SAPIENZA
UNIVERSITÀ DI ROMA

DOTTORATO DI RICERCA IN BIOCHIMICA
CICLO XXIV (A.A. 2008-2011)

THE FOLDING PROBLEM SIMPLIFIED: PROTEINS WITH NEARLY IDENTICAL SEQUENCE BUT DIFFERENT STRUCTURE AND FUNCTION

Dottoranda
Angela Morrone

Docente guida
Prof. Maurizio Brunori

Coordinatore
Prof. Paolo Sarti

Dicembre 2011

Acknowledgements

This thesis owes its existence to the help, support, and inspiration of many people. In the first place, I would like to express my deep and sincere gratitude to my supervisor Prof. Maurizio Brunori for his excellent guidance, constant encouragement, patience and care during the entire course of my Ph.D.

I would like to thank also the coordinator of the Ph.D. Program in Biochemistry, Prof. Paolo Sarti for his kind and expert guidance during these three years.

I am very grateful to Prof. Carlo Travaglini-Allocatelli for his important support. His help and constructive advices were extremely precious to me.

I am deeply indebted to Prof. Stefano Gianni. I greatly benefited from his punctual insights and scientific experience and I really appreciated his continuous help during all the time of the research and writing of this thesis. Thank you Stefano-sir!

I wish to thank also Prof. Philip N. Bryan and Prof. Valerie Daggett for their precious contribution to our project.

Special thanks are addressed to Prof. Per Jemth for kindly hosting me in his laboratory in Uppsala (Sweden). These thanks are also for the people belonging to his group; even though I spent only a month with them, they made me feel welcome and part of their group. In particular I wish to thank Dr. Celestine Chi, for his precious help during my stay there.

I wish to thank present and past members of the lab, in particular Maria Antonietta Carillo, her friendship has been a special gift from this Ph.D., Eva Di Silvio for her company and for being there for me, especially during the tough times, and Rajanish Giri, my "G_A and G_B partner", who shared with me a big part of the work presented in this thesis with joy and cheerfulness.

I would also like to thank all the people with whom I shared my days on the "Soppalco": Giorgio Giardina, Valentina Stelitano, Nicoletta Castiglione, Serena Rinaldo and Manuela Caruso. Thank you all for supporting me,

Acknowledgements

especially in the last months. Extra special thanks go to Giorgio: his wonderful cover figure has given a surplus value to my thesis.

Many thanks are also addressed to all the professors, researchers and students of the "piano terra", among them I am particularly grateful to Linda Celeste Montemiglio for her helpful encouragement and hugs.

I owe an acknowledgement also to Prof. Fabio Altieri and Dr. Caterina Grillo, I would have never started the Ph.D. without them.

Much has happened and changed in the time I have been involved with this Ph.D. project. I devote my deepest gratitude to my family, for their never-ending support, love and care. Finally, this thesis is dedicated to my husband for his complete and unconditional love. I could not have completed this journey without him by my side.

TABLE OF CONTENTS

INTRODUCTION

1.1 The Folding Problem

The ability of a protein to fold rapidly and efficiently into its intricate and highly specific structure is an essential part of the conversion of genetic information into cellular activity.
Despite over five decades of work, understanding protein folding still remains one of the major challenges in modern biochemistry and biophysics.
Pioneering work in the 1950s by Christian Anfinsen, on the folding of ribonuclease, demonstrated that the primary structure of a protein "encodes" the information necessary for a nascent polypeptide to fold into its native, physiologically active, three-dimensional conformation (Anfinsen, et al. 1961). After denaturation, indeed, small globular proteins can fold back, into proper shape, simply removing the denaturant agent. The spontaneity of the folding reaction led to the concept that the native conformation of a protein is the most stable accessible state and corresponds to the lowest free energy (Pace 1990).
Afterwards, Cyrus Levinthal pointed out that folding can not occur by a stochastic search among all possible conformations; because otherwise even a small protein would need a unrealistically long time to fold (Levinthal 1968). The attractive and repulsive forces between neighbouring amino acid residues, favouring certain conformations of individual amino acids in the polypeptide chain, dramatically reduce the conformational space and the number of possible folding pathways available.
Therefore, to understand how proteins fold, it is necessary to understand the thermodynamics of the forces stabilizing the native state and also the dynamic mechanism whereby the conformational search to the native state is achieved.

1.2 In vitro Folding studies

In order to draw general rules about this complex and fascinating process, most folding studies are carried out by characterizing *in vitro* the reversible folding mechanism of simple protein systems. Small globular proteins, involving a limited experimentally accessible number of states, are generally used.
A fundamental experimental approach is represented by the so-called perturbation method: altering some environmental conditions, it is possible to alter the equilibrium of the system and the relative populations of native and denatured species. A small perturbation, obtained by chemical or physical agents such as urea, guanidine hydrochloride or heat, is associated to a relaxation process to a new equilibrium. It is possible to monitor this process by a variety of spectroscopic techniques, employing as optical probes absorbance, fluorescence, circular dichroism, small angle X-ray scattering and nuclear magnetic resonance.
As described below, the current approach to study experimentally protein folding consists of characterizing folding reactions under equilibrium and transient conditions.

1.2.1 Equilibrium studies

Equilibrium studies provide information about the folding process in terms of stability and co-operativity. It has been empirically determined that the stability of proteins can be expressed as a linear function of the denaturant concentration (see inset panel of Fig. 1.1) (Tanford 1968). Consequently, in order to obtain the thermodynamic stability of a protein, expressed by the change in free energy between the native (N) and the denatured (D) state ($\Delta G_{D\text{-}N}$), the linear extrapolation method is routinely used. In these studies, the protein is unfolded with a chemical denaturant and a spectroscopic signal, generally fluorescence or circular dichroism, is recorded as a function of denaturant concentration. Following this approach, curves with sigmoidal shape are typically obtained (Figure 1.1). By applying simplified assumptions on the effect of a denaturant (Myers, et al. 1995), it is possible to estimate the protein stability in the absence of denaturant with some confidence (Pace 1986).
Moreover, this analysis allows to obtain a parameter, called $m_{D\text{-}N}$ value, which describes the co-operativity of the process. The $m_{D\text{-}N}$ value is the slope of the variation of the $\Delta G_{D\text{-}N}$ with denaturant concentration; it has been

shown, on a purely empirical base, that this value is correlated with the number of residues and the surface area exposed to solvent during unfolding (Tanford 1968). $m_{D\text{-}N}$ value is expressed in kcal mol^{-1} M^{-1}.

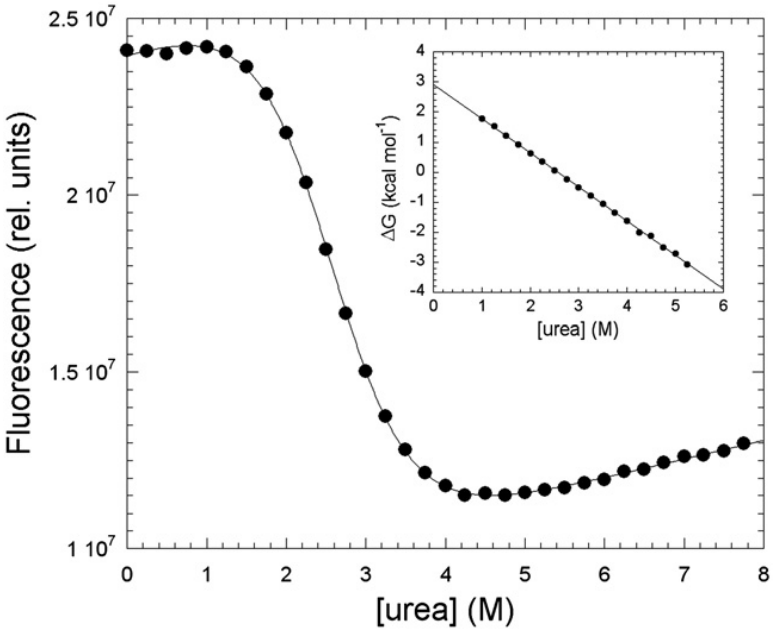


Figure 1.1. *Equilibrium unfolding of the second PDZ domain from PTP-BL (PDZ2) monitored by fluorescence (Gianni, et al. 2005). Inset Panel: Linear free energy extrapolation. A quantitative analysis of the observed spectroscopic signal as a function of denaturant allows to estimate the stability of the protein in the absence of denaturant and to calculate the $m_{D\text{-}N}$ value, which describes the co-operativity of the process.*

1.2.2 Kinetic studies

The kinetic approach provides detailed information about the folding pathway of a protein and the reaction dynamics. Kinetic studies are based on rapid perturbations of the equilibrium conditions that change the energetic of the system. The perturbation induces a relaxation process to a new equilibrium which allows to estimate the relaxation rate of the reaction.
Folding kinetics of small fast-folding proteins often exhibit two-state folding behaviour, as described by Jackson and Fersht (Jackson and Fersht 1991).

Due to the extremely fast folding of these small proteins, stopped-flow mixing devices with dead times in the millisecond range are usually required, and the folding reaction is monitored by following the change in circular dichroism or Trp fluorescence signal. The speed of the folding reaction is measured by rapidly diluting the protein from a high denaturant concentration, where it is completely denatured, into a low denaturant concentration where the folding reaction will occur. On the other hand, the speed of the unfolding reaction is measured by diluting the native protein from an aqueous buffer into a solution with high denaturant concentration. Folding and unfolding rate constants at a variety of denaturant concentrations, are determined by fitting the folding and unfolding curves to exponential equations. In the case of two-state folding, only a single exponential folding phase is observed (Figure 1.2), whereas multi-state systems may lead complex kinetics.

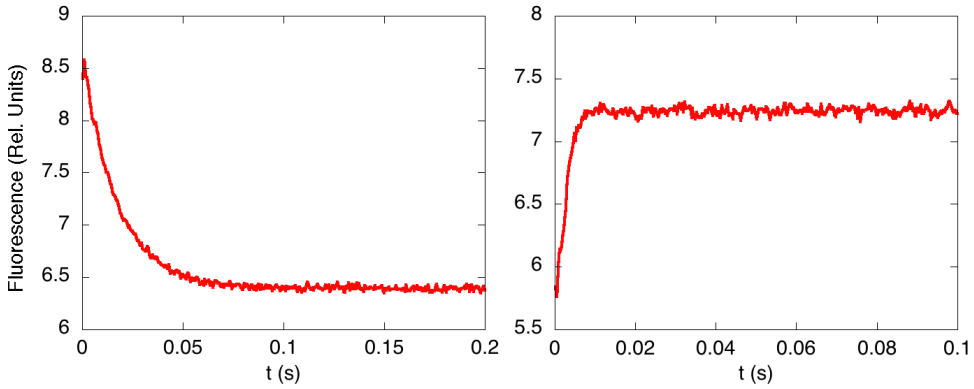


Figure 1.2. Stopped-flow time courses of the engineered protein G_A88. Unfolding (left panels) and refolding (right panels) reactions were monitored by the change in tryptophan fluorescence at pH 7.2 and 10 °C. The excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. The unfolding transition was initiated by a jump from urea 0 to 7.27 M; the refolding transition was initiated by a jump from urea 7 to 1.10 M. Both the curves are described by a single-exponential function.

If a protein folds following a two-state model (scheme 1.1), only the denatured and the native states are significantly populated (Buchner and Kiefhaber, 2005).

$$D \underset{k_U}{\overset{k_F}{\rightleftharpoons}} N \qquad \text{(Scheme 1.1)}$$

When a perturbation is imposed on the system, the observed rate constant k_{obs} is represented by the following equation:

$$k_{obs} = k_F + k_U \qquad \text{(Eq. 1.1)}$$

where k_U and k_F represent the unfolding and refolding rate constants respectively. In a two-state reaction, the logarithm of the observed (un)folding rate constants is linearly dependent on denaturant concentration (Jackson and Fersht 1991). Figure 1.3 shows the folding/unfolding rate constants expected for a two-state system. Because of its classical V-shaped appearance, this kind of semilogaritmic plot is currently called “chevron” plot by the protein folding community. From the analysis of a chevron plot, it is possible to calculate (a) k_F^{H2O} and k_U^{H2O}, which represent the extrapolated folding and unfolding rate constants in absence of denaturant and (b), m_F and m_U which reflect their dependence on denaturant concentration and correlate with the change in accessible surface area between the two ground states and the transition state in between.

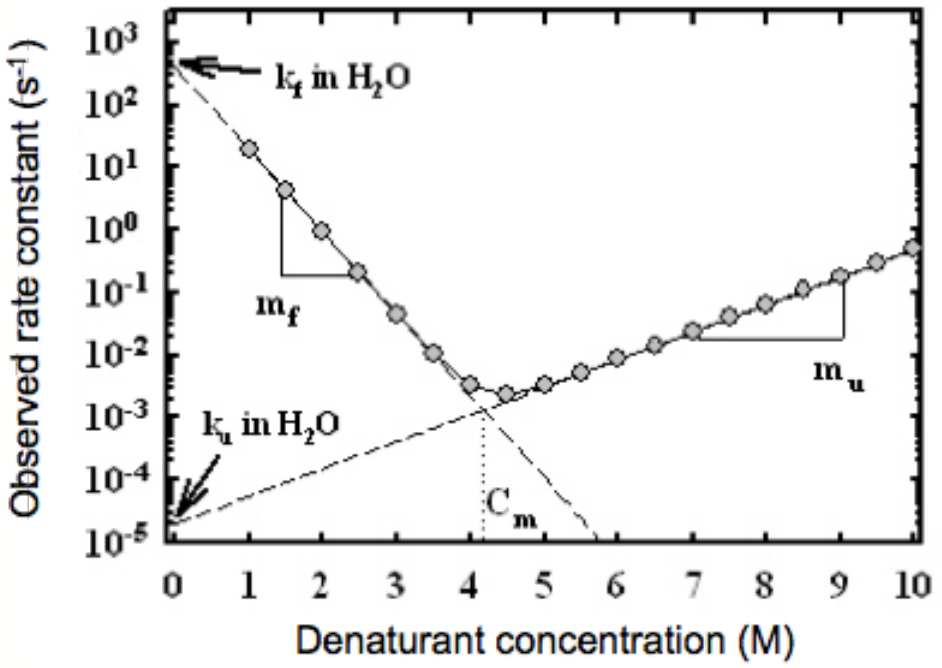


***Figure 1.3**. Semilogarithmic plot of the observed rate constants as a function of denaturant concentration (chevron plot); simulated data set. Dashed lines represent the best fit to a two-state model. As described in the text, for a two-state system, the observed rate constant is the sum of the folding and unfolding microscopic rate constants.*

1.3 Transition State Theory

Protein folding can be described as a unimolecular chemical reaction, in which the "reactant" (an unfolded protein) is converted to a "product" (the folded state). Unimolecular chemical reactions are typically governed by a single rate-limiting step, when the system passes through a high free-energy barrier called "transition state (TS)". Correspondingly, the presence of two, well-defined, thermodynamic macrostates in a folding reaction implies the existence of an energy barrier in between. Two-state folding transitions (Figure 1.4) are, thus, adequately described by the transition state theory (TST).

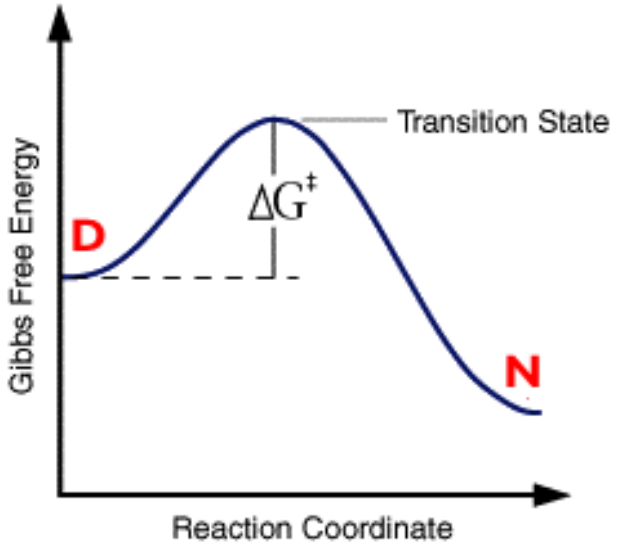


Figure 1.4. *Transition State Theory (TST). Schematic illustration of free energy profile for a folding reaction. The transition state is located between the denatured state D and the native (folded) state N of the protein. The folding rate constant is proportional to the exponential of the (negative) activation free energy, the free energy difference between the transition state and the denatured state D.*

Transition state theory was first proposed by Eyring in 1935 to explain chemical reaction rates (Eyring 1935). Equation 1.2 is used to describe this theory:

$$k = \left(\frac{\kappa K_b T}{h}\right) e^{(-E_a/RT)} \qquad \text{(Eq. 1.2)}$$

where k is the rate constant of the reaction, κ is called trasmission coefficient, K_b is Boltzmann's constant, h is Plank's constant, E_a is the activation energy for the reaction, R is the universal gas constant and T is the absolute temperature.
The term ($\kappa K_b T/h$) is a pre-exponential factor interpreted as the vibrational frequency, which refers to the maximal theoretical reaction rate ($\sim 10^{13}$ sec^{-1}) for simple molecular reactions.

According to Eyring's equation the observed folding rate constant, k_f, is proportional to the free energy of activation (i.e., the free energy difference per mole between the denatured and the transition states, $\Delta G^{\ddagger}$).
One standard extension of the TST assumes that the reaction's progress can be described by a reaction coordinate. In essence, an ideal reaction coordinate would be any degree of freedom that connects reactant**(**s**)** and product**(**s**)** along the lowest free energy continuous path on the free energy surface of the reaction; the highest free energy point along this path is the transition state (Figure 1.4). Moreover, while in ordinary chemical reactions, such as bond breaking in the gas phase, the TS corresponds to a well defined molecular structure, in protein folding the TS is described as an ensemble of high free energy conformations; it is therefore referred to as the transition state ensemble (J. N. Onuchic 1996; J.N. Onuchic 2004).

