and all the instrument corrections were applied.

As a final test of amyloid formation, a seeding experiment was performed.[109] An aliquot of Sa dimer was incubated with wt Sa at pH 3.5 for eighteen hours. As a control, the wt Sa monomer was also incubated with wt Sa at pH 3.5. Also, the isolated dimer was incubated without wt Sa at pH 3.5. The seeding sample showed an increased CR and ThT binding compared to both negative controls, showing that the dimer was able to nucleate the formation of more amyloid (Figure 54).

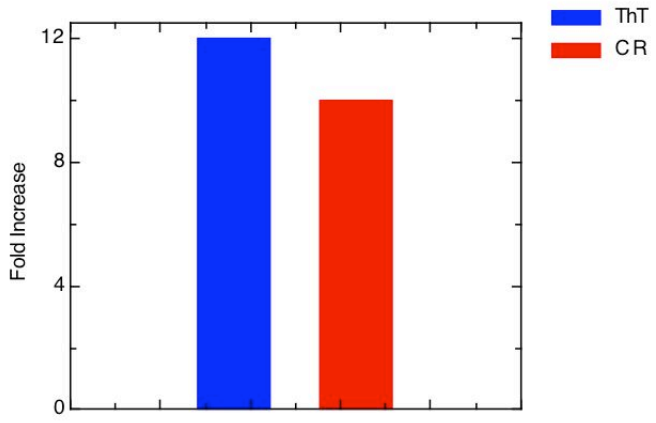

Figure 54. Seeding of amyloid formation by a RNase Sa dimer. These values represent the end point of incubation of RNase Sa dimer with wt Sa at pH 3.5. The fold increase was calculated from the control of RNase Sa monomer incubated with RNase Sa at pH 3.5.

DISCUSSION

Purification and Characterisation

To my knowledge this is the first on-pathway intermediate of amyloid ever isolated. The results from the purifications of the intermediates are shown in Table 7. We identified three characteristics of amyloid for the intermediates: 1) Congo red binding and Thioflavine T binding were determined to increase for the intermediates 2)

The dimer was shown to seed the formation of amyloid and 3) the intermediates show CD spectra consistent with a β sheet conformation.

The intermediates are capable of binding the amyloid specific dyes Congo red and Thioflavine T. Previously assays that utilized these dyes could not reveal the nature of the dye-binding conformation of the protein.[115] With these experiments, we have been able to narrow the possibilities to two: the dimeric intermediate, and the amyloid competent monomer. We speculate that the β-sheet enriched monomer is not capable of binding congo red and thioflavine T. This is due to the strong correlation between the kinetics of conformational change (data not shown) and amyloid formation followed by the amyloid specific dyes (Figure 43). We monitored the CD spectrum of the monomeric protein as a function of time and observed a conformational change that occurred after 3 hours (data not shown). This could be due to the formation of a β rich dimer or the formation of a β rich monomer. Recovery results suggest that the dimeric intermediate is only weakly populated, suggesting that the β rich monomer is the nucleation compenent spicies. This hypothesis suggests that in order to study the production of the amyloid competent intermediate, we will need more specific probe for the various conformations.

We estimate the conversion of RNase Sa to amyloid to be ~ 20% (Table 7). The formation of the on-pathway intermediate requires at least two elements 1) a β rich structure and 2) a charge neutral protein. We estimate by CD that the population of the β rich structure to be > 80%. Even when the pH and TFE are maximized for amyloid formation, the kinetics of amyloid formation are still slow, taking greater than 6 hours to

complete. In contrast, non productive aggregation is very fast, often occurring within minutes of addition of protein.

The isolated intermediates were able to seed the formation of amyloid. The ability of isolated intermediates to seed the formation of amyloid helps to identify the nucleus of fiber formation. Several studies have suggested that the lag phase is attributed to the rate limiting formation of the dimer, while others suggest that the formation of an amyloid competent monomer is the rate limiting step .[143] Our results suggest that the dimer is capable of inducing fiber formation, suggesting that at least a dimer is required to seed fiber formation. Further experiments will be necessary to address this hypothesis.

Stability & Solubility of the Intermediates

Amyloid fibers are known to be insoluble and extremely stable structures,[144; 145; 146] however, to date the stability and solubility of the intermediate forms or proto-fibers are unknown. Figure 48 shows an SDS gel of several well-studied proteins which were exposed to 30% TFE at pH = pI. Two major points can be made from this gel: 1) the full length aggregates are resistant to denaturation by SDS and boiling and 2) some of the intermediates are resistant to denaturation by SDS and boiling. This result suggests that the intermediates have an increased conformational stability. Another possibility is that SDS stabilizes the intermediates. For this to be a possibility, the intermediate would most likely have an increased exposure of hydrophobic residues.