1.4 Protein Folding Intermediates

Understanding the role and structure of partially folded intermediates is of fundamental mechanistic importance for protein folding studies. Earlier work suggested that the folding of small single domain proteins generally conforms to an all-or-none behavior (Jackson 1991), involving simultaneous formation of secondary and tertiary structure (Itzhaki, et al. 1995). Following this view, folding occurs in a two-state fashion, via condensation around a marginally stable nucleus, and discrete intermediates tend to be avoided (Fersht 1995). When the inherent stability of folding nuclei is increased, however, even very simple protein systems appear to fold in a more complex fashion, with population, for example, of a partially folded intermediate, which may either transiently accumulate leading to multi-phasic kinetics, or be a high energy species en-route to the native state (Gianni, et al. 2003). In both cases, the system is represented by a three-state model (scheme 1.2).

$$D \underset{k_{ID}}{\overset{k_{DI}}{\rightleftarrows}} I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftarrows}} N \qquad \text{(Scheme 1.2)}$$

The presence of such local minima in the landscape is very difficult to

address experimentally (Bryngelson, et al. 1995) and intermediates may sometimes escape detection.

1.4.1 Identification of Folding Intermediates: The roll-over effect

The analysis of chevron plots is a common and powerful tool for detecting protein folding intermediates. A deviation from its classical V-shaped appearance may indicate that the system under analysis is kinetically more complex than a simple two-state reaction. This deviation, known as roll-over effect, in either folding or unfolding branches of chevron plots, may result from different scenarios, including accumulation of intermediates (Figure 1.5 a) (Matouschek, et al. 1990; Parker, et al. 1995; Ferguson, et al. 1999) or changes in the rate-limiting step between two discrete barriers, due to a high-energy never accumulating intermediate (Figure 1.5 b) (Oliveberg 1998; I.E. Sanchez 2003; Gianni, et al. 2009).
It is sometimes possible to detect the accumulation of low-energy intermediates by observing multiphasic kinetics and/or by analyzing fluorescence amplitudes (Khorasanizadeh, et al. 1996; Ferguson, et al. 1999; Capaldi, et al. 2001; Jemth, et al. 2004). On the other hand, the mechanisms involving a change in the rate-limiting step (i.e., involving a high-energy intermediate) are more difficult to be detected. They often result kinetically equivalent to another event which could be responsible for a roll-over effect: a structural change in one single transition state, involving Hammond effects on smooth energy barriers (Figure 1.4 c) (Otzen, et al. 1999; Ternstrom, et al. 1999).

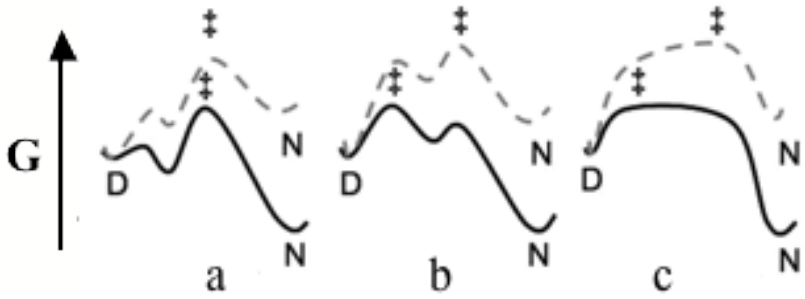


Figure 1.5. *Schematic free energy diagrams predicted for (a) three-state folding involving a low-energy accumulated intermediate; (b) three-state folding involving a high-energy never accumulating intermediate; (c) two-state folding involving a broad energy barrier. Solid black lines and dashed gray lines represent schematic energy diagrams for the two different scenarios expected at low and high denaturant concentrations, respectively. While the mechanisms involving a change in the rate-limiting step (b) or a smooth barrier (c) are generally indistinguishable, accumulation of a folding intermediate (a) results in double exponential time courses and may be addressed using ultrafast mixing/relaxation techniques (Shastry, et al. 1998; Capaldi, et al. 2001; Jemth, et al. 2004).*

A classical roll-over effect is present in the chevron plot reported in Figure 1.6. In this example, the observed rate constants follow a two-state behavior in the refolding and unfolding arms until 5.5 M guanidine concentration, whilst at higher denaturant concentrations, a deviation from linearity is detected.
The roll-over effect can be more or less pronounced depending on the difference between the *m* values of each state. In particular if the intermediate state is native like, i.e. highly compact, the roll-over can be described by a kink in the refolding branch. However in many cases, the folding intermediate can display unfolded-like properties; in such a case, the deviation from linearity is much more difficult to detect and deviation from two-state folding may be detected by comparison between the thermodynamic parameters obtained by equilibrium and kinetic experiments (Matouschek, et al. 1989; Fersht 1999; Gianni, et al. 2003).

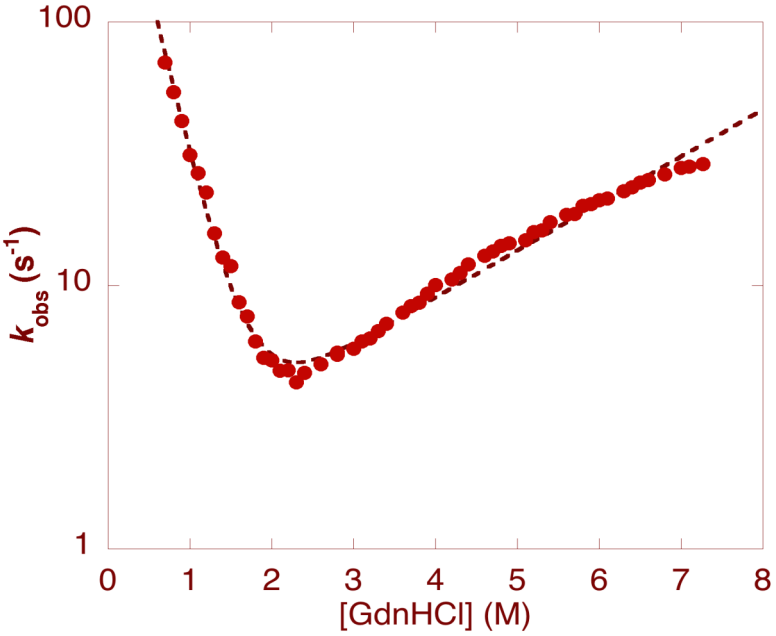


***Figure 1.6**. Chevron plot of the B1 IgG-binding domain of streptococcal protein G (generally called GB1) measured at pH 9.0 and 25 °. It is clearly evident that the data do not fit well to a two-state model (dashed line), since a deviation from linearity in the unfolding arm is present (roll-over effect).*

1.4.2 Alternative explanations for multi-state kinetics

Roll-over effects may result from events other than the presence of a high or low-energy intermediate state; thus analysis of chevron plots must consider all possibilities. In particular, a deviation from linearity in a chevron can result either from changes in the position of the transition state along the reaction coordinate or from transient aggregation of the denatured protein.
Changes in transition state ensembles consist in movements of the transition state position along the reaction coordinate as the denaturant concentration is increased. As shown in Figure 1.5 c, this instance is represented by a two-state model involving a broad energy barrier, assuming the protein to display plastic folding pathways characterized by a malleable TS (Ternstrom, et al. 1999; Cellmer, et al. 2007; Lindberg and Oliveberg 2007;). In other cases, the TS is surprisingly robust and maintains its structural features when the system is perturbed, for example by altering solvent conditions or by

mutagenesis (Jackson and Fersht 1991; Itzhaki, et al. 1995). In spite of the remarkable conceptual differences invoked by mechanisms implying either a malleable or a robust TS, it is extremely difficult, if not impossible, to unequivocally discriminate between these two models (Scott, et al. 2004). This is not surprising, because these two models correspond to extreme manifestations of a more complex scenario, whereby folding is characterized by a rough energy landscape (Hyeon and Thirumalai 2003) as detailed in "*The Energy Landscape Theory*" section (paragraph 1.7.2).

Another mechanism underlying complexity is due to the presence of transient aggregates i.e. association of two or more non-native protein molecules. This event driven by hydrophobic interactions, results in the formation of amorphous structures that lack long-range order. Because aggregation is sensitive to protein concentration, monitoring the kinetics as a function of concentration is a mandatory control to exclude aggregation artefacts (Figure 1.7).

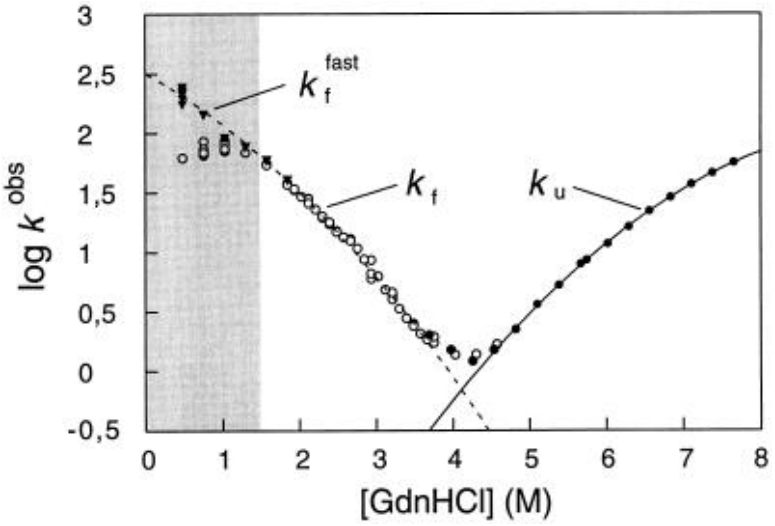


Figure 1.7. *An example of complexity due to aggregation and broad activation barrier. GdnHCl dependence of the rate constants for folding and unfolding of the spliceosomal protein U1A (Silow and Oliveberg 1997). The deviation from two-state folding observed at low [GdnHCl] (○), found also for other proteins, may be mistaken as accumulation of an intermediate. With U1A, this deviation was showed to be caused by transient aggregation of the denatured protein under refolding conditions (conditions where aggregation occurs are marked grey). At low protein concentrations the denatured protein remains monomeric during the refolding process while at higher protein concentrations the denatured protein aggregates in the dead-time of the stopped-flow instrument, giving rise to a retardation of the refolding rate. A part from transient aggregation commented above, it is evident the presence of a pronounced specular curvature in both the arms of U1A chevron plot. Silow and Oliveberg demonstrated that specular curvature is due to a broad activation energy barrier separating the native and the denatured states.*

1.5 Experimental strategies for Folding studies

In analogy with classical organic chemistry, the best strategy to unravel the mechanism of the folding reactions would be to isolate all the intermediate states and characterize the transition states in between. Indeed, many advances in protein folding studies have been achieved by isolating

intermediates and studying their structure. However isolation and characterization of folding intermediates is often impossible because they are short-lived and at very low concentration (due to cooperativity). Indeed, for many single domain proteins, only the fully native and the fully denatured states may be populated at equilibrium. In these cases, a key role in protein folding studies has been played by the description and the characterization of folding transition states. A great contribution is due to A. Fersht and co-workers (Fersht, et al. 1992) who introduced in the late 1980s the so-called Φ value analysis, contributing to the description of protein folding mechanism at nearly atomic resolution.

1.5.1 Φ–Value analysis

The protein engineering approach, termed Φ value analysis, was developed in the laboratory of Alan Fersht with the purpose of unveiling the structure of the protein folding transition states (Fersht, et al. 1992). Following this method, the degree of structure formation of individual residues in the transition state is actually inferred from analyzing the effect of single-site mutations on folding rates and stability. The Φ value analysis has been extensively used to investigate the transition state of many proteins (Matouschek, et al. 1989; Gianni, et al. 2007a; Zarrine-Afsar, et al. 2010; Banachewicz, et al. 2011).
The Φ value is calculated as the ratio of the energetic perturbation induced on the transition state versus that induced in the native folded state, introducing a non-disruptive mutation, intended to cause a small perturbation:

$$\Phi = \frac{\Delta\Delta G^{*}_{TS-D}}{\Delta\Delta G_{N-D}} \qquad \text{(Eq. 1.3)}$$

where $\Delta\Delta G_{N-D}$ is the change induced by mutation in the free energy of folding and $\Delta\Delta G^{*}_{TS-D}$ that induced in the activation energy of folding.
Φ values normally range from 0 to 1. A Φ-value near unity indicates that the TS is energetically perturbed upon mutation as much as the native state. This effect indicates that in the TS the mutated residue is engaged in fully native contacts (i.e. it has all its native interactions established) (Figure 1.8 right panel). On the other hand, a Φ-value near zero is taken as evidence that the TS is not energetically perturbed by the mutation, while the native state is (i.e. the mutated residue is as unstructured in the TS as in the denatured

ensemble) (figure 1.8 left panel).

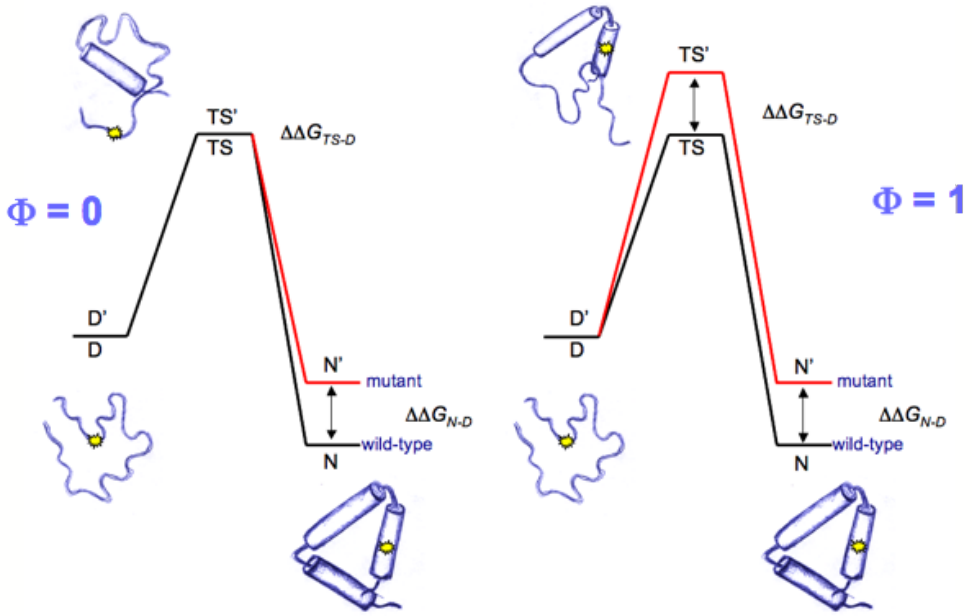


Figure 1.8. *Φ-value analysis. The depicted energy profiles represent two different scenarios in which a hypothetical probed residue (shown in yellow on the structures) gives rise to different effects in the folding activation energy and in the stability of the protein, depending on the structure of the transition state. Left panel: the site of mutation is in an unstructured region of the transition state (Φ=0). Right panel: the probed residue is highly native-like in the transition state (Φ=1).*

Experimentally determined Φ values are generally in between 0 and 1 (most often between 0.3 and 0.7). The traditional interpretation of fractional Φ values is, however, not straightforward as they might indicate the existence of multiple folding pathways or a unique transition state ensemble with genuinely weakened interactions (Oliveberg and Fersht 1996).
Also the interpretation of the so-called non-classical Φ values (Φ> 1 and Φ < 0), that are seldom observed, is not straightforward. While, in some cases, these values are due to an incorrect mutation, in some others, they appear genuine. Negative Φ values can be observed when the native state is

destabilized while the transition state is stabilized and vice-versa. Φ values higher than 1 may be detected when the transition state is affected by the mutation more than the native state (Gianni 2012 in press). Because unusual Φ values are often indicative of non-native contact formation (Ozkan, et al. 2001), they can be used to detect local misfolding in transient intermediates either *en-route* to the productive folding pathway, as in the case of the Im7 protein (Capaldi, et al. 2002), or acting as off-pathway kinetic traps, like in the case of the circularly permuted PDZ domain of D1p protein (Gianni, et al. 2010).

1.5.2 Comparative studies on Protein Folding: Homologous Proteins

Current knowledge on the protein folding reaction has been achieved by extensively characterizing the folding mechanisms of simple globular proteins (Jackson 1998). However, given the diversity of protein folds and especially amino acid sequences, it is extremely difficult to draw general rules by studying folding pathways of individual proteins. In fact, when considering the folding of different proteins at least three key variables may jeopardize a comparison: i) sequence composition, ii) native and iii) denatured states structure variability.

A powerful strategy to elucidate some of the relationships between sequence information and folding mechanism is to study proteins that differ in sequence but share the same overall fold (Chiti, et al. 1999b; Clarke, et al. 1999a; Martínez and Serrano 1999; Riddle, et al. 1999; Friel, et al. 2003; Travaglini-Allocatelli, et al. 2003; Travaglini-Allocatelli, et al. 2005; Chi, et al. 2007; Calosci, et al. 2008). This strategy assumes that general correlations between amino acid sequences and folding pathways may be extrapolated by comparing folding processes for different members of a given protein family. The final goal is to identify the limited number of sequence determinants to achieve the common fold. Thus, the protein folding problem is often investigated using proteins with a low degree of sequence identity, yet adopting essentially the same fold.

Over the years, it has been found that the mechanism of folding is, generally, conserved in protein families (Chiti, et al. 1999a; Gianni, et al. 2001a; Zarrine-Afsar, et al. 2005) suggesting that the native topology is one of the main factors controlling protein folding (Baker 2000).

An interesting example of comparative study is represented by the cytochrome c (cyt c) family, which has been used extensively in different

laboratories, as a model system for folding studies (Elove, et al. 1994; Bai, et al. 1995; Colon, et al. 1996; Gianni, et al. 2001a; Brunori, et al. 2003; Travaglini-Allocatelli, et al. 2003). The c-type cythocromes are a large family of globular proteins with a characteristic α-helical fold and a covalently bound heme group. In 2004, using a comparative approach, it was shown that very different members of the cyt c protein family share a common folding mechanism (Travaglini-Allocatelli, et al. 2004). By comparing the folding kinetics of cytochrome c_{551} from the mesophilic bacterium *Pseudomonas aeruginosa* (Pa-cyt c_{551}) and cytochrome c_{552} from the thermophilic bacterium *Hydrogenobacter thermophilus*, it was demonstrated that the folding transition states of these two proteins share some similarities, in spite of large differences in thermodynamic stabilities. Analysis of extensive kinetic data available on many eukaryotic cyt c has indicated that similar species are populated along the folding pathway, enabling to propose a consensus mechanism.

The PDZ domains represent another example of comparative folding study. They constitute a large family of protein-interaction modules that mediate protein-protein recognition by binding to short amino acid sequences (Fanning and Anderson 1999). Different classes of PDZ domains recognize specific C-terminal sequences (PDZ motifs) on a variety of protein substrates. Analyzing and comparing the kinetic folding mechanisms of five related but distinct PDZ domains, it has been found that, despite their low sequence identity and apparent folding complexity, the folding reactions for PDZ domains can be explained by a model with an intermediate and two transition states that are rather conserved with regard to their positions along the folding reaction coordinate (Chi, et al. 2007). In particular, through a combination of Φ-value analysis and molecular dynamics simulations, it was found that the late transition states are much more structurally similar than the early transition states (Calosci, et al. 2008). Surprisingly, in a further study on a topological PDZ mutant (cpPDZ2), such as a circular permutant, where the native N- and C-termini were joined and the sequence cleaved in a different position (Ivarsson, et al. 2008), it was shown that, although circular permutation introduces a significant destabilisation of the native state, the folding kinetics of cpPDZ2 reveal a remarkable stabilisation of the folding intermediate, which accumulates transiently during folding.

1.5.3 A complementary approach in Protein Folding studies: Proteins with high sequence identity but different fold

The analysis of heteromorphic proteins, that is proteins with high sequence identity but different structure, can be considered as a novel and alternative approach complementary to the folding studies on protein families.
While the analysis of protein families allows to control one of the main factors governing protein folding, i.e. native-state topology, the study of heteromorphic proteins allows to control another relevant parameter: the amino acid sequence of a protein.
It is known that, proteins with a significant similarity in their amino acid sequence are expected to have the same fold. In fact, analysis of the protein data bank (PDB) reveals that a sequence similarity of 40% nearly always leads to a conserved three-dimensional structure and function (Wilson, et al. 2000). This observation provoked Rose and Creamer in 1994 to issue the ‘‘Paracelsus Challenge’’, whereby the protein folding community was charged with the task of designing two proteins that were at least 50% identical but possessed different folds (Rose and Creamer 1994). Amazingly, this goal was fully achieved in only 3 years, when Dalal and co-workers designed a sequence that, in spite of being 50% identical to a mostly β-sheet protein, folded into a four-helix bundle (Dalal, et al. 1997). Since then, several other scientists have achieved similarly impressive feats of design (Davidson 2008). In 2008, ambitious work by Bryan and co-workers led to the design of a pair of proteins with an extraordinarily high degree of sequence identity but different structure and function (Alexander, et al. 2007; He, et al. 2008). In particular, the sequences of two domains from streptococcal protein G were subjected to an iterative design of heteromorphic proteins leading the authors to produce three pairs of protein G variants with an increasing level of sequence identity (30%, 77% and 88% respectively) (Alexander, et al. 2007; He, et al. 2008). Two proteins, sharing 88% sequence identity (49 out of 56 amino acids), are called G_A88, which is mostly α-helical (the 3 helix bundle protein A fold), and G_B88, displaying the α+β protein G fold (Figure 1.9); yet they display divergent structures and functions that were similar to the respective wild-type proteins. The study of this remarkable protein engineering achievement offers some unique opportunities for a complementary analysis on protein folding mechanisms and allows to pose two key questions: (1) At which stage of its folding pathway does a protein commit to a given topology? and (2) Which residues are crucial in directing folding to a given structure?.

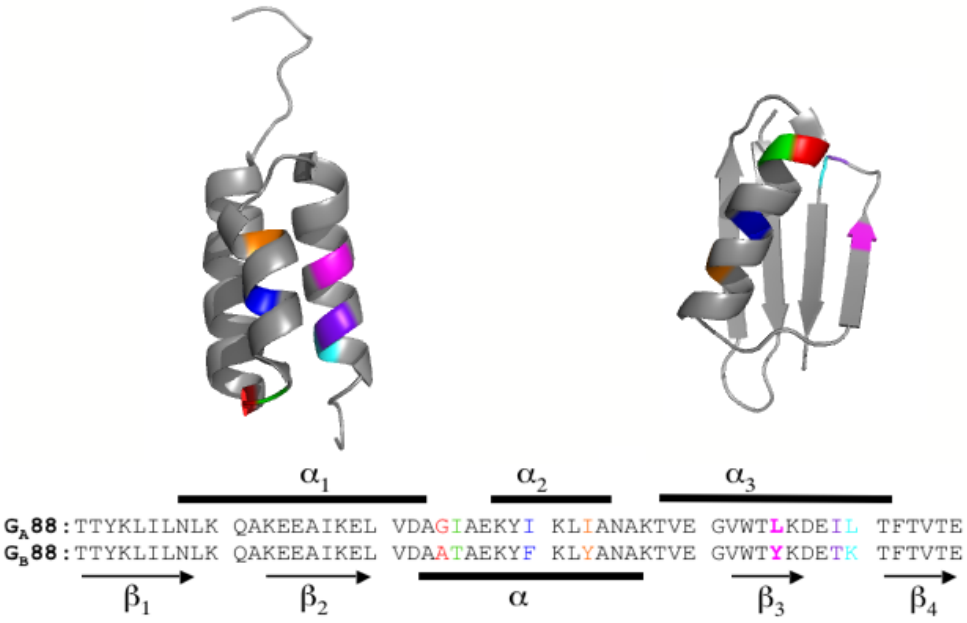


Figure 1.9. *G_A88 and G_B88 structures. Sequence alignment and secondary structure are shown below. The 7 residues that differ between the two proteins are shown in different colours: 24 (red), 25 (green), 30 (blue), 33 (orange), 45 (magenta), 49 (violet) and 50 (cyan).*

1.6 Theoretical studies

A crucial approach in protein folding studies is represented by the comparison between experimental and theoretical results. Theoretical models and computer simulations have greatly advanced our understanding of protein folding and have helped the interpretation of experimental data. There is a synergy between theory and experiment, the former providing testable models and the latter aiming at validating hypothesis (Szilagyi 2007).
One of the principal tools in the theoretical approach to the structure and function of biological molecules is the method of molecular dynamics (MD) simulations (Karplus and McCammon 2002). Simulations can help in identifying or predicting transition and intermediate states along the folding pathway, they provide estimates of the rate of folding and in some cases, predict the final, folded structure.
Small proteins typically fold in the several microseconds to seconds

timescale; detailed atomistic simulations, however, are currently limited to the nanosecond to microsecond regime. Therefore, simulation of folding requires either simplified models or special sampling methods, both of which introduce some approximations. To simulate unfolding, the simplest method of increasing sampling is to increase the temperature of the simulation to 500 K or more where the native structure of the protein is melting within a few nanoseconds. This approach has been applied to several small proteins, including bovine pancreatic trypsin inhibitor (Kazmirski and Daggett 1998b), lysozyme (Kazmirski and Daggett 1998a), myoglobin (Tirado-Rives and Jorgensen 1993), barnase (Wong, et al. 2000), ubiquitin (Alonso and Daggett 1998), the SH3 domain (Day and Daggett 2003), etc. Features of the unfolding process, such as the transition-state ensemble or the unfolded ensemble, have in some cases shown remarkable agreement with experimental results (Tsai, et al. 1999).

1.7 Folding mechanisms

As mentioned previously, in 1968 Levinthal postulated the existence of folding pathways. The concept of folding pathways gave rise to a number of models describing the folding process, which are briefly summarized below. The various models of protein folding are not necessarily exclusive; they try to grasp different aspects of folding.

In 1973, the nucleation model (Wetlaufer 1973) tried to resolve Levinthal's paradox by assuming that the rate-limiting step in the folding process is a nucleation event, presumably the formation of smaller structural units, and once nucleation occurs the nuclei grow fast and the folding process rapidly completes. This model is not consistent with the large number of observations where folding intermediates were observed.

The ‘‘diffusion–collision model (DC)" (Karplus and Weaver 1976) implies fluctuating microdomains (portions of secondary structure or hydrophobic clusters) that move diffusively and repeatedly collide with each other. Productive collisions lead to coalescence into intermediates, which may involve microdomains that are not necessarily contiguous along the protein sequence. Folding proceeds as a series of coalescence events that might either follow a unique order (sequential folding pathways) or explore different routes (parallel folding). Also the so-called ‘‘jigsaw puzzle model’’ (Harrison and Durbin 1985) denied the necessity of a unique, directed folding pathway and stated that each protein molecule can follow a different route to the

native structure, just like there are multiple ways to solve a jigsaw puzzle.
The introduction of the ‘‘hydrophobic collapse model’’ (Dill 1985), based on the intuition of C. Tanford, led to the view that the hydrophobic effect is the main driving force of folding, and the process starts with a rapid collapse of the chain, expulsion of the water and formation of the secondary structure.
The ‘‘framework model’’ (Baldwin 1989) stated that the folding process is hierarchical, starting with the formation of the secondary structure elements, and the docking and the organization of the preformed substructures is the rate-limiting step.

The model named ‘‘nucleation–condensation (NC)’’ represents an attempt to unify the features of both the framework and the hydrophobic collapse mechanisms (Fersht 1995; Fersht 1997). In this model, secondary structure and hydrophobic interactions form nearly simultaneously and synergistically (Daggett and Fersht 2003), leading to the formation of a weak structured local nucleus. The nucleus is composed of a set of adjacent residues, stabilized by long-range interactions that are formed as the rest of the protein collapses around the nucleus: formation of the nucleus (nucleation) (Wetlaufer 1973) is coupled with a more extended formation of structure (condensation).

1.7.1 Nucleation-Condensation and Diffusion-Collision: Extreme manifestations of a common mechanism for Protein Folding

Among the models for protein folding, Nucleation-Condensation (NC) and Diffusion-Collision (DC) mechanisms are, currently, the two models generally used to describe the folding of small single-domain proteins. A critical test to distinguish between them is represented by the analysis of the magnitude and distribution of Φ values for a given protein.
The transition state of a protein which folds according to the DC model, displays heterogeneous structure localization, with regions having Φ values close to 1 and others displaying Φ values close to 0, distributed in contiguous blocks, indicative of preformed secondary structure elements (or independent microdomains) (Gianni, et al. 2003).
On the other hand, in the case of the NC model, the nucleus may be identified by the few residues displaying higher Φ values; the native-like structure in the TS should gradually decrease with a smooth gradient of decreasing Φ values from the nucleus (Itzhaki, et al. 1995).
In 2007, a work on a PDZ domain (second PDZ repeat from Protein Tyrosine

Phosphatase-Bas Like, PDZ2) (Gianni, et al. 2007a), performed in our laboratory, suggested that NC and DC models may represent extreme manifestations of an underlying common mechanism and that proteins may appear to fold by either NC or DC depending on the inherent stability of their secondary structure elements. As reported above, it was already shown that the folding pathway of PDZ2 proceeds through two consecutive TS barriers (Gianni, et al. 2005). The interactions formed in the two distinct TSs were mapped by Φ-value analysis. Surprisingly, the first TS was characterized by many mutants displaying $\Phi = 0$ and only some mutants having fractional Φ values, suggestive of a NC mechanism. On the other hand, the second TS displayed characteristics of the DC mechanism with several Φ-values close to 1, and the rest displaying fractional Φ. The folding of PDZ2 has been suggested as a paradigmatic example in which the two extreme models are manifested in one single protein. According to this unifying folding mechanism, the folding of small globular proteins is suggested to involve three major events: (1) formation of a weak nucleus that determines the native-like topology of the structure, (2) a global collapse of the entire polypeptide chain, and finally (3) consolidation of the remaining partially structured regions to achieve the native state conformation (Gianni, et al. 2007a).

1.7.2 The Energy Landscape Theory

Since the process whereby a protein acquires its native three-dimensional shape involves formation of many non-covalent weak bonds, a classical one-trajectory view of protein folding is likely to be an over-simplification of the underlying mechanism. An original viewpoint emerged when the concept of energy landscape for a protein (Frauenfelder et al. 1991) was extended to folding. Energy landscapes are mathematical devices that help to understand the microscopic behavior of a molecular system (Bryngelson, et al. 1995). Although energy landscapes are, by definition, high-dimensional surfaces, they are often pictured as a surface in three-dimensions. In these pictures, the vertical axis represents the free energy and the horizontal axes represent the conformational degrees of freedom of the polypeptide chain. The statistical energy landscape perspective describes folding as taking place on a rugged free-energy landscape in which free-energy barriers separate ensembles of states displaying different levels of structural heterogeneity, such as the folded (F) and the unfolded (U) states. This view likens the energy landscape of a protein to a funnel, with the native structure at its global minimum

(Figure 1.10).

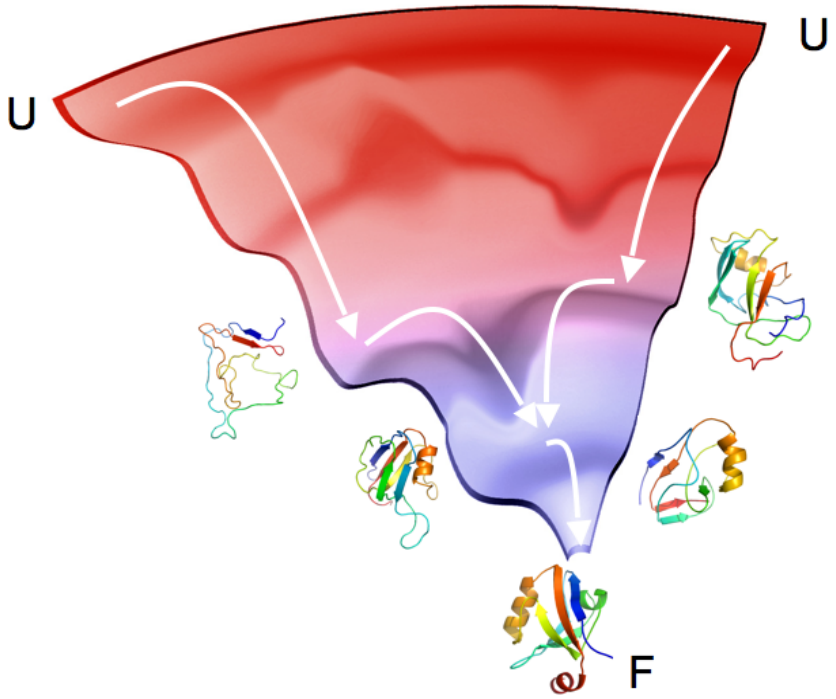


Figure 1.10. *Schematic representation of a folding funnel: the free energy drives the polypeptide towards the folded state (F), the most stable conformation, and the conformational entropy dramatically decreases as the native state is approached. Implicit in this view is that folding may occur via alternative parallel pathways. This concept is graphically depicted using the structures of the early and late folding transition states for different PDZ domains, as obtained in ref. (Calosci, et al. 2008).*

According to the landscape theory, the plasticity of folding pathways assumes that proteins can be rerouted through the energy landscape by mutational (Wright, et al. 2003), topological (Lindberg and Oliveberg 2007) or solvent perturbations (Gianni, et al. 2007b; Otzen and Fersht 1998). Thus, the polypeptide chain may fold by different pathways, potentially adopting multiple partially folded ensembles *en route* to the native state (J.N. Onuchic 2004). Alternative folding pathways, involving different nucleation motifs, can be selectively stabilized *via* loop entropy perturbations, such as circular permutation. Over-stabilization of a nucleus may lead to frustration of the

folding landscape, involving the segregation into local minima that compete for producing the native state. An example of such a scenario is represented by the D1pPDZ domain, a naturally occurring circularly permuted variant that displays an off-pathway kinetic trap characterized by a misfolded N-terminal hairpin incorrectly docked on an otherwise native-like structure (Gianni, et al. 2010).
Whilst the presentation of the funnel model provided a novel outlook on protein folding, a detailed experimental description of such a complex scenario is still a challenge.

1.8 The role of the Denatured State in Protein Folding

A key issue for the proper evaluation of folding kinetics and mechanism is the starting state of the process, i.e., the denatured protein (Dill and Shortle 1991). The denatured state of proteins has received significantly less attention than the folded state since biological function is a property of the native state. Further, the denatured state is generally complex and poorly defined and, cannot be studied by standard structural approaches, excluding sophisticated NMR methods. It is important to emphasise the distinction between the denatured state which prevails under conditions that favour folding being a transient species *en-route* to the folded state, and the completely unfolded state which is a more expanded chain populated at high concentrations of denaturant or high temperatures, which is behaving as a true random coil. For example, in a typical stopped-flow refolding experiment, the unfolded state refers to the protein in the presence of high denaturant concentrations, while the denatured state is generally the entity into which it is rapidly converted after the denaturant has been rapidly diluted and the protein is beginning to fold.
Over the last decade, an increasing number of experiments provided evidence for a highly dynamic unfolded state (Dyson and Wright 2002; Dyson and Wright 2004) bearing distinct loosely structured conformational states (Chattopadhyay, et al. 2005; Pletneva, et al. 2005), under denaturing conditions. There is, indeed, considerable evidence that even in strong denaturants such as 6M guanidine and 9M urea, some dynamic structural constraints may remain in disordered polypeptides (Dill and Shortle 1991).
As the starting state of the folding reaction, the structural and thermodynamic biases of the denatured state may hold important clues into the ‘folding

code', which (unlike the genetic code) has proven difficult to decipher because of its redundancy (Bowler 2011).