Amyloid fibers are known to be insoluble aggregates. The intermediates are exceptionally prone to aggregation most likely due to an increased hydrophobicity. This statement is supported by 3 observations: 1) a ~ 5 fold increase in ANS binding of the isolated intermediate and wt Sa (data not shown). 2) Intermediates can be stained by a methanol-based stain, but soon disappear from the gel and improved staining occurred when stained by a water-based stain. This suggests that these intermediates have an increased solubility in the methanol-based stain. 3) Since we can induce fiber formation by exposing RNase Sa to 30% TFE, a hydrophobic solvent, it seems likely that the intermediates would have an increased hydrophobicity relative to the native protein. These results suggest that the intermediates may be stable, yet insoluble structures. However, further study is needed to confirm these observations.

CHAPTER VII

SUMMARY

This dissertation consists of two very different projects. The first was to perform a comparison of the equilibrium and kinetic folding for the HPr proteins from *Bacillus subtilis* and *E. coli*. These proteins exhibit super protection. Super protection is the lack of agreement between conformational stability determined by hydrogen exchange ΔG_{HD} and traditional methods ΔG_{UDC}. It has been shown that with the proper corrections, agreement between ΔG_{HD} and ΔG_{UDC} can be achieved for a variety of different proteins. [8] We show that the HPr proteins do not fold in a simple two-state manner. We show that removing a single charged side chain (Asp 69) is enough to convert the folding transition for HPr to a simple two-state mechanism. While some of the data presented here have been previously analyzed using a two-state analysis (urea and thermal denaturations followed by CD),[14] recent data suggests that a three-state analysis better describes the unfolding of wt *bs*HPr and wt *ec*HPr. We have measured and calculated the equilibrium conformational stability ΔG_U by equilibrium and kinetic methods. The data that most supports a three-state unfolding mechanism are the lack of agreement between traditional equilibrium methods and hydrogen exchange. In both cases, a kinetic intermediate was detected which could explain the difference between the traditional methods and hydrogen exchange. It would appear that hydrogen exchange is capable of detecting the true thermodynamic conformational stability of a protein. These studies

illustrate that a two-state model should not be solely determined based on a conformational probe, but rather a combination of techniques and probes should be utilized.

The second project investigates the relationship between protein solubility, stability, and amyloid fiber formation using ribonuclease Sa (RNase Sa) as a model system. RNase Sa is a small acidic protein (96 residues, pI=3.5) with a mixed $\alpha + \beta$ structure. We have recently made variants that replaced solvent-exposed acidic residues with lysine residues. The mutant with three such charge reversals is denoted 3K, and that with fives is 5K. By reversing the charge at these sites, we are able to study the relationship between stability, solubility and fibril formation in a single protein over a wide pH range. Our results suggest that the solubility of a polypeptide chain is a major factor that determines the *in vitro* conversion of globular proteins into amyloid fibrils under non-pathological conditions and may contribute to fibril formation *in vivo*. The results also suggest that increasing the solubility of a protein should prevent the formation of fibrils. We are in the process of testing this hypothesis with RNase Sa and other proteins.

REFERENCES

1. Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10-9.
2. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44-45.
3. Baldwin, R. L. & Rose, G. D. (1999). Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci* **24**, 26-33.
4. Baldwin, R. L. & Rose, G. D. (1999). Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci.* **24**, 77-83.
5. Dill, K. A., Bromberg, S., Yue, K., Fiebig, K. M., Yee, D. P., Thomas, P. D. & Chan, H. S. (1995). Principles of protein folding--a perspective from simple exact models. *Protein Science* **4**, 561-602.
6. Chan, H. S., Bromberg, S. & Dill, K. A. (1995). Models of cooperativity in protein folding. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **348**, 61-70.
7. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *FASEB J.* **10**, 75-83.
8. Huyghues-Despointes, B. M., Scholtz, J. M. & Pace, C. N. (1999). Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat Struct Biol* **6**, 910-2.

9. Eftink, M. R. & Shastry, M. C. (1997). Fluorescence methods for studying kinetics of protein-folding reactions. *Methods Enzymol* **278**, 258-86.