The earliest study on protein denatured states was carried out in the late '60s, when Tanford asked whether denatured proteins are completely random coils in 6M guanidine, which unfolds nearly all water-soluble proteins (Tanford 1968). He used hydrodynamic properties such as intrinsic viscosity to characterize the overall dimensions of the polypeptide chain, after reducing any disulfide bonds present. Moreover, optical rotatory dispersion, the forerunner of circular dichroism, was used to detect secondary structure. Tanford found that denatured proteins in 6M guanidine can be described as random coils in the sense of being devoid of all secondary and tertiary structure. He pointed out, however, that thermally denatured proteins in water have optical rotatory dispersion spectra indicative of some secondary structure, possibly residual native structure.

Since then, an increasing number of examples of proteins that maintain the gross features of their native topology in the denatured state have emerged (Bond, et al. 1997; Shortle and Ackerman 2001; van Gunsteren, et al. 2001; Fersht and Daggett 2002).

The advent of site-directed mutagenesis methods revolutionized the study of the thermodynamics of protein folding. Careful analysis of important interactions in the native state became possible. However, some mutagenesis studies produced results that were difficult to reconcile with effects on native state interactions. Studies on Staphylococcal Nuclease (SNase) showed that the equilibrium denaturation $m_{D\text{-}N}$-value could vary dramatically with single-site mutations, both increasing and decreasing, by up to 50% (Shortle and Meeker 1986; Shortle 1995). As pointed out earlier, theoretical (Schellman 1978) and empirical (Tanford 1968; Myers, et al. 1995) studies have demonstrated that the *m*-value is correlated to the change in solvent-exposed surface area upon unfolding. While the decrease of $m_{D\text{-}N}$-value could be explained by a more labile native state with greater solvent exposure, it was more difficult to rationalize the highly compact native state becoming more compact in the mutant variants which showed an increased $m_{D\text{-}N}$-value (Bowler 2007). Thus, it was proposed that these mutants might be characterized by changes in the permanence of residual structure in the denatured state. A number of spectroscopic studies, including NMR, supported these conclusions (Shortle 1995).

The Nucleophosmin C-terminal domain (Cter- NPM1), a small 3-helix bundle protein, represents a recent example of protein with a residual structure in the denatured state. It has been clearly demonstrated that its denatured state retains some malleable, native-like, residual structure (Scaloni, et al. 2009). A very innovative method has been applied to this

protein to probe the importance of different amino acids in stabilizing its collapsed denatured state. The method combines two unrelated perturbations, i.e. mutations and stabilization by salt. The results show that individual amino acids of two α-helices give a significant contribution to the stability of the collapsed denatured state (Scaloni, et al. 2010).

Deviations from random-coil behavior in the denatured state are of practical importance if they influence protein folding, stability, or function, or if they compromise the analysis and the interpretation of experiments (Cho and Raleigh 2009). Moreover, non-random native-like structure in the denatured state of fast-folding proteins is suggestive of an important role for the denatured state in efficient folding (Meng, et al. 2009; Neuweiler, et al. 2009). Flickering structure in the denatured state could directly provide the elements of the transition state; residual structure in the denatured state could also provide a template for assembling more disordered parts of the denatured state in the transition state. The uncertainties surrounding the role of residual structure in the denatured state in promoting efficient folding indicate that this is an area in need of further investigation (Bowler 2011).

AIM OF THE THESIS

The problem of spontaneous folding of a polypeptide into compact, highly organized three-dimensional structures represents a fundamental challenge in modern Biochemistry (Kennedy and Norman 2005).
The sequence of a protein contains all of the information for folding to the functionally competent state (Anfinsen 1961). However, not all amino acids of a polypeptide chain are equally important in specifying which fold is adopted. The concept of "key residues" was previously proposed to describe the formation of "key contacts" in specific transition states of folding as a critical step to trigger "downhill" backbone collapse (Abkevich, et al. 1994; Vendruscolo, et al. 2001). The design of two proteins, called G_A88 and G_B88, with an extraordinarily high degree of sequence identity (i.e. 88%) but different structure and function (Alexander, et al. 2007), offered to us a unique opportunity of investigating the role of key residues in the mechanism of protein folding.
The first part of the experimental work, described in this thesis, is focused on a detailed characterization of the folding and unfolding of both G_A88 and G_B88, the first being α helical and the second mostly β sheet. The results obtained under a variety of solvent conditions indicate the presence of a pH-sensitive residual structure in the denatured state of G_B88, which is not observed for G_A88. Protein topology is, therefore, committed very early along their folding pathway being "imprinted" in the residual structure of the denatured state, and this weak, loosely defined topology is sufficient for dictating the folding pathway. By integrating our kinetic results with molecular dynamics studies, it appears that, while only a few residues are responsible for the selective stabilization of the two topologies, no residue acts as a unique gatekeeper in the selection of protein topology on the folding of G_A88 and G_B88. Both experiments and simulations on the folding of these two proteins suggest that native topology might be already pre-sculpted in the denatured state, where incipient nuclei are present.
The surprising finding that the folding pathway of G_A88 and G_B88 diverges as early as in their denatured state, prompted us to carry out a systematic and very extensive analysis of each of the heteromorphic variants designed by Philip N. Bryan. In fact, as reported in the Introduction, G_A88 and G_B88 were obtained by an iterative design of heteromorphic pairs using genetic engineering (Alexander, et al. 2007). All the pairs of G_A and G_B proteins

produced by Bryan and co-workers, display structure and function similar to the respective wild-type proteins which are both known to interact specifically with two different macromolecules (i.e. serum albumin and IgG, respectively). The study of these protein variants allowed us to explore the sequence space associated with a given protein structure. To this purpose, we performed a detailed characterization of the folding pathway of all six G_A and G_B proteins, i.e. G_A30, G_B30, G_A77, G_B77, G_A88 and G_B88, considering also the natural G_B domain, called GB1, a popular system for protein folding studies.

The results were obtained by exploring the folding process of GB1 at different pH values, and by performing an extensive Φ-value analysis of the engineered G_A and G_B variants. Analysis of data shows that the mechanism of folding of all these G_A variants conforms to a co-operative two-state model with a structurally conserved transition state. On the other hand, the folding pathway of the G_B variants appears less co-operative and more complex, revealing the presence of an intermediate and the existence of a variable organization in the transition states. Therefore, while the G_A-fold is populated via a single pathway with one conserved intrinsic nucleus, the G_B-fold can be reached by formation of more than one nucleus, each being selectively stabilized by altering the amino acid sequence via site-directed mutagenesis.

METHODS

3.1 Site-directed Mutagenesis

G_A and G_B genes were cloned into the vector pG58, which encodes an engineered subtilisin pro-sequence as the N terminus of the fusion protein (Ruan, et al. 2004). pG58 plasmid containing the different G_A and G_B genes was generously provided by Prof. Philip N. Bryan (University of Maryland, USA). These genes were used as templates to perform site-directed mutagenesis. All mutants were obtained by using the QuikChange mutagenesis kit (Stratagene) according to the manufacturer's instructions.

3.2 Protein Expression and Purification

Expression of the wild-type proteins and their mutants was obtained in Luria Bertani (LB) medium containing 100μg/ml ampicillin. Cultures (1 l in 2 l flasks) were shaken at 180 rpm and grown at 37°C until OD600 arrived to 0,5; then protein expression was induced with 1 mM IPTG (isopropyl ß-d-thiogalactoside). After induction, cells were grown for 20 hours at 25°C and then collected by centrifugation.
Cells were resuspended in 50 mM $NaPO_4$ buffer (pH 7.2) and sonicated. After sonication, cell extract was centrifuged 30' at 13000 rpm to remove any insoluble material.
Proteins were purified employing an affinity-chromatography previously developed (Ruan, et al. 2004). Soluble cell extract of pro-domain fusion proteins was injected on a 5-ml Bio-Scale™ Mini Profinity eXact cartridge at 5 ml/min to allow binding and then washed with 10-column volumes of 100 mM $NaPO_4$ (pH 7.2) to remove impurities. To cleave and elute the purified target proteins, 15 ml of 100 mM NaF in the presence of 100 mM $NaPO_4$ (pH 7.2) were injected at 5 ml/min. After the first 10 ml, the flow was stopped and the column incubated for 30 minutes to allow complete cleavage. The purified proteins were then dialyzed to remove sodium fluoride. Purity was

checked by SDS/PAGE.

3.3 Equilibrium unfolding

Equilibrium denaturations were followed by Far-UV circular dichroism (CD) and fluorescence spectroscopy to monitor changes in secondary and tertiary structure, respectively.
CD measurements were carried out on a JASCO spectropolarimeter (Jasco, Inc., Easton, MD, USA), in a 1 cm quartz cuvette (Schellman). The spectra were recorded between 250 and 200 nm.
Fluorescence measurements were carried out on a Fluoromax single photon counting spectrofluorometer (Jobin-Yvon) with a cuvette of 1 cm light path. Tryptophan fluorescence emission spectra were recorded between 300 and 400 nm. The excitation wavelength was 280 nm.
Protein concentrations were typically 6 μM.

3.4 Stopped-flow measurements

Single mixing kinetic folding experiments were carried out on a Pi-star or on an SX-18 stopped-flow instruments (Applied Photophysics, Leatherhead, UK). The excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. In all experiments, performed at 25°C and 10°C, refolding and unfolding were initiated by a 11-fold dilution of the denatured or the native protein with the appropriate buffer. Final protein concentrations were typically 1 μM. The observed kinetics were always independent of protein concentration (from 0.5 to 5 μM), as expected from monomolecular reactions without effects due to transient aggregation (Silow and Oliveberg 1997).

3.5 Temperature-jump measurements

The observed kinetic rate constants of some G_A mutants, turning out to fast for stopped-flow apparatus, were obtained by Temperature-jump technique. The relaxation kinetics were measured as a function of guanidine or urea by using a Hi-Tech PTJ-64 capacitor-discharge T-jump apparatus (Hi-Tech, Salisbury, U. K.). Temperature was rapidly changed from 18 °C to 25°C and

from 4 °C to 10 °C with a jump-size of 7°C and 6°C respectively. 10 to 20 individual traces were averaged at given denaturant concentrations. The fluorescence change of N-acetyltryptophanamide (NATA) was used in control measurements. Protein concentration was typically 20 μM. Degassed and filtered samples were slowly pumped through the 0.5 x 2 mm quartz flow cell at 5 μl min during data acquisition. The excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter.

3.6 pH variation

Equilibrium and kinetic experiments were performed exploring a wide range of pH, from 2.0 to 10, using the following buffers: 50 mM Glycine /NaOH from pH 10 to 9.0, 50 mM Tris/HCl from pH 9.0 to 7.2, 50 mM sodium phosphate from pH 8.0 to 6.3, 50 mM Bis-Tris/HCl from pH 7.0 to 6.0, 50 mM sodium acetate from pH 5.5 to 3.8, 50 mM sodium formate from pH 3.4 to 3.0 and 50 mM sodium phosphate/phosphoric acid from pH 2.8 to 2.0. All reagents were of analytical grade.

3.7 Data analysis

3.7.1 Quantitative analysis of two-state equilibrium transitions

The folding-unfolding transition in globular proteins can be described, in general, as a two-state equilibrium process. Although there is now increasing evidence for complex behavior within such simple systems (Brockwell and Radford 2007), many small monomeric proteins show simple, unfolding transitions between native and denatured states. Under these conditions, if a given optical probe (y) is used to monitor unfolding, the observed signal y_{obs} will reflect the fractional population of the native and denatured state, being:

$$\frac{[D]}{[D]+[N]} = \frac{y_{obs} - y_N}{y_D - y_N} \qquad \text{(Eq. 3.1)}$$

$$\frac{[N]}{[D]+[N]} = 1 - \frac{y_{obs} - y_N}{y_D - y_N} \qquad \text{(Eq. 3.2)}$$

where y_{obs} is the optical measure at a given denaturant concentration, y_D and y_N are the signals of the denatured and native states respectively.

By applying the linear free energy extrapolation assumption, it can be postulated that:

$$\Delta G_{D-N} = \Delta G^{0}{}_{D-N} - m_{D-N}[denaturant] \quad \text{(Eq. 3.3)}$$

where $\Delta G_{D\text{-}N}$ is the stability of the protein at different denaturant concentrations and $\Delta G^{0}{}_{D\text{-}N}$ is the stability in water.
Moreover, according to the mass action law, for a two-state system:

$$K_{eq} = \frac{[D]}{[N]} \quad \text{(Eq. 3.4)}$$

and

$$\Delta G_{D-N} = -RT\ln\left(K_{eq}\right) \quad \text{(Eq. 3.5)}$$

where K_{eq} is the equilibrium constant of the reaction, $\Delta G_{D\text{-}N}$ is the stability of the native state to the denatured state, R is the gas constant and T is the temperature. Combining Eq. 3.1 with Eq. 3.4 and Eq. 3.5, the equation that defines the dependence of a given spectroscopic signal from the denaturant concentration can be derived:

$$y_{obs} = \frac{y_N + y_D e^{\left(\frac{(m_{D-N}[denaturant] - \Delta G^{0}{}_{D-N})}{RT}\right)}}{1 + e^{\left(\frac{(m_{D-N}[denaturant] - \Delta G^{0}{}_{D-N})}{RT}\right)}} \quad \text{(Eq. 3.6)}$$

In practice, this equation is complicated by the dependence of the intrinsic spectroscopic signal for both the native and the denatured state on the denaturant concentration (namely the native and denatured baselines (Santoro and Bolen 1988). Eq. 3.6, based on the two-state assumption, is now routinely used for the equilibrium studies of folding. A useful test to verify the two-state assumption is to calculate the thermodynamic folding parameters using different spectroscopic probes.

3.7.2 Quantitative analysis of folding kinetics: the two-state model

Analysis was performed by non-linear least-squares fitting of single

exponential phases using the fitting procedures provided in the Applied Photophysics software. In a two-state reaction, as consequence of a perturbation induced on the system, the observed rate constant k_{obs} is represented by the following equation:

$$k_{obs} = k_F + k_U \qquad \text{(Eq. 3.7)}$$

where k_U and k_F represent the unfolding and refolding rate constants respectively. The logarithm of the observed (un)folding rate constants is linearly dependent on the denaturant concentration. Such a relationship is formalized in Eq. 3.8 (Jackson and Fersht 1991):

$$k_{obs} = k_F^{H_2O} e^{(-m_F[denaturant]/RT)} + k_U^{H_2O} e^{(m_U[denaturant]/RT)} \qquad \text{(Eq. 3.8)}$$

where k_F^{H2O} and k_U^{H2O} are the extrapolated folding and unfolding rate constants in absence of denaturant and, m_F and m_U reflect their dependence on denaturant concentration and correlate with the change in accessible surface area between the two ground states and the transition state between them.
Kinetic analysis of chevron plot allows the determination of the stability of a protein in the absence of denaturant agent ($\Delta G^0_{D\text{-}N}$):

$$\Delta G^0_{D-N} = -RT \ln\left(\frac{k_F}{k_U}\right) \qquad \text{(Eq. 3.9)}$$

where R is the gas constant and T the temperature.
Moreover, the algebric sum of the two kinetic *m*-values, m_U and m_F, allows to calculate the total $m_{D\text{-}N}$:

$$m_{D-N} = m_F + m_U \qquad \text{(Eq. 3.10)}$$

A comparison between the equilibrium and kinetic parameters $\Delta G^0_{D\text{-}N}$ and $m_{D\text{-}N}$ is crucial to verify the validity of the two-state model.

3.7.3 Quantitative analysis of folding kinetics: the three-state model

As reported in the Introduction Section,, if a partially folded intermediate is present, the folding kinetics can be described by the following scheme:

$$D \underset{k_{ID}}{\overset{k_{DI}}{\rightleftarrows}} I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftarrows}} N \qquad \text{(Scheme 3.1)}$$

where k_{DI} is the microscopic rate constant for the formation of the intermediate from the denatured state, k_{ID} is the microscopic rate constant for the unfolding of the intermediate to the denatured state, k_{IN} is the microscopic rate constant for the formation of the native state from the intermediate and k_{NI} is the microscopic rate constant for the unfolding of the native state to the intermediate state.

Curved chevron plots were fitted, using the Kaleidagraph software package, by numerical analysis based on a three-state model following the equation:

$$k_{obs} = k_{DI} + \frac{k_{NI}}{++K_{part}} \qquad \text{Equ. 3.11}$$

where K_{part} is the partition factor k_{ID}/k_{IN} proportional to the difference between the activation barriers for the intermediate state to revert to the reagents rather than proceeding to the products. The logarithm of each microscopic rate constant was assumed to vary linearly with denaturant concentration (the slope of each dependence yielding the corresponding *m*-value).

RESULTS

4.1 Folding characterization of G_A88 and G_B88

To study the folding mechanism of G_A88 and G_B88, we carried out both equilibrium and kinetic experiments. The pathways of folding of G_A88 and G_B88 were extensively characterized at 10°C, under different solvent conditions by exploring a wide range of pH from 2.0 to 10. In the case of G_B88, the experiments were performed also at 25°C, in presence of the stabilizing agent sodium sulfate (0.4 M) as described in paragraph 4.1.2.

4.1.1 Equilibrium unfolding of G_A88 and G_B88

Urea-induced equilibrium denaturations of G_A88 and G_B88, measured at 10°C, pH 7.2 in 50mM sodium phosphate buffer and monitored by far UV-CD spectroscopy, are reported in Figure 4.1.
The observed transitions follow a simple two-state behaviour, suggesting the absence of stable equilibrium intermediates for both proteins and indicating that these designed variants encounter cooperative (un)folding reactions. Furthermore, the reaction was fully reversible at all conditions explored. At physiological pH and in the absence of denaturant the unfolding free energy of G_A88, derived from a two-state analysis and a global fit of 60 wavelengths (from 250 to 220 nm), is $\Delta G_{D\text{-}N}$ = 3.00 ± 0.18 kcal/mol, with a $m_{D\text{-}N}$ value of 0.62 ± 0.04 kcal mol^{-1} M^{-1}. In the case of G_B88, the $\Delta G_{D\text{-}N}$ = 2.35 ± 0.30 kcal/mol, and the $m_{D\text{-}N}$ value is 1.10 ± 0.08 kcal mol^{-1} M^{-1}.

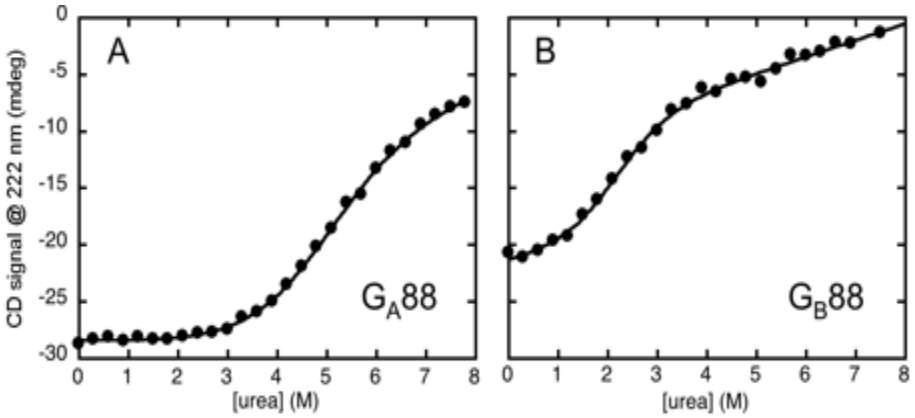


Figure 4.1. Equilibrium denaturations of G_A88 (A) and G_B88 (B) monitored by CD in 50mM sodium phosphate buffer at pH 7.2 and 10°C. Lines are the best fit to a two state unfolding mechanism.

Considering that the G_B88 construct contains circa 10 structured residues more than G_A88, both $m_{D\text{-}N}$ values are consistent with those expected from proteins of this size, according to the BPPred database (Geierhaas, et al. 2007). Hence, the seemingly different cooperativity, as reflected by the different $m_{D\text{-}N}$ values, can be explained by considering the difference in the number of structured residues between the two proteins.

4.1.2 Folding and unfolding kinetics of G_A88 and G_B88

In an effort to unveil the folding and unfolding mechanism of G_A88 and G_B88, we carried out extensive kinetic experiments on both proteins under a variety of different experimental conditions. In particular, the folding and unfolding kinetics were investigated at several pH values, ranging from 10 to 2. In the case of G_A88, it was not possible to measure reliable folding and unfolding rate constants at 25°C over a wide range of denaturant concentration because the time-courses were too fast for our stopped flow apparatus. Thus, kinetic folding data for the two proteins were recorded at 10°C to slow the process for G_A88. In all cases folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final denaturant concentration (representative folding and unfolding time courses of the two proteins are reported in Figure 4.2).

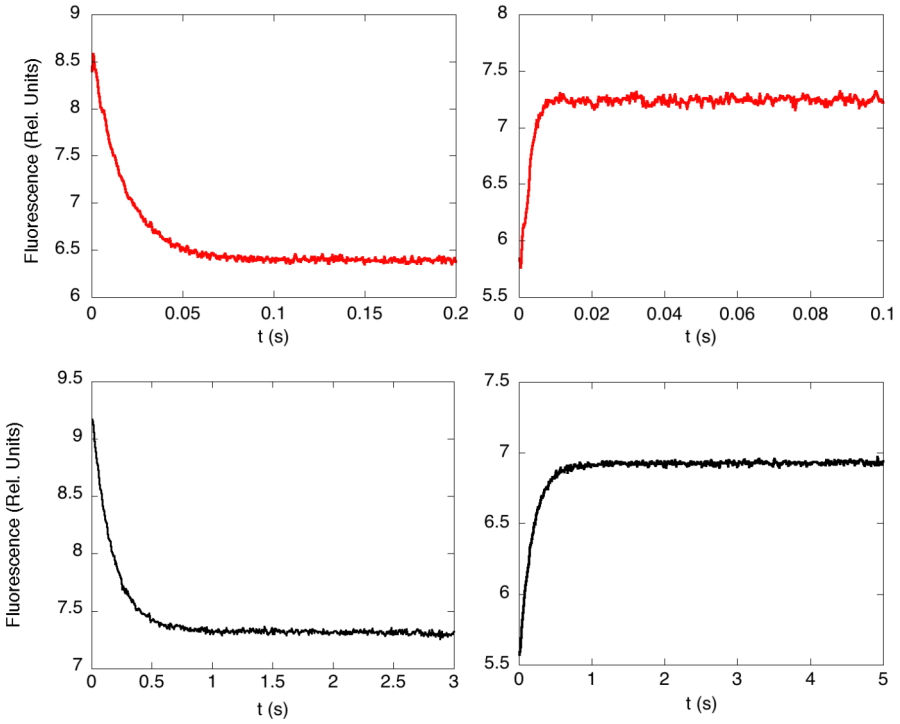


Figure 4.2. Folding kinetics of G_A88 (red) and G_B88 (black) measured at pH 7.2 and 10 °C. Unfolding (left panels) and refolding (right panels) time courses, monitored by the change in Trp fluorescence, were measured at 7.27 and 1.10 M urea, respectively.

A semi-logarithmic plot of the observed folding/unfolding rate constants of G_A88 and G_B88 versus denaturant concentration (chevron plot) recorded at pH 7.2 is reported in Figure 4.3.

Both proteins displayed a V-shaped chevron plot, hallmark of two-state folding (Jackson 1991). In the case of G_A88 there was excellent agreement between the thermodynamic parameters obtained by equilibrium and kinetic data. On the other hand, in the case of G_B88, there was a detectable difference between the $m_{D\text{-}N}$ value obtained from equilibrium experiments being 1.10 ± 0.08 kcal mol^{-1} M^{-1}, and that calculated from chevron plot analysis, 0.90 ± 0.05 kcal mol^{-1} M^{-1}. As observed previously for other small single domain proteins (Mayor, et al. 2003; Religa, et al. 2005; White, et al. 2005), a significant deviation of $m_{D\text{-}N}$ obtained from equilibrium and kinetic

data would suggest the presence of residual structure and/or changes in the exposure of non-polar residues in the denatured state of the protein.

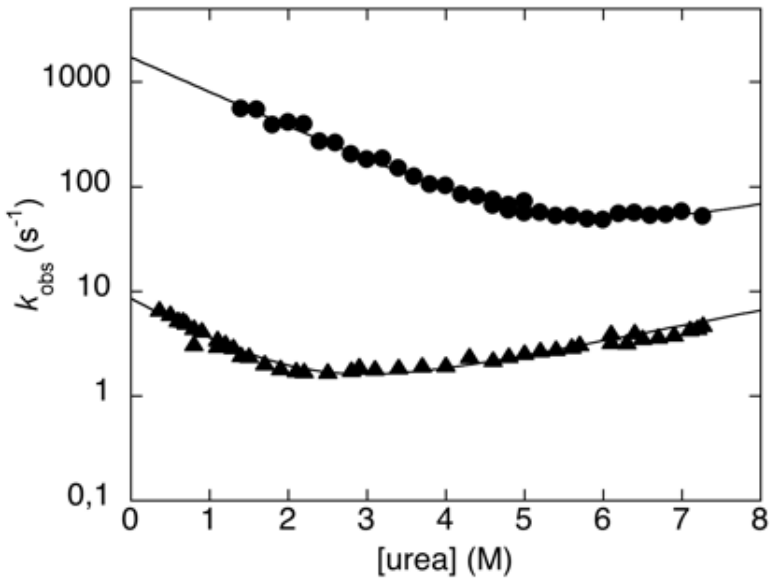


Figure 4.3. Semilogarithmic plot of the observed rate constants for folding and unfolding of G_A88 (●) and G_B88 (▲) versus [urea]. Each constant was measured at 10° C, in 50mM sodium phosphate, at pH 7.2. The (un)folding reactions were monitored by fluorescence emission.

We inferred the presence of residual structure in the denatured state by performing additional experiments at various experimental conditions.
Following Sanchez and Kiefhaber (Sanchez and Kiefhaber 2003), the measurement of chevron plots at different experimental conditions represents a powerful method to address the global properties of folding transition and denatured states. In fact, since the *m*-values (slopes of the unfolding and refolding arms of the chevron plots) reflect the change in accessible surface area upon (un)folding, analysis of their dependence when the experimental conditions are varied, may be of diagnostic value to identify transition state movements along the reaction co-ordinate, as well as denatured state collapse or residual structure. We compared the folding kinetics of G_A88 and G_B88 at various pH values, ranging from 10 to 2.0. Inspection of Figure 4.4 reveals that, while both the stability and *m*-values of G_A88 are essentially insensitive to pH, G_B88 is destabilized at pH < 5.

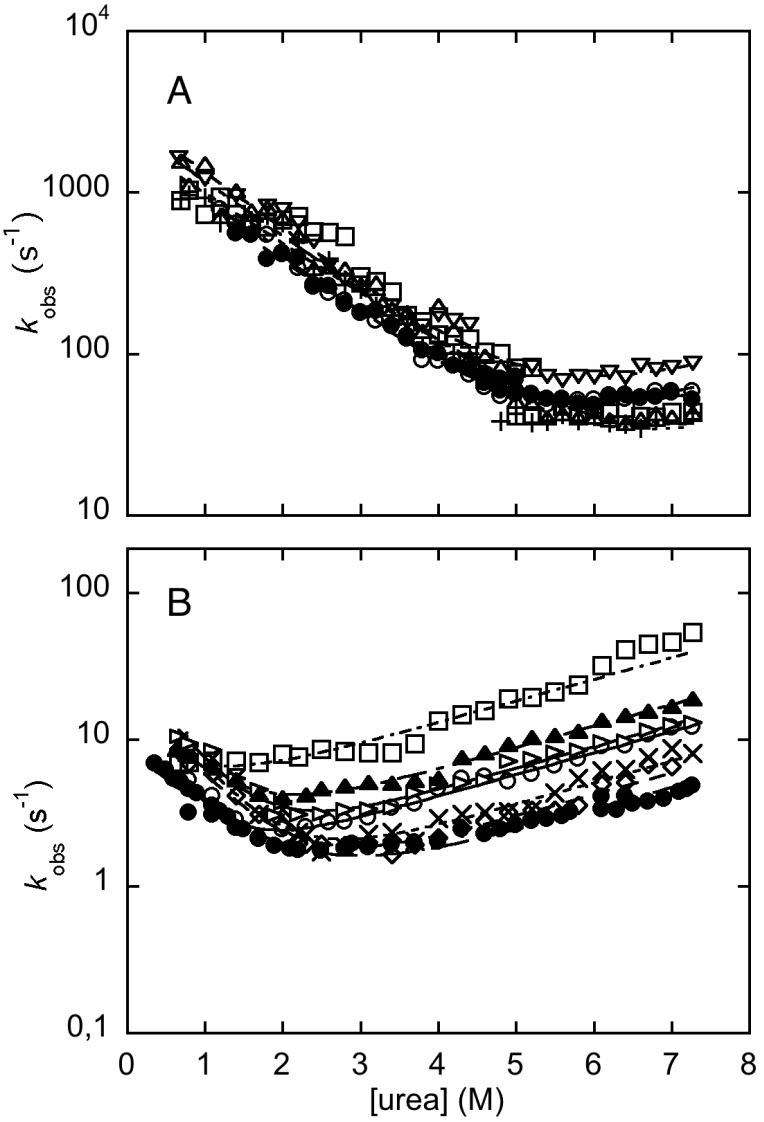


Figure 4.4. Chevron plots of G_A88 and G_B88 at different pH values. A: Semilogarithmic plot of the observed rate constant for folding and unfolding of G_A88 versus [urea] measured at different pH values and 10 °C (▽ pH 10.0, ○ pH 8.0, ● pH 7.2, □ pH 4.0, + pH3.0, △ pH 2.0). B: Semilogarithmic plot of the observed rate constant for folding and unfolding of G_B88 versus [urea] measured at different pH values and 10 °C (○ pH 8.0, ● pH 7.2, ◊ pH 6.2, ✕ pH5.5, ▷ pH 4.7, ▲ pH 4.3, □ pH 4.0). The lines are the best fit to a two-state model.

However, the low stability of the G_B88 protein and the poor definition of the observed refolding arms prevented a quantitative analysis of the observed *m*-values. Therefore, as detailed below, we resorted to further investigate the folding of G_B88 under stabilizing conditions.

Certain inorganic salts, such as phosphates and sulfates, favor compact protein conformations because of preferential exclusion of solvent from the protein surface (Timasheff 1993); this makes them potent stabilizers of both the native and the partially folded states. Chevron plots of G_B88 measured at different pH values and in the presence of 0.4 M sodium sulfate are provided in Figure 4.5 A. As expected, the stabilizing salt allows for a better definition of the refolding arms of the chevron plots. Consequently, we carried out a quantitative analysis of folding parameters over a wide range of pH conditions in the presence of salt. Figure 4.5 B shows the dependence on pH of calculated $m_{D\text{-}N}$, m_F and m_U values for G_B88. The data fit to the protonation of a single titratable group with an apparent $pK_a \sim 5$. Interestingly, as reported in Table 4.1, the $m_{D\text{-}N}$ decreases with decreasing pH values, suggesting that the denatured state of this small single domain protein becomes more compact at acidic conditions. Importantly, however, even at pH 7.2, the total $m_{D\text{-}N}$ value, calculated as algebric sum of folding and unfolding *m* values, is lower than that calculated from equilibrium experiments (i.e. 0.93 ± 0.05 kcal mol^{-1} M^{-1}), suggesting the presence of residual structure in the denatured state under more "physiological" conditions.

Additional details regarding the structural features of the denatured state of G_B88 were obtained by performing molecular dynamics (MD) simulations conducted in collaboration with Valerie Daggett. In Figure 4.6 are presented the snapshots from three independent thermal unfolding simulations performed for G_A88 and G_B88 at 498 K at neutral pH. Interestingly, in the case of G_B88, the calculated denatured state displays detectable residual structure in the α-helix.

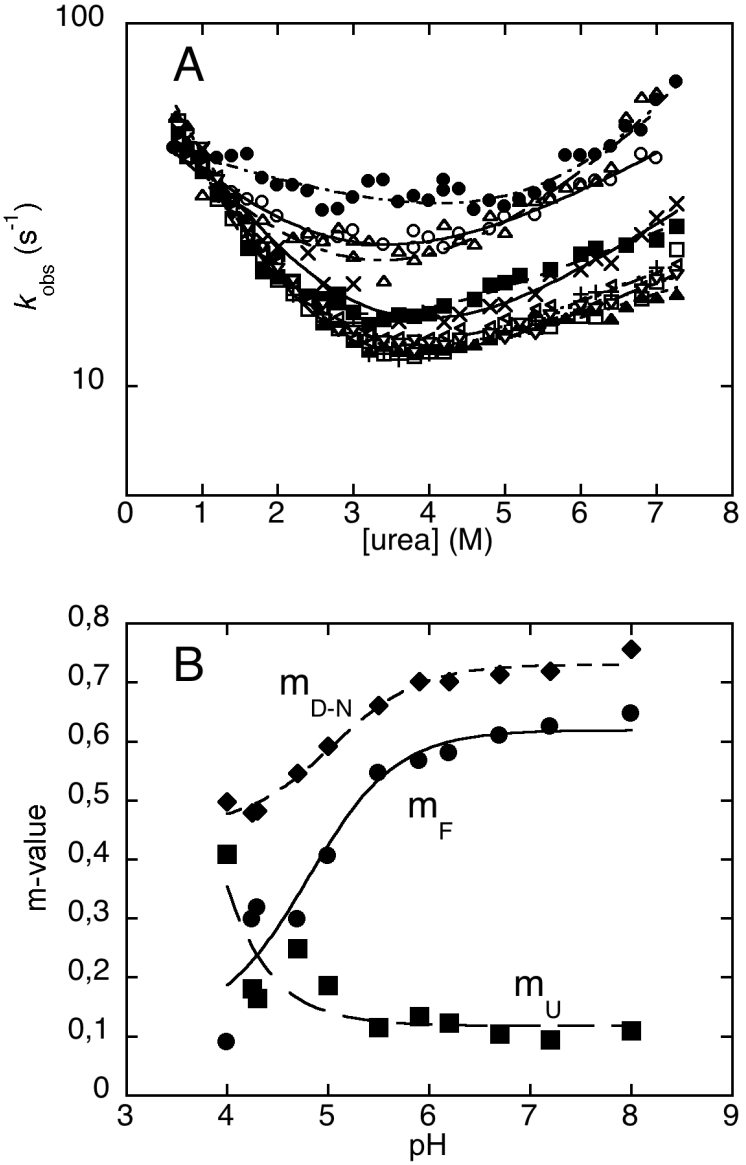


Figure 4.5. A: Semilogarithmic plot of the observed rate constants for folding and unfolding of G_B88 versus [urea] measured at different pH values, in presence of 0.4 M sodium sulfate and at 25 °C (■ pH 8.0, ▲ pH 7.2, ▽ pH 6.7, □ pH 6.2, + pH 5.9, ◁ pH 5.5, ✕ pH 5.0, △ pH 4.7, ○ pH 4.3, ● pH 4.0). The lines are the best fit to a two-state model. B: Dependence of m_F (●), m_U (■) and m_{D-N} (◆) on pH. Lines are the best fit to an equation implying a single protonation site.