10. Vivian, J. T. & Callis, P. R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys J* **80**, 2093-109.

11. Woody, R. W. (1995). Circular dichroism. *Methods Enzymol* 246, 34-71.

12. Aronsson, G., Brorsson, A. C., Sahlman, L. & Jonsson, B. H. (1997). Remarkably slow folding of a small protein. *FEBS Lett* **411**, 359-64.

13. Nicholson, E. M. & Scholtz, J. M. (1996). Conformational stability of the *Escherichia coli* HPr protein: Test of the linear extrapolation method and a thermodynamic characterization of cold denaturation. *Biochemistry* **35**, 11369-11378.

14. Scholtz, J. M. (1995). Conformational stability of HPr: the histidine-containing phosphocarrier protein from *Bacillus subtilis*. *Protein Science* **4**, 35-43.

15. Taddei, N., Chiti, F., Paoli, P., Fiaschi, T., Bucciantini, M., Stefani, M., Dobson, C. M. & Ramponi, G. (1999). Thermodynamics and kinetics of folding of common-type acylphosphatase: comparison to the highly homologous muscle isoenzyme. *Biochemistry* **38**, 2135-42.

16. van Nuland, N. A., Meijberg, W., Warner, J., Forge, V., Scheek, R. M., Robillard, G. T. & Dobson, C. M. (1998). Slow cooperative folding of a small globular protein HPr. *Biochemistry* **37**, 622-37.

17. Greene, R. F. & Pace, C. N. (1974). Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, α-chymotrypsin, and β-lactoglobulin. *J. Biol. Chem.* **249**, 5388-5393.

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

19. Pace, C. N. & Laurents, D. V. (1989). A new method for determining the heat capacity change for protein folding. *Biochemistry* **28**, 2520-2525.

20. Freire, E. (1995). Thermal denaturation methods in the study of protein folding. *Methods Enzymol.* **259**, 144-168.

21. Freire, E. (1995). Differential scanning calorimetry. In *Protein Folding and Stability: Theory and Practice* (Shirley, B. A., ed.), Vol. **40**, pp. 191-218. Humana Press, Totowa, NJ.

22. Eftink, M. E. & Ionescu, R. (1997). Thermodynamics of protein unfolding: questions pertinent to testing the validity of the two-state model. *Biophys. Chem.* **64**, 175-197.

23. Grimsley, J. K., Scholtz, J. M., Pace, C. N. & Wild, J. R. (1997). Organophosphorus hydrolase is a remarkably stable enzyme that unfolds through a homodimeric intermediate. *Biochemistry* **36**, 14366-14374.

24. Botelho, M. G., Gralle, M., Oliveira, C. L., Torriani, I. & Ferreira, S. T. (2003). Folding and stability of the extracellular domain of the human amyloid precursor protein. *J. Biol. Chem.* **250**, 5389-5390.

25. Freire, E. (1995). Thermodynamics of partly folded intermediates in proteins. *Annual Review of Biophysics & Bio molecular Structure* **24**, 141-65.

26. Scholtz, J. M. & Robertson, A. D. (1995). Hydrogen exchange techniques. In *Protein Stability and Folding. Theory and Practice* (Shirley, B. A., ed.), Vol. **40**, pp. 291-311. Humana Press, Totowa. NJ.

27. Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. (1996). Mechanisms and uses of hydrogen exchange. *Curr Opin Struct Biol* 6, 18-23.

28. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins* **20**, 4-14.

29. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange I. *Proteins: Structure, Function, and Genetics* **17**, 75-86.

30. Englander, S. W., Bai, Y., Milne, J. S. & Mayne, L. (1993). Primary structure effects on peptide group hydrogen exchange II. *Proteins: Structure, Function, and Genetics* **17**, 86-98.

31. Makhatadze, G. I., Clore, G. M. & Gronenborn, A. M. (1995). Solvent isotope effect and protein stability. *Nature Struct. Biol.* **2**, 852-5.

32. Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M. & Fischer, G. (1998). Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J Mol Biol* **279**, 449-60.

33. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-230.

34. Kuwajima, K., Yamaya, H. & Sugai, S. (1996). The burst-phase intermediate in the refolding of β-lactoglobulin studied by stopped-flow circular dichroism and absorption spectroscopy. *J Mol Biol* **264**, 806-822.

35. Sosnick, T. R., Shtilerman, M. D., Mayne, L. & Englander, S. W. (1997). Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl. Acad. Sci.* **94**, 8545-50.