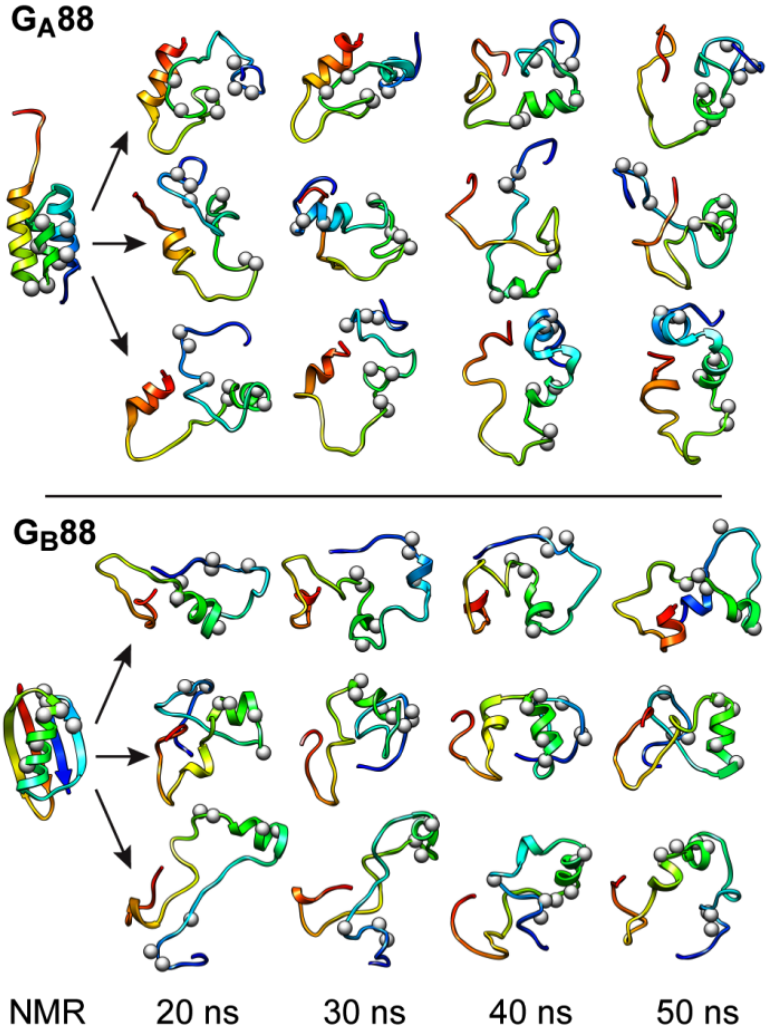


Figure 4.6. Three independent thermal unfolding simulations of G_A88 (above) and G_B88 (below) performed at 498 K and neutral pH. The proteins are colored red → blue from the N → C. The Cα atoms of the 7 different residues between G_A88 and G_B88 (positions 24, 25, 30, 33, 45, 49, 50) are shown as gray balls. The NMR structures of the native states are shown on the left, and the calculated structures at 20, 30, 40, and 50 nanoseconds for each of the three long 498 K simulations (1 - top, 2 - middle, 3 - bottom) are depicted.

pH	k_f (s^{-1})	k_u (s^{-1})	m_f (kcal $mol^{-1}M^{-1}$)	m_u (kcal $mol^{-1}M^{-1}$)	ΔG (kcal mol^{-1})	$m_{D\text{-}N}$ (kcal $mol^{-1}M^{-1}$)
8.0	88± 12	7.3± 1.2	0.65± 0.08	0.11± 0.02	1.5± 0.1	0.76± 0.08
7.2	105±12	5.9± 1.1	0.62± 0.06	0.09± 0.01	1.7± 0.1	0.72± 0.06
6.7	107± 20	5.5± 1.2	0.60± 0.08)	0.10± 0.02	1.7± 0.2	0.71± 0.08
6.2	89± 10	4.7± 1.1	0.60± 0.07	0.12± 0.02	1.7± 0.1	0.70± 0.06
5.9	94± 12	4.2± 1.0	0.57± 0.06	0.13± 0.02	1.8± 0.2	0.70± 0.07
5.5	84± 9	5.2± 1.2	0.55± 0.05	0.11± 0.02	1.6± 0.1	0.66± 0.06
5.0	75± 8	3.0± 1.1	0.41± 0.06	0.19± 0.03	1.9± 0.2	0.59± 0.05
4.7	56± 6	2.9± 1.2	0.30± 0.08	0.25± 0.04	1.7± 0.1	0.55± 0.08
4.3	55± 8	6.1± 2.6	0.32± 0.08	0.16± 0.03	1.3± 0.1	0.48± 0.06
4.2	56± 7	6.3± 3.6	0.30± 0.10	0.18± 0.05	1.3± 0.1	0.48± 0.10
4.0	51± 4	5.8± 0.5	0.21± 0.03	0.18± 0.05	1.3± 0.1	0.39± 0.03

Table 4.1. Folding parameters of G_B88 as a function of pH. Chevron plots were fitted to a two-state model. Kinetic $m_{D\text{-}N}$ represents the sum of two kinetics m values for folding and unfolding reactions.

4.2 Folding characterization of the different G_A and G_B variants

The results described above, suggest the folding pathway of G_A88 and G_B88 to diverge as early as in their denatured state. This surprising result promoted us to perform a systematic analysis of each of the heteromorphic variants designed by Bryan and co-workers (Alexander, et al. 2007; He, et al. 2008), considering also the natural G_B domain, known as GB1.

4.2.1 Folding characterization of GB1

To study the folding mechanism of GB1, we carried out both equilibrium and kinetic experiments. GB1 folding pathway was extensively characterized at 25 °C, under different pH conditions, exploring a wide range from 2.0 to 9.6.

4.2.1.1 Equilibrium denaturations.

Guanidine-induced equilibrium denaturations of GB1, monitored by fluorescence spectroscopy, were obtained at 25°C exploring a wide range of pH, from 2.0 to 9.6 (Figure 4.7). Equilibrium denaturations were fitted both individually and globally with shared $m_{D\text{-}N}$ value. Values obtained from the global analysis are listed in Table 4.2. They were consistent within error with the values obtained by fitting individually each independent equilibrium experiments, as well as with the value calculated from kinetic experiments, confirming the two-state nature of the equilibrium unfolding transition of GB1. The denaturation profiles were all consistent with a two-state unfolding and returned an *m*-value of 1.75 ± 0.2 kcal mol^{-1} M^{-1}.

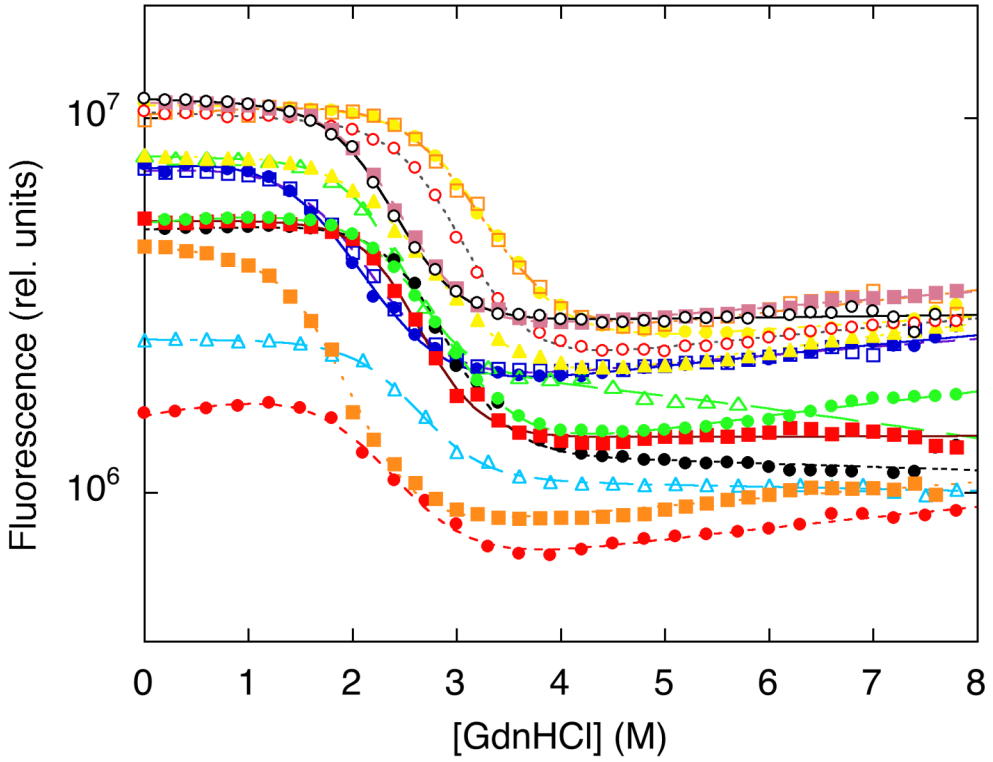


Figure 4.7. Equilibrium denaturations of GB1 monitored by fluorescence in a wide range of pH, from 2.0 to 9.6, at 25°C (●, pH 2.0; △, pH 2.5; □, pH 3.0; ●, pH 3.5; ▲, pH 4.0; △, pH 4.5; ●, pH 5.0; □, pH 6.0; ●, pH 6.5; ○, pH 7.0; ■, pH 7.5; ●, pH 8.0; ■, pH 8.5; ○, pH 9.0; ■, pH 9.6). Lines represents the best fit to a two state unfolding transition.

4.2.1.2 Kinetic experiments

The folding and unfolding kinetics of GB1 were investigated at several pH values, ranging from 2.0 to 9.6. In all cases, folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final denaturant concentration. Each rate constant was obtained from the average of at least five independent shots in stopped-flow experiments. Semi-logarithmic plots of the observed folding/unfolding rate constants of GB1 versus guanidine concentration (chevron plots) at the different pH values are presented in Figure 4.8.

Surprisingly, the unfolding arm of the chevron plots at pH values higher than 6.0 shows a deviation from linearity that becomes evident at high guanidine concentrations (roll over effect). This effect escaped previous studies probably because of the restricted range of experimental conditions, limited to [GdnHCl] < 5.5 M (Park, et al. 1997; McCallister, et al. 2000).

Indeed, if we were to ignore the data we recorded for [GdnHCl] > 5.5 M, the unfolding arm of the chevron plots would appear essentially linear but would display a puzzling change in slope in the different experimental conditions.

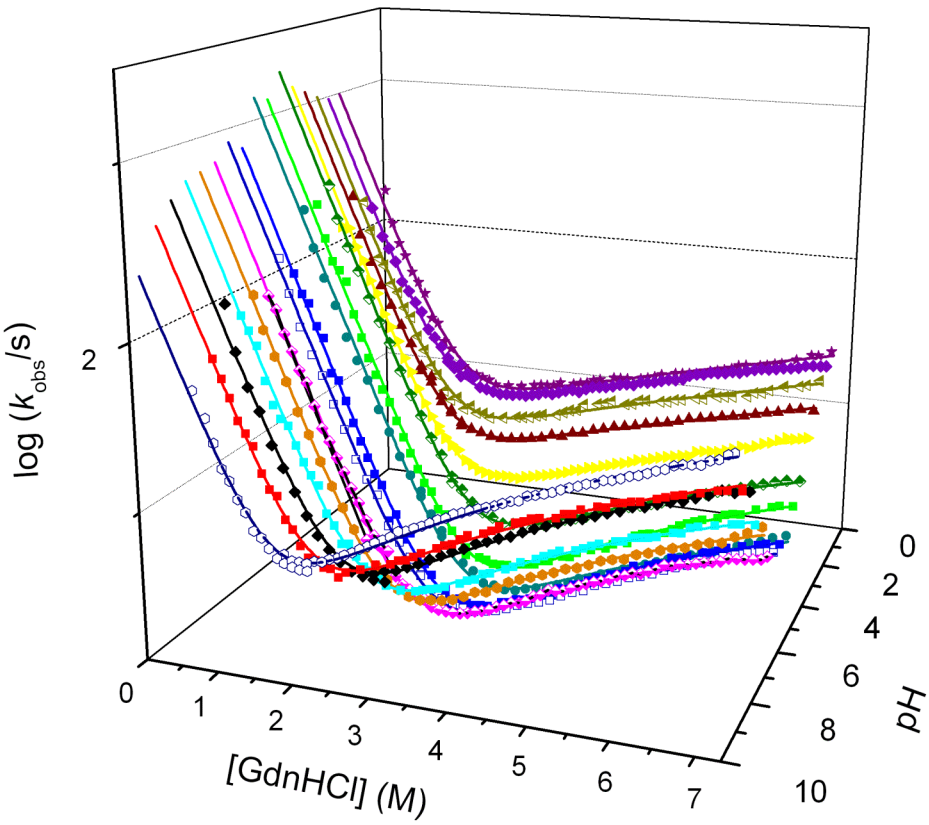


Figure 4.8. Chevron plots of GB1 measured from pH 2.0 to pH 9.6. Three-dimensional graph was obtained with Origin software. Lines are the best global fit to a three-state equation with shared m-values (Gianni, et al. 2007b). Exclusion of data [GdnHCl] > 5.5 M would result in a quasi-linear unfolding arm with an apparent change in unfolding m-values.

The deviation from linearity of the chevron plots of GB1is highlighted in Figure 4.9 where the chevron plot obtained at pH 9.0 is reported together with the residuals of the fit, showing a clear systematic deviation from the expected values for a two-state behavior.
As reported in the Introduction Section, a deviation from linearity in either the folding or the unfolding branches may be considered of diagnostic value for the identification of intermediates (Parker, et al. 1995; Wildegger and Kiefhaber 1997). If a partially folded intermediate is present, the folding

kinetics can be described by a three-state mechanism and the observed chevron plots can be fitted by numerical analysis based on a three-state model following the Equation 3.11 reported in Methods Section.

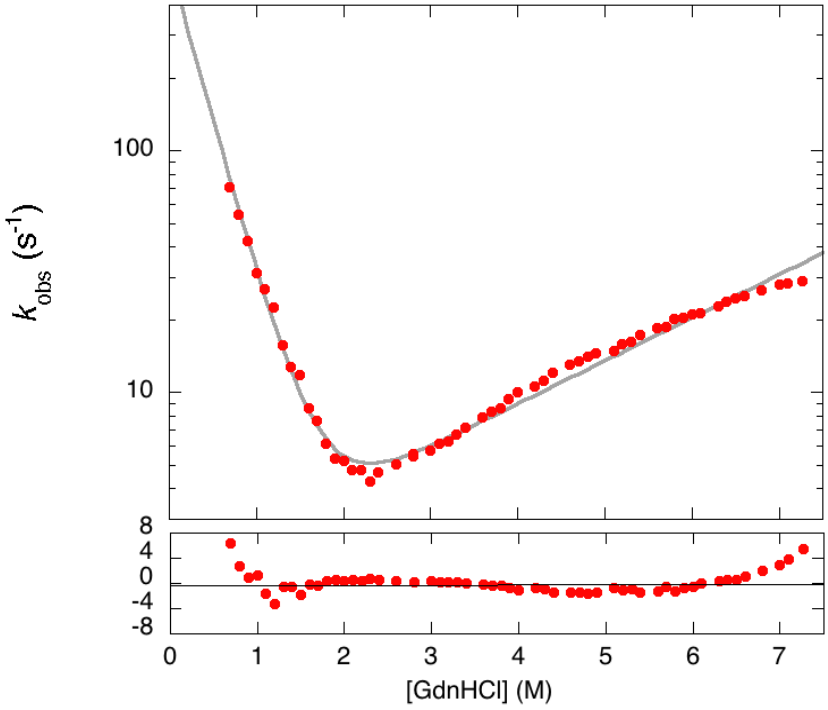


Figure 4.9. Chevron plot of wild-type GB1 measured at pH 9.0. The grey line is the best fit to a two-state equation. The residuals of the fit, showing a clear systematic deviation from the expected values for a two-state behavior, are reported below the graph. Inclusion of data at [GdnHCl] > 5.5 M are critical to detect the roll-over effect.

Parameters calculated from global analysis (listed in Table 4.2) allow the identification of the relative positions of the two activation barriers along the reaction coordinate in terms of their relative accessible surface area (Tanford β-value), resulting in a β_T-value of 0.76 ± 0.04 for the transition state *TS1* and 0.93 ± 0.04 for the native-like activation barrier *TS2*.

The excellent statistical parameters of the global analysis indicate that the two activation barriers are robust to changes in pH conditions and display a

conserved solvent accessible surface area when pH and protein stability are altered.

pH	k_{DI} (s^{-1})	$k_{NI}/(1+K_{part})$ (s^{-1})	K_{part}	$\Delta G_{D\text{-}N}$ [a] (kcal mol^{-1})	$\Delta G_{D\text{-}N}$ [b] (kcal mol^{-1})
2.0	1000 ± 90	11.0 ± 3.0	1.7 ± 0.2	2.7 ± 0.3	3.9 ± 0.8
2.5	1000 ± 90	19.2 ± 4.8	3.0 ± 0.4	2.3 ± 0.2	3.0 ± 0.7
3.0	1050 ± 95	5.2 ±1.3	0.9 ±0.1	3.1 ± 0.4	3.8 ± 0.4
3.5	1200 ± 120	5.1 ±1.3	1.2 ±0.1	3.2 ± 0.2	3.9 ± 0.4
4.0	1400 ± 120	1.3 ± 0.3	0.42 ± 0.09	4.1± 0.2	4.5 ± 0.4
4.5	1800 ± 170	0.4 ± 0.1	0.20 ± 0.04	5.0 ± 0.2	4.2 ± 0.4
5.0	1300 ± 120	0.15 ± 0.04	0.09 ± 0.02	5.3 ± 0.2	4.9 ± 0.5
5.5	1370 ± 100	0.15 ± 0.04	0.13 ± 0.03	5.4 ± 0.1	4.6 ± 0.5
6.0	720 ± 70	0.12 ± 0.03	0.09 ± 0.02	5.1 ± 0.1	5.2 ± 0.5
6.5	830 ± 70	0.14 ± 0.04	0.10 ± 0.02	5.1 ± 0.2	4.7 ± 0.5
7.0	670 ± 70	0.19 ± 0.05	0.12 ± 0.02	4.8 ± 0.2	5.3 ± 0.5
7.5	630 ± 70	0.27 ± 0.07	0.11 ± 0.02	4.6 ± 0.2	4.3 ± 0.4
8.0	600 ± 60	0.4 ± 0.1	0.12 ± 0.03	4.3 ± 0.3	4.3 ± 0.4
8.5	500 ± 40	0.6 ± 0.2	0.11 ± 0.02	3.9 ± 0.2	3.8 ± 0.4
9.0	390 ± 40	0.9 ± 0.2	0.13 ± 0.03	3.6 ± 0.2	3.3 ± 0.3
9.6	220 ± 25	1.3 ± 0.3	0.11 ± 0.02	3.0 ± 0.2	3.0 ± 0.3

Table 4.2. Folding parameters of GB1 as a function of pH. [a]Calculated from chevron plot analysis. The chevron plots were fitted globally to a three-state model with shared m values. k_{DI} is the microscopic rate constant for the formation of the intermediate from the denatured state; k_{NI} is the microscopic rate constant for the unfolding of the native state to the intermediate state; K_{part} is the partitioning factor k_{ID}/k_{IN} reflecting the difference between the activation barriers for the intermediate to revert to the reagents rather than proceeding to the products. The analysis returned a total $m_{D\text{-}N}$ = 1.95 ± 0.2 kcal mol^{-1} M^{-1}. The Tanford β values for the two transition states were β_{TS1} = 0.76 ± 0.04 and β_{TS2} = 0.93 ± 0.04. [b]Calculated from equilibrium experiments. Equilbrium denaturation curves were fitted both individually and globally with shared $m_{D\text{-}N}$ value. The global analysis returned $m_{D\text{-}N}$ = 1.75 ± 0.2 kcal mol^{-1} M^{-1}. This value was consistent within error with the values obtained by fitting individually each independent equilibrium experiments, as well as with the value calculated from kinetic experiments.

4.2.1.3 The effect of pH on the folding kinetics of GB1

The analysis of the chevron plots reported in Figure 4.8 allowed measuring the folding and unfolding rate constants of GB1 over a very wide range of pH. The folding rate constants display a negligible dependence on pH (data not shown). A plot of the logarithm of apparent unfolding rate constants from the native state to *TS2* (k_{NI}) and *TS1* (formally equivalent to $k_{NI}/(1+K_{part})$) as a function of pH are reported in Figure 4.10.

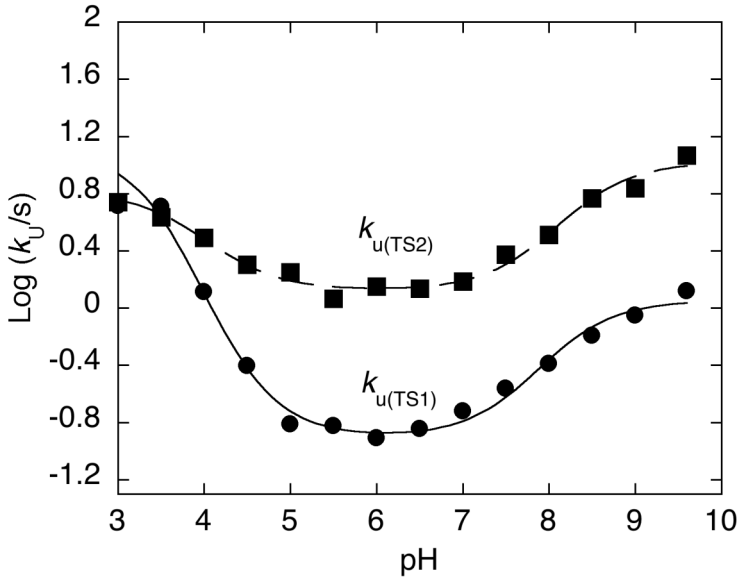


Figure 4.10 Unfolding rate constants versus pH. Logarithm of calculated unfolding rate constants from the native state N to the first transition state k_u(TS1) and from the native state N to the second transition state k_u(TS2) as a function of pH. The lines are the best fit to a model involving the protonation of two groups. In both cases, we obtained the same approximate pK_a of ~ 4 and ~ 8.

Interestingly, both *TS1* and *TS2* display sigmoidal transitions at acidic and alkaline pH consistent with protonation of at least two groups in the native state with apparent pK_a of ~ 4 and ~ 8. Importantly, the acid transition for the unfolding rate constant of *TS1* (changing by almost two orders of magnitude) is more pronounced than that for the unfolding rate constant of *TS2* (changing by less than an order of magnitude), suggesting the contribution of a salt bridge that is weak or not formed in *TS1*, but is consolidated in *TS2*.

4.2.2 Φ-value analysis of the G_A and G_B variants

In order to study the folding pathway of each of the heteromorphic pairs designed in the laboratory of Philip N. Bryan, i.e. G_A30 and G_B30, G_A77 and G_B77, G_A88 and G_B88, an extensive Φ-value analysis was performed.

The Φ-value analysis represents a powerful strategy to obtain structural information about intermediate and transition states. Indeed, carrying out a systematic comparison of the folding and unfolding rate constants for a series of mutational variants, it is possible to map out the interaction patterns in the intermediate and transition states (Fersht, et al. 1992). We introduced the recommended non-disruptive mutations into the wild-type proteins and investigated their effects on the kinetic and thermodynamic properties. 128 mutants were constructed; unfortunately 18 of them expressed poorly and could not be included in the analysis. The folding and unfolding kinetics of the remaining mutants were investigated in 50mM sodium phosphate buffer, at pH 7.2. Kinetic and thermodynamic parameters obtained by these experiments were used to calculate Φ-values. Following Fersht and Sato, the mutants with a $\Delta\Delta G_{D-N}$ too low (<0.35 kcal M^{-1}) to calculate reliable Φ-values, were excluded from the analysis (Fersht and Sato, 2004).

The chevron plots obtained for G_A and G_B mutants are shown in Figures 4.11 and 4.12 respectively. Interestingly, whilst the observed chevron plots of G_A proteins displayed a V-shaped behavior, hallmark of two-state folding (Jackson 1991), the folding pathway of the G_B variants, in analogy to GB1, appeared more complex, consistent with the presence of an intermediate. Parameters calculated from global analysis of G_A30, G_A77 and G_A88 are listed in Tables 4.3, 4.4 and 4.5 respectively; while these obtained for G_B30, G_B77 and G_B88 are listed in tables 4.6, 4.7 and 4.8, respectively. The calculated Φ-values were divided into three categories and reported on the structures of G_A and G_B (Figures 4.13 and 4.14, respectively) using the following color code: red, $0 < \Phi < 0.30$; magenta, $0.30 < \Phi < 0.70$; blue, $0.70 < \Phi < 1$. A comprehensive description of the folding pathways for each different protein, as obtained by Φ-value analysis, is reported in the Discussion.

Since G_A and G_B proteins show different stabilities and their rate constants span different orders of magnitude, it was necessary to use two different temperatures (10°C and 25°C) and two different denaturant agents (guanidine and urea) to characterize all the proteins. Moreover, the rate constants of some G_A mutants, which were too fast for stopped-flow apparatus, were measured by employing a T-Jump instrument (in collaboration with the

laboratory of Prof. Per Jemth at the University of Uppsala, where I spent a period of one month during the Ph.D. work). In order to compare the Φ-values calculated at different conditions, kinetic experiments of some mutants were performed at 10°C and 25°C with both guanidine and urea. The Φ-values obtained at different conditions were approximately the same (data not shown). The folding mechanism of G_A and G_B proteins is therefore not affected by the nature of the denaturant agent used to obtain the data.

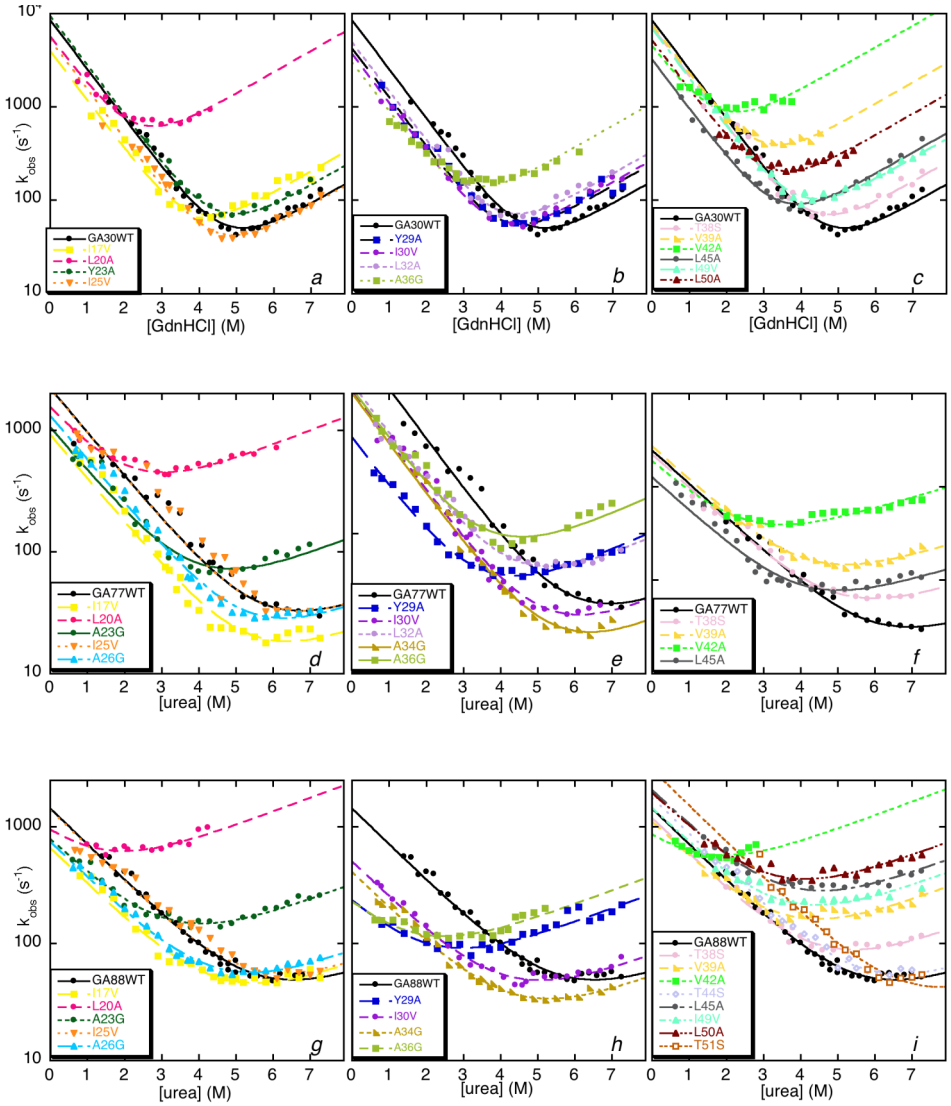


Figure 4.11. Folding kinetics of wild-type (WT) G_A proteins (black filled circles) and their respective mutants. Observed rate constants were measured as a function of denaturant concentration at pH 7.2. a,b,c) Chevron plots of G_A30 mutants measured at 25°C. d,e,f) Chevron plots of G_A77 mutants measured at 10°C. f,h,i) Chevron plots of G_A88 mutants measured at 10°C. The mutants G_A30_L20A (panel a), G_A30_V39A and G_A30_V42A (panel c), G_A88_L20A (panel g) and G_A88_V42A (panel i) were measured by employing the Temperature-jump methodology (in collaboration with Prof. Per Jemth). All chevron plots, displaying a V-shaped behavior, were fitted by numerical analysis based on a two-state model. Lines are the best fit to this model.

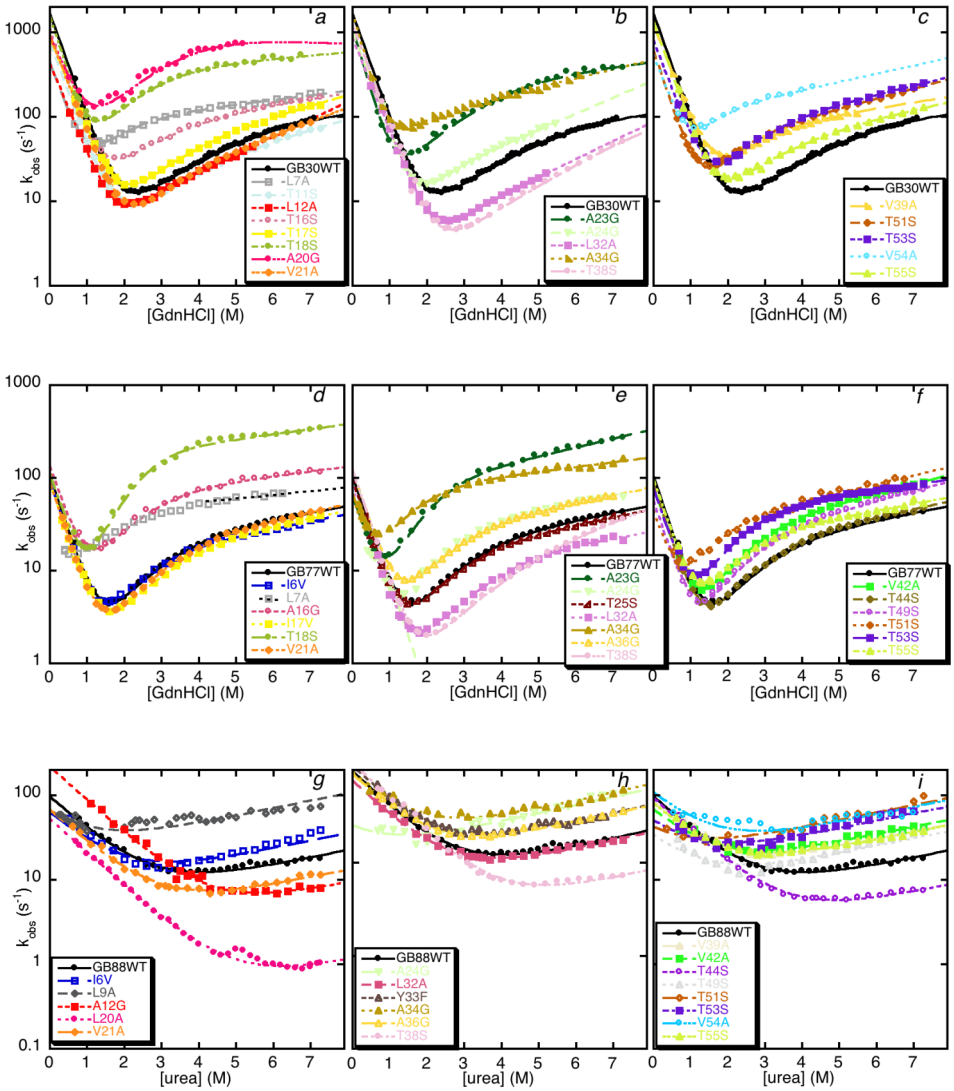


Figure 4.12. Folding kinetics of wild-type (WT) G_B proteins (black filled circles) and their respective mutants. Observed rate constants were measured as a function of denaturant concentration at pH 7.2 and 25°C. a,b,c) Chevron plots of G_B30 mutants. d,e,f) Chevron plots of G_B77 mutants. g,h,i) Chevron plots of G_B88 mutants. All the chevron plots of G_B30 and G_B77 WT and mutants, showing a roll-over effect hallmark of an on-pathway intermediate, were fitted to a three-state equation; whilst the chevron plots of G_B88 WT and mutants, displaying a V-shaped behavior, were fitted by numerical analysis based on a two-state model. Lines are the best fit to these models.

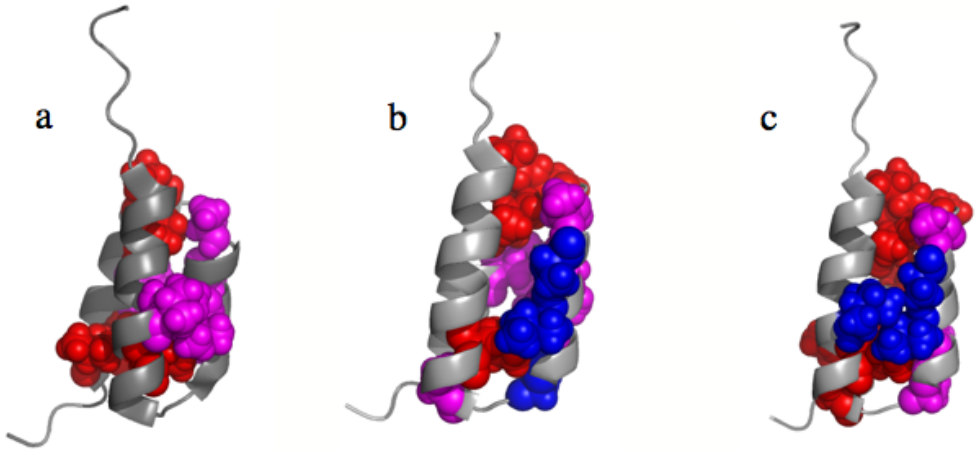


Figure 4.13. Structural distribution of the measured Φ-values on the native structures of G_A30 (a), G_A77 (b) and G_A88 (c). The experimentally determined Φ values were divided into three categories and reported on the structure of G_A variants using the following color code: red, $0 < \Phi < 0.30$; magenta, $0.30 < \Phi < 0.70$; blue, $0.70 < \Phi < 1$. A conserved nucleus between the helices α 1and α 2 is clearly evident in all the G_A variants.