36. Tezcan, F. A., Findley, W. M., Crane, B. R., Ross, S. A., Lyubovitsky, J. G., Gray, H. B. & Winkler, J. R. (2002). Using deeply trapped intermediates to map the cytochrome c folding landscape. *Proc. Natl. Acad. Sci.* **99**, 8626-30.

37. Baldwin, R. L. (1995). The nature of protein folding pathways: the classical versus the new view. *J. Biomol. NMR* **5**, 103-9.

38. Kiefhaber, T. (1995). Protein folding kinetics. *Methods in Molecular Biology* **40**, 313-41.

39. Yan, Y. B., Jiang, B., Zhang, R. Q. & Zhou, H. M. (2001). Two-phase unfolding pathway of ribonuclease A during denaturation induced by dithiothreitol. *Protein Science* **10**, 321-8.

40. Wildegger, G. & Kiefhaber, T. (1997). Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J Mol Biol* **270**, 294-304.

41. Roux, P., Delepierre, M., Goldberg, M. E. & Chaffotte, A. F. (1997). Kinetics of secondary structure recovery during the refolding of reduced hen egg white lysozyme. *J. Biol. Chem* **272**, 24843-9.

42. Parker, M. J. & Marqusee, S. (1999). The cooperativity of burst phase reactions explored. *J Mol Biol* **288**, 1597-608

43. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C. & Radford, S. E. (1999). Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J Mol Biol* **286**, 1597-608.

44. Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E. & Fersht, A. R. (1995). Movement of the position of the transition state in protein folding. *Biochemistry* **34**, 13656-62.

45. Fersht, A. R., Itzhaki, L. S., Elmasry, N. F. & Matthews, J. M. (1994). Single versus paralell pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl. Acad. Sci. USA* **91**, 10426-10429.

46. Matthews, C. R. (1993). Pathways of protein folding. *Annual Review of Biochemistry* **62**, 653-683.

47. Utiyama, H. & Baldwin, R. L. (1986). Kinetic mechanisms of protein folding. *Methods Enzymol* **131**, 51-70.

48. Baldwin, R. L. (1996). On-pathway versus off-pathway folding intermediates. *Folding & Design* **1**, R1-R8.

49. Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr Opin Struct Biol* **7**, 15-28.

50. Roder, H. & Shastry, M. R. (1999). Methods for exploring early events in protein folding. *Curr Opin Struct Biol* **9**, 620-6.

51. Oliveberg, M. & Fersht, A. R. (1996). Thermodynamics of transient conformations in the folding pathway of barnase: reorganization of the folding intermediate at low pH. *Biochemistry* **35**, 2738-49.

52. Matouschek, A., Matthews, J. M., Johnson, C. M. & Fersht, A. R. (1994). Extrapolation to water of kinetic and equilibrium data for the unfolding of barnase in urea solutions. *Protein Engineering* **7**, 1089-95.

53. Otzen, D. E., Kristensen, O., Proctor, M. & Oliveberg, M. (1999). Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. *Biochemistry* **38**, 6499-511.

54. Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14**, 4953-63.

55. Kiefhaber, T. & Schmid, F. X. (1992). Kinetic coupling between protein folding and prolyl isomerization. *J Mol Biol* **224**, 231-240.

56. Mayr, L. M., Odefey, C., Schutkowski, M. & Schmid, F. X. (1996). Kinetic analysis of the unfolding and refolding of ribonuclease T1 by stopped-flow double-mixing technique. *Biochemistry* **35**, 5550-5561.

57. Schindler, T. & Schmid, F. X. (1996). Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry* **35**, 16833-42.

58. Capaldi, A. P., Shastry, M. C., Kleanthous, C., Roder, H. & Radford, S. E. (2001). Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat Struct Biol* **8**, 68-72.

59. Kiefhaber, T., Quaas, R., Haan, U. & Schmid, F. X. (1990). Folding of ribonuclease T1. 2. Kinetic models for the folding and unfolding reaction. *Biochemistry* **29**, 3061-3070.

60. Kiefhaber, T., Quaas, R., Hahn, U. & Schmid, F. X. (1990). Folding of ribonuclease T1. 1. existence of multiple unfolded states created by proline isomerization. *Biochemistry* **29**, 3053-3061.

61. Kiefhaber, T., Schmid, F. X., Willart, K., Engelborghs, Y. & Chaffotte, A. (1992). Structure of a rapidly formed intermediate in ribonuclease T1 folding. *Protein Science* **1**, 1162-1172.

62. Waygood, E. B. (1998). The structure and function of HPr. *Biochemistry & Cell Biology* **76**, 359-367.

63. Jia, Z., Quail, J. W., Waygood, E. B. & Delbaere, L. T. (1993). The 2.0-Å resolution structure of Escherichia coli histidine-containing phosphocarrier protein HPr. A redetermination. *J. Biol. Chem* **268**, 22490-22501.

64. Pelton, J. G., Torchia, D. A., Meadow, N. D. & Roseman, S. (1992). Structural comparison of phosphorylated and unphosphorylated forms of IIIGlc, a signal-transducing protein from *Escherichia coli*, using three-dimensional NMR techniques. *Biochemistry* **31**, 5215-24.

65. Herzberg, O., Reddy, P., Sutrina, S., Saier, M., Reizer, J. & Kapadia, G. (1992). Structure of the histidine-containing phosphocarrier protein HPr from *Bacillus subtilis* at 2.0 Å resolution. *Proc. Natl. Acad. Sci., USA* **89**, 2499-2503.

66. Rajagopal, P., Waygood, E. B. & Klevit, R. E. (1994). Structural consequences of histidine phosphorylation: NMR characterization of the phosphohistidine form of histidine-containing protein from *Bacillus subtilis* and *Escherichia coli. Biochemistry* **33**, 15271-82.

67. van Nuland, N. A., Boelens, R., Scheek, R. M. & Robillard, G. T. (1995). High-resolution structure of the phosphorylated form of the histidine-containing phosphocarrier protein HPr from *Escherichia coli* determined by restrained molecular dynamics from NMR-NOE data. *J Mol Biol* **246**, 180-93.

68. van Nuland, N. A., Grotzinger, J., Dijkstra, K., Scheek, R. M. & Robillard, G. T. (1992). Determination of the three-dimensional solution structure of the histidine-containing phosphocarrier protein HPr from *Escherichia coli* using multidimensional NMR spectroscopy. *European Journal of Biochemistry* **210**, 881-91.

69. van Nuland, N. A., Hangyi, I. W., van Schaik, R. C., Berendsen, H. J., F., v. G. W., Scheek, R. M. & Robillard, G. T. (1994). The high-resolution structure of the

histidine-containing phosphocarrier protein HPr from *Escherichia coli* determined by restrained molecular dynamics from nuclear magnetic resonance nuclear Overhauser effect data. *J Mol Biol* **237**, 544-59.

70. Nicholson, E. M., Peterson, R. W. & Scholtz, J. M. (2002). A partially buried site in homologous HPr proteins is not optimized for stability. *J Mol Biol* **321**, 355-362.

71. Peterson, R. W., Nicholson, E. M., Thapar, R., Klevit, R. E. & Scholtz, J. M. (1999). Increased helix and protein stability through the introduction of a new tertiary hydrogen bond. *J Mol Biol* **286**, 1609-1619.

72. Thapar, R., Nicholson, E. M., Waygood, E. B., Scholtz, J. M. & Klevit, R. E. (1996). Influence of N-cap mutations on the structure and stability of *E. coli* HPr. *Biochemistry* **35**, 11268-11277.

73. Waygood, E. B., Erickson, E., El-Kabbani, O. A. L. & Delbaere, L. T. J. (1985). Characterization of phosphorylated histidine-containing protein (HPr) of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *Biochemistry* **24,** 6938-6945.

74. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59.

75. Sambrook, J., Fritsch, E. F. & Manniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. 2nd edit (Nolan, C., Ed.), Cold Springs Harbor Press, Cold Springs Harbor, NY.

76. Anderson, J. W., Bhanot, P., Georges, F., Klevit, R. E. & Waygood, E. B. (1991). Involvement of the carboxy-terminal residue in the active site of the histidine-containing protein, HPr, of the phosphoenolpyruvate:sugar transport system of *Escherichia coli*. *Biochemistry* **30**, 9601-9607.

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

78. Synergy Software, P. I. (1996). KaleidaGraph Fourth Edition 4.0 edit. Synergy Software, PA.

79. Pace, C. N. & Scholtz, J. M. (1997). Measuring the conformational stability of a protein. In *Protein Structure: A Practical Approach* 2nd edition. (Creighton, T. E., ed.), pp. 299-321. IRL Press, Oxford.

80. Oas, T. G. & Toone, E. J. (1997). Thermodynamic solvent isotope effects and molecular hydrophobicity. In *Adv. Biophys. Chem.* (Bush, C., ed.), Vol. **6**, pp. 1-52. JAI Press, INC, Greenwich, CT.

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

82. Peterson, R. W. (2001). Examination of the propagation of stabilizing interactions in HPr by multiple techniques, and incorporation of designe rules into a computer prediction algorithm. Doctor of Philosophy, Texas A&M. College Station,TX.

83. Matthews, C. R. (1987). Effect of point mutations on the folding of globular proteins. *Methods Enzymol* **154**, 498-511.

84. Software. (1998). DSC Data analysis software, MicroCal Software, New York.

85. Klunk, W. E., Pettegrew, J. W. & Abraham, D. J. (1989). Two simple methods for quantifying low-affinity dye-substrate binding. *J Histochem Cytochem* **37**, 1293-7.

86. Klunk, W. E., Pettegrew, J. W. & Abraham, D. J. (1989). Quantitative evaluation of congo red binding to amyloid-like proteins with a beta-pleated sheet conformation. *J Histochem Cytochem* **37**, 1273-81.

87. Bowers, P. M. & Klevit, R. E. (1996). Hydrogen bonding and equilibrium isotope enrichment in histidine-containing proteins. *Nat Struct Biol* **3**, 522-31.

88. Santoro, M. M. & Bolen, D. W. (1992). A test of the linear extrapolation of unfolding free energy changes over an extended denaturant concentration range. *Biochemistry* **31**, 4901-4907.

89. Schellman, J. A. & Gassner, N. C. (1996). The enthalpy of transfer of unfolded proteins into solutions of urea and guanidinium chloride. *Biophysical Chemistry* **59**, 259-275.

90. Smith, J. S. & Scholtz, J. M. (1996). Guanidine hydrochloride unfolding of peptide helices: separation of denaturant and salt effects. *Biochemistry* **35**, 7292-7297.

91. Huyghues-Despointes, B. M. P., Scholtz, J. M. & Pace, C. N. (1999). Protein conformational stabilities can be determined from hydrogen-exchange rates. *Nat. Struct. Biol.* **6**, 910-912.

92. Freire, E. (1995). Differential scanning calorimetry. *Methods Mol Biol* **40**, 191-218.

93. Pace, C. N. (1975). The stability of globular proteins. *CRC Crit Rev Biochem* **3**, 1-43.

94. Agashe, V. R. & Udgaonkar, J. B. (1995). Thermodynamics of denaturation of barstar: Evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry* **34**, 3286-3299.

95. Hammen, P. K., Scholtz, J. M., Anderson, J. W., Waygood, E. B. & Klevit, R. E. (1995). Investigation of a side-chain-side-chain hydrogen bond by mutagenesis, thermodynamics, and NMR spectroscopy. *Protein Science* **4**, 936-44.

96. Linderstrøm-Lang, K. (1958). On the ionisation of proteins. In *Symposium on Protein Structure* (Neuberger, A., ed.). Methuen, London.

97. Englander, S. W. & Kallenbach, N. R. (1984). Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Quarterly Rev. Biophys.* **16**, 521-655.

98. Huyghues-Despointes, B. M., Pace, C. N., Englander, S. W. & Scholtz, J. M. (2001). Measuring the conformational stability of a protein by hydrogen exchange. In *Methods in Molecular Biology* (Murphy, K. P., ed.), Vol. **168**, pp. 69-92. Humana Press, Totowa. NJ.

99. Grantcharova, V. P. & Baker, D. (1997). Folding dynamics of the src SH3 domain. *Biochemistry* **36**, 15685-92.

100. Khorasanizadeh, S., Peters, I. D. & Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nature Struct. Biol.* **3**, 193-205.

101. Khare, D., Alexander, P. & Orban, J. (1999). Hydrogen bonding and equilibrium protium-deuterium fractionation factors in the immunoglobulin G binding domain of protein G. *Biochemistry* **38**, 3918-3925.

102. Chiti, F., Taddei, N., Stefani, M., Dobson, C. M. & Ramponi, G. (2001). Reduction of the amyloidogenicity of a protein by specific binding of ligands to the native conformation. *Protein Science* **10**, 879-86.

103. Cohen, A. S. (1967). Amyloidosis. *New England Journal of Medicine* **277**, 522-30 contd.

104. Cohen, A. S. (1967). Amyloidosis (concluded). *New England Journal of Medicine* **277**, 628-38.

105. Glenner, G. G. (1980). Amyloid deposits and amyloidosis. The beta-fibrilloses (first of two parts). *New England Journal of Medicine* **302**, 1283-92.

106. Rochet, J. C. & Lansbury, P. T., Jr. (2000). Amyloid fibrillogenesis: themes and variations. *Curr Opin Struct Biol* **10**, 60-8.

107. Lai, Z., Colon, W. & Kelly, J. W. (1996). The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. *Biochemistry* **35**, 6470-82.

108. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat Struct Biol* **6**, 1005-9.

109. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. & Dobson, C. M. (1999). Designing conditions for *in vitro* formation of amyloid protofilaments and fibrils. *Proc Natl Acad Sci U S A* **96**, 3590-4.

110. Krebs, M. R., Wilkins, D. K., Chung, E. W., Pitkeathly, M. C., Chamberlain, A. K., Zurdo, J., Robinson, C. V. & Dobson, C. M. (2000). Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the beta-domain. *J Mol Biol* **300**, 541-9.

111. Ramirez-Alvarado, M., Merkel, J. S. & Regan, L. (2000). A systematic exploration of the influence of the protein stability on amyloid fibril formation *in vitro*. *Proc Natl Acad Sci U S A* **97**, 8979-84.

112. Fändrich, M., Fletcher, M. A. & Dobson, C. M. (2001). Amyloid fibrils from muscle myoglobin. *Nature* **410**, 165-6.

113. O'Nuallain, B. & Wetzel, R. (2002). Conformational Abs recognizing a generic amyloid fibril epitope. *Proc Natl Acad Sci U S A* **99**, 1485-90.

114. Kelly, J. W. (1996). Alternative conformations of amyloidogenic proteins govern their behavior. *Curr Opin Struct Biol* **6**, 11-7.

115. Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G. & Dobson, C. M. (2000). Mutational analysis of the propensity for amyloid formation by a globular protein. *Embo J* **19**, 1441-9.

116. Khurana, R., Uversky, V. N., Nielsen, L. & Fink, A. L. (2001). Is Congo red an amyloid-specific dye? *J Biol Chem* **276**, 22715-21.

117. Luo, P. & Baldwin, R. L. (1997). Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix–forming properties of peptides from trifluoroethanol:water mixtures back to water. *Biochemistry* **36**, 8413-8421.

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

119. Jiang, X., Smith, C. S., Petrassi, H. M., Hammarstrom, P., White, J. T., Sacchettini, J. C. & Kelly, J. W. (2001). An engineered transthyretin monomer that is nonamyloidogenic, unless it is partially denatured. *Biochemistry* **40**, 11442-52.

120. Pepys, M. B., Hawkins, P. N., Booth, D. R., Vigushin, D. M., Tennent, G. A., Soutar, A. K., Totty, N., Nguyen, O., Blake, C. C., Terry, C. J. & et al. (1993). Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature* **362**, 553-7.

121. Souillac, P. O., Uversky, V. N., Millett, I. S., Khurana, R., Doniach, S. & Fink, A. L. (2002). Effect of association state and conformational stability on the

kinetics of immunoglobulin light chain amyloid fibril formation at physiological pH. *J. Biol. Chem* **277**, 12657-65.

122. Nielsen, L., Khurana, R., Coats, A., Frokjaer, S., Brange, J., Vyas, S., Uversky, V. N. & Fink, A. L. (2001). Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. *Biochemistry* **40**, 6036-46.

123. Litvinovich, S. V., Brew, S. A., Aota, S., Akiyama, S. K., Haudenschild, C. & Ingham, K. C. (1998). Formation of amyloid-like fibrils by self-association of a partially unfolded fibronectin type III module. *J Mol Biol* **280**, 245-58.

124. Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem Sci* **24**, 329-32.

125. Schein, C. H. (1990). Solubility as a function of protein structure and solvent components. *Biotechnology (N Y)* **8**, 308-17.

126. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. (2001). The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Science* **10**, 1206-15.

127. Wood, S. J., Maleeff, B., Hart, T. & Wetzel, R. (1996). Physical, morphological and functional differences between ph 5.8 and 7.4 aggregates of the Alzheimer's amyloid peptide Abeta. *J Mol Biol* **256**, 870-7.

128. Hoyer, W., Antony, T., Cherny, D., Heim, G., Jovin, T. M. & Subramaniam, V. (2002). Dependence of alpha-synuclein aggregate morphology on solution conditions. *J Mol Biol* **322**, 383-93.

129. Hebert, E. J., Grimsley, G. R., Hartley, R. W., Horn, G., Schell, D., Garcia, S., Both, V., Sevcik, J. & Pace, C. N. (1997). Purification of ribonucleases Sa, Sa2, and Sa3 after expression in *Escherichia coli*. *Prot. Expr. Purif.* **11**, 162-8.

130. Pace, C. N., Hebert, E. J., Shaw, K. L., Schell, D., Both, V., Krajcikova, D., Sevcik, J., Wilson, K. S., Dauter, Z., Hartley, R. W. & Grimsley, G. R. (1998). Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3. *J Mol Biol* **279**, 271-286.

131. Pace, C. N., Laurents, D. V. & Thomson, J. A. (1990). pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. *Biochemistry* **29**, 2564-2572.

132. Pfeil, W. & Privalov, P. L. (1976). Thermodynamic investigations of proteins. I. Standard functions for proteins with lysozyme as an example. *Biophys Chem* **4**, 23-32.

133. Harrison, P. M., Bamborough, P., Daggett, V., Prusiner, S. B. & Cohen, F. E. (1997). The prion folding problem. *Curr Opin Struct Biol* **7**, 53-9.

134. Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C. M. & Stefani, M. (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507-11.

135. Hammarstrom, P., Jiang, X., Deechongkit, S. & Kelly, J. W. (2001). Anion shielding of electrostatic repulsions in transthyretin modulates stability and

amyloidosis: insight into the chaotrope unfolding dichotomy. *Biochemistry* **40**, 11453-9.

136. Hurle, M. R., Helms, L. R., Li, L., Chan, W. & Wetzel, R. (1994). A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proc Natl Acad Sci U S A* **91**, 5446-50.

137. Niraula, T. N., Haraoka, K., Ando, Y., Li, H., Yamada, H. & Akasaka, K. (2002). Decreased thermodynamic stability as a crucial factor for familial amyloidotic polyneuropathy. *J Mol Biol* **320**, 333-42.

138. Scrocchi, L. A., Chen, Y., Waschuk, S., Wang, F., Cheung, S., Darabie, A. A., McLaurin, J. & Fraser, P. E. (2002). Design of peptide-based inhibitors of human islet amyloid polypeptide fibrillogenesis. *J Mol Biol* **318**, 697-706.

139. Cairo, C. W., Strzelec, A., Murphy, R. M. & Kiessling, L. L. (2002). Affinity-based inhibition of beta-amyloid toxicity. *Biochemistry* **41**, 8620-9.

140. Mirzabekov, T. A., Lin, M. C. & Kagan, B. L. (1996). Pore formation by the cytotoxic islet amyloid peptide amylin. *J. Biol. Chem* **271**, 1988-92.

141. Olofsson, A., Ippel, H. J., Baranov, V., Horstedt, P., Wijmenga, S. & Lundgren, E. (2001). Capture of a dimeric intermediate during transthyretin amyloid formation. *J. Biol. Chem* **276**, 39592-9.

142. Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. (2001). A domain-swapped RNase A dimer with implications for amyloid formation. *Nat Struct Biol* **8**, 211-4.

143. Liu, Y., Hart, P. J., Schlunegger, M. P. & Eisenberg, D. (1998). The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1- A resolution. *Proc Natl Acad Sci U S A* **95**, 3437-42.

144. Knaus, K. J., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W. K. & Yee, V. C. (2001). Crystal structure of the human prion protein reveals a mechanism for oligomerization. *Nat Struct Biol* **8**, 770-4.

145. Bellotti, V., Stoppini, M., Mangione, P., Sunde, M., Robinson, C., Asti, L., Brancaccio, D. & Ferri, G. (1998). Beta2-microglobulin can be refolded into a native state from ex vivo amyloid fibrils. *European Journal of Biochemistry* **258**, 61-7.

146. Tito, P., Nettleton, E. J. & Robinson, C. V. (2000). Dissecting the hydrogen exchange properties of insulin under amyloid fibril forming conditions: a site-specific investigation by mass spectrometry. *J Mol Biol* **303**, 267-78.

VITA

Jason Schmittschmitt
16100 West Bay Drive, Unit 247
Jupiter, Fl. 33477

Education

University of Kentucky, Thomas Hunt Morgan School of Biological Sciences
Lexington , Kentucky 1992-1996
Bachelor of Science in Biology

Texas A&M Department of Biochemistry & Biophysics
College Station Texas 1998-2003
Doctor of Philosophy