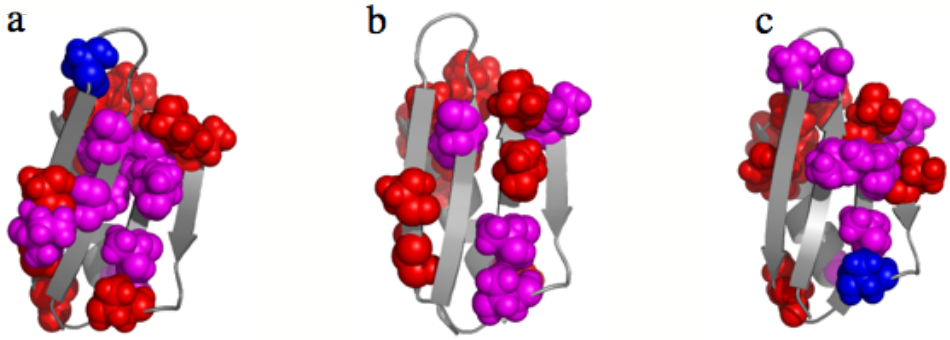


Figure 4.14. Structural distribution of the Φ-values, measured for the first transition state, on the native structures of G_B30 (a), G_B77 (b) and G_B88 (c). The experimentally determined Φ values were divided into three categories and reported on the structure of G_B variants using the following color code: red, $0 < \Phi < 0.30$; magenta, $0.30 < \Phi < 0.70$; blue, $0.70 < \Phi < 1$. As detailed in the Discussion, by considering G_B30 and G_B88, it is possible to observe, a shift of the medium-high Φ-values from the first β-hairpin to the second, with G_B77 displaying an intermediate behavior.

	k_F (s^{-1})	k_U (s^{-1})	ΔG_{D-N} (kcal mol^{-1})	$\Delta\Delta G_{D-TS}$ (kcal mol^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})
WT	7800 ± 250	2.4 ± 0.1	4.5 ± 0.02		
I17V	4400 ± 130	5.1 ± 0.2	3.8 ± 0.02	0.32 ± 0.05	0.75 ± 0.01
L20A	5900 ± 100	98 ± 2	2.3 ± 0.01	0.16 ± 0.03	2.2 ± 0.01
Y23A	8500 ± 350	3.7 ± 0.2	4.3 ± 0.02	0.05 ± 0.04	0.19 ± 0.01
I25V	5900 ± 300	2.4 ± 0.2	4.4 ± 0.03	0.15 ± 0.05	0.17 ± 0.01
Y29A	4480 ± 100	3.4 ± 0.1	4.0 ± 0.01	0.31 ± 0.04	0.51 ± 0.01
I30V	3800 ± 100	4.2 ± 0.1	3.8 ± 0.01	0.41 ± 0.05	0.72 ± 0.01
L32A	5300 ± 200	4.7 ± 0.2	3.9 ± 0.02	0.22 ± 0.04	0.60 ± 0.01
A36G	3300 ± 100	16 ± 0.4	3.0 ± 0.01	0.48 ± 0.05	1.5 ± 0.01
T38S	8200 ± 250	3.8 ± 0.2	4.3 ± 0.02	0.03 ± 0.03	0.24 ± 0.01
V39A	7700 ± 100	45 ± 1	2.9 ± 0.01	0.01 ± 0.02	1.7 ± 0.01
V42A	3600 ± 100	210 ± 10	1.6 ± 0.01	0.43 ± 0.05	2.9 ± 0.02
L45A	3420 ± 100	7.6 ± 0.2	3.4 ± 0.01	0.46 ± 0.05	1.1 ± 0.02
I49V	6700 ± 120	7.0 ± 0.2	3.8 ± 0.01	0.08 ± 0.03	0.69 ± 0.02
L50A	5100 ± 110	21 ± 1	3.1 ± 0.01	0.24 ± 0.04	1.5 ± 0.01

A

	Φ
WT	
I17V	**0.43** ± 0.06
L20A	**0.07** ± 0.01
Y23A	*
I25V	*
Y29A	**0.61** ± 0.08
I30V	**0.56** ± 0.06
L32A	**0.36** ± 0.07
A36G	**0.31** ± 0.03
T38S	*
V39A	**0.00** ± 0.01
V42A	**0.15** ± 0.02
L45A	**0.42** ± 0.04
I49V	**0.12** ± 0.04
L50A	**0.17** ± 0.03

B

Table 4.3. A. Kinetic and thermodynamic parameters of wild-type (WT) and mutant G_A30 proteins calculated from chevron plot analysis. The chevron plots were fitted globally to a two-state model. The Tanford β-value for the transition state was $\beta_{TS} = 0.70 \pm 0.04$.

*B. Experimentally determined **Φ**-values for the G_A30 mutants.*

** Mutants with $\Delta\Delta G_{D-N}$ too low (<0.35 kcal mol^{-1}) to calculate a reliable **Φ** value.*

	k_F (s^{-1})	k_U (s^{-1})	ΔG_{D-N} (kcal mol^{-1})	$\Delta\Delta G_{D-TS}$ (kcal mol^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})
WT	2460 ± 50	3.8 ± 0.2	3.6 ± 0.01		
I17V	990 ± 20	2.4 ± 0.1	3.4 ± 0.01	0.51 ± 0.04	0.26 ± 0.01
L20A	1540 ± 30	136 ± 10	1.4 ± 0.01	0.26 ± 0.03	2.3 ± 0.02
A23G	1100 ± 60	14 ± 1	2.4 ± 0.01	0.45 ± 0.03	1.2 ± 0.01
I25V	2870 ± 100	3.2 ± 0.3	3.8 ± 0.02	0.09 ± 0.02	0.17 ± 0.01
A26G	1320 ± 70	4.0 ± 0.1	3.2 ± 0.01	0.35 ± 0.03	0.38 ± 0.01
Y29A	480 ± 30	12 ± 1	2.1 ± 0.00	0.91 ± 0.06	1.5 ± 0.01
I30V	1120 ± 50	3.8 ± 0.2	3.2 ± 0.01	0.44 ± 0.03	0.45 ± 0.01
L32A	1200 ± 50	11 ± 0.3	2.7 ± 0.01	0.40 ± 0.03	0.98 ± 0.01
A34G	990 ± 50	2.9 ± 0.1	3.3 ± 0.01	0.51 ± 0.03	0.36 ± 0.01
A36G	1100 ± 50	21 ± 0.6	2.2 ± 0.01	0.46 ± 0.03	1.4 ± 0.01
T38S	2300 ± 80	10 ± 0.3	3.1 ± 0.01	0.04 ± 0.02	0.56 ± 0.02
V39A	2800 ± 100	27 ± 2	2.6 ± 0.01	0.07 ± 0.01	1.0 ± 0.04
V42A	1600 ± 40	119 ± 11	1.5 ± 0.01	0.24 ± 0.02	2.2 ± 0.01
T44S	3000 ± 50	3.1 ± 0.1	3.9 ± 0.01	0.12 ± 0.02	0.22 ± 0.01
L45A	1370 ± 50	15 ± 0.6	2.5 ± 0.01	0.33 ± 0.03	1.1 ± 0.01

A

	Φ
WT	
I17V	*
L20A	**0.12** ± 0.01
A23G	**0.37** ± 0.03
I25V	*
A26G	**0.91** ± 0.07
Y29A	**0.59** ± 0.04
I30V	**0.98** ± 0.07
L32A	**0.41** 0.03
A34G	**1.40** ± 0.09
A36G	**0.32** ± 0.02
T38S	**0.07** ± 0.03
V39A	**0.07** ± 0.01
V42A	**0.11** ± 0.01
T44S	*
L45A	**0.30** ± 0.03

B

Table 4.4. Kinetic and thermodynamic parameters of wild-type (WT) and mutant G_A77 proteins calculated from chevron plot analysis. The chevron plots were fitted globally to a two-state model The Tanford β value for the transition state was β_{TS} = 0.75 ± 0.04.

*B. Experimentally determined **Φ**-values for the G_A77 mutants.*

** Mutants with $\Delta\Delta G_{D-N}$ too low (<0.35 kcal mol^{-1}) to calculate a reliable **Φ** value.*

.

	k_F (s^{-1})	k_U (s^{-1})	ΔG_{D-N} (kcal mol^{-1})	$\Delta\Delta G_{D-TS}$ (kcal mol^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})
WT	1500 ± 100	6.4 ± 0.8	3.0 ± 0.02		
A12G	5000 ± 780	1000 ± 90	0.9 ± 0.02	0.69 ± 0.01	2.1 ± 0.02
I17V	700 ± 30	7.7 ± 0.8	2.5 ± 0.02	0.41 ± 0.07	0.52 ± 0.01
L20A	670 ± 110	280 ± 20	0.5 ± 0.00	0.44 ± 0.00	2.5 ± 0.02
A23G	770 ± 40	37 ± 3	1.7 ± 0.03	0.36 ± 0.07	1.3 ± 0.02
I25V	1500 ± 30	7.6 ± 0.9	3.0 ± 0.01	0.01 ± 0.04	0.09 ± 0.00
A26G	760 ± 30	10 ± 1.0	2.4 ± 0.02	0.37 ± 0.07	0.62 ± 0.01
Y29A	210 ± 20	32 ± 3	1.1 ± 0.05	1.09 ± 0.22	2.0 ± 0.09
I30V	500 ± 20	9.4 ± 0.9	2.2 ± 0.02	0.60 ± 0.10	0.81 ± 0.01
A34G	410 ± 3	6.1 ± 0.1	2.4 ± 0.00	0.71 ± 0.09	0.68 ± 0.01
A36G	200 ± 4	42 ± 1	0.9 ± 0.01	1.13 ± 0.20	2.2 ± 0.03
T38S	1200 ± 20	16 ± 0.3	2.4 ± 0.01	0.11 ± 0.04	0.61 ± 0.01
V39A	950 ± 20	39 ± 1	1.8 ± 0.01	0.24 ± 0.05	1.2 ± 0.01
V42A	610 ± 10	260 ± 25	0.5 ± 0.01	0.49 ± 0.07	2.6 ± 0.06
T44S	1770 ± 60	7.1 ± 0.5	3.1 ± 0.01	0.11 ± 0.03	0.05 ± 0.01
L45A	1930 ± 80	63 ± 7	1.9 ± 0.04	0.15 ± 0.02	1.9 ± 0.02
I49A	1210 ± 40	51 ± 5	1.8 ± 0.01	0.11 ± 0.04	1.3 ± 0.01
L50A	1970 ± 30	89 ± 8	1.7 ± 0.01	0.17 ± 0.03	1.3 ± 0.01
T51S	3050 ± 50	4.3 ± 0.1	3.7 ± 0.01	0.41 ± 0.02	0.63 ± 0.01

A

	Φ
WT	
A12G	**0.32** ± 0.00
I17V	**0.80** ± 0.14
L20A	**0.17** ± 0.00
A23G	**0.27** ± 0.06
I25V	*
A26G	**0.60** ± 0.11
Y29A	**0.55** ± 0.11
I30V	**0.74** ± 0.12
A34G	**1.04** ± 0.14
A36G	**0.52** ± 0.09
T38S	**0.19** ± 0.06
V39A	**0.19** ± 0.04
V42A	**0.19** ± 0.03
T44S	*
L45A	**0.08** ± 0.01
I49A	**0.09** ± 0.03
L50A	**0.13** ± 0.02
T51S	**0.65** ± 0.03

B

Table 4.5. A. Kinetic and thermodynamic parameters of wild-type (WT) and mutant G_A88 proteins calculated from chevron plot analysis. The chevron plots were fitted globally to a two-state model. The Tanford β value for the transition state was β_{TS} = 0.72 ± 0.04.

*B. Experimentally determined **Φ**-values for the G_A88 mutants.*

** Mutants with $\Delta\Delta G_{D-N}$ too low (<0.35 kcal mol^{-1}) to calculate a reliable **Φ**-value.*

	k_{DI} (s^{-1})	$k_{NI}/(1+Kpart)$ (s^{-1})	$Kpart$	ΔG_{D-N} (kcal mol^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})
WT	1600 ± 120	1.6 ± 0.1	0.03 ± 0.001	4.1 ± 0.3	
L5A	48 ± 3	35 ± 3	0.08 ± 0.004	0.18 ± 0.1	3.9 ± 0.2
I6V	1400 ± 110	1.9 ± 0.1	0.04 ± 0.001	3.9 ± 0.4	0.17 ± 0.06
L7A	285 ± 25	27 ± 3	0.31 ± 0.010	1.4 ± 0.1	2.7 ± 0.2
L12A	560 ± 30	1.5 ± 0.1	0.05 ± 0.001	3.5 ± 0.4	0.57 ± 0.01
T16S	960 ± 50	7.7 ± 0.1	0.10 ± 0.002	2.8 ± 0.1	1.2 ± 0.1
T17S	1100 ± 100	2.4 ± 0.1	0.04 ± 0.001	3.6 ± 0.3	0.46 ± 0.01
T18S	800 ± 80	37 ± 3	0.13 ± 0.005	1.8 ± 0.1	2.3 ± 0.3
A20G	950 ± 90	47 ± 4	0.08 ± 0.003	1.8 ± 0.1	2.3 ± 0.2
A23G	550 ± 40	11 ± 1	0.05 ± 0.002	2.3 ± 0.3	1.7 ± 0.2
A24G	980 ± 80	2.8 ± 0.5	0.04 ± 0.001	3.5 ± 0.1	0.61 ± 0.30
T25S	1300 ± 100	1.6 ± 0.4	0.04 ± 0.001	4.0 ± 0.5	0.12 ± 0.02
A34G	930 ± 90	27 ± 1	0.19 ± 0.005	2.1 ± 0.1	2.0 ± 0.3
T38S	1500 ± 120	0.4 ± 0.01	0.01 ± 0.001	4.9 ± 0.1	0.81 ± 0.02
V39A	1370 ± 130	8.4 ± 0.3	0.12 ± 0.001	3.0 ± 0.3	1.1 ± 0.1
T44S	1170 ± 110	1.7 ± 0.1	0.04 ± 0.001	3.9 ± 0.1	0.22 ± 0.04
T49S	880 ± 90	26 ± 2	0.03 ± 0.001	2.1 ± 0.1	2.0 ± 0.1
T51S	550 ± 40	8.2 ± 0.4	0.08 ± 0.004	2.5 ± 0.2	1.6 ± 0.2
T53S	810 80	6.7 ± 0.5	0.06 ± 0.001	2.8 ± 0.3	1.3 ± 0.2
V54A	380 30	34 ± 2	0.20 ± 0.004	1.4 ± 0.4	2.6 ± 0.3
T55S	2000 150	3.2 ± 0.3	0.06 ± 0.001	3.8 ± 0.5	0.28 ± 0.02

A

	Φ1	**Φ2**
WT		
L5A	**0.53** ± 0.13	**0.66** ± 0.10
I6V	*	*
L7A	**0.38** ± 0.01	**0.87** ± 0.12
L12A	**1.08** ± 0.03	**1.07** ± 0.03
T16S	**0.24** ± 0.01	**0.76** ± 0.05
T17S	**0.44** ± 0.01	**0.50** ± 0.04
T18S	**0.18** ± 0.01	**0.53** ± 0.10
A20G	**0.13** ± 0.01	**0.36** ± 0.06
A23G	**0.36** ± 0.01	**0.51** ± 0.05
A24G	**0.46** ± 0.02	**0.71** ± 0.05
T25S	*	*
A34G	**0.16** ± 0.01	**0.67** ± 0.08
T38S	**0.05** ± 0.02	**0.56** ± 0.02
V39A	**0.08** ± 0.01	**0.76** ± 0.06
T44S	*	*
T49S	**0.17** ± 0.01	**1.04** ± 0.10
T51S	**0.39** ± 0.02	**0.70** ± 0.05
T53S	**0.31** ± 0.01	**0.58** ± 0.03
V54A	**0.32** ± 0.01	**0.71** ± 0.03
T55S	**0.28** ± 0.02	**0.80** ± 0.02

B

*Table 4.6. A. Kinetic and thermodynamic parameters of wild-type (WT) and mutant G_B30 proteins. The chevron plots were fitted globally to a three-state model. k_{DI} is the rate constant for the formation of the intermediate from the denatured state; k_{NI} is the rate constant for the unfolding of the native state to the intermediate state; K_{part} is the partitioning factor k_{ID}/k_{IN}. The Tanford β value for the two transition states were $\beta_{TS1} = 0.77 \pm 0.04$ and $\beta_{TS2} = 0.97 \pm 0.04$. B. Experimentally determined Φ-values for the G_B30 mutants calculated for the first (Φ1) and the second (Φ2) transition state. * Mutants with $\Delta\Delta G_{D-N}$ too low (<0.35 kcal mol^{-1}) to calculate a reliable Φ-value.*

	k_{DI} (s^{-1})	$k_{NI}/(1+Kpart)$ (s^{-1})	$Kpart$	ΔG_{D-N} (kcal mol^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})
WT	107 ± 10	0.5 ± 0.01	0.03 ± 0.001	3.1 ± 0.3	
I6V	88 ± 6	0.7 ± 0.00	0.05 ± 0.001	2.9 ± 0.2	0.25 ± 0.02
L7A	22 ± 1	8.3 ± 0.1	0.27 ± 0.003	0.6 ± 0.1	2.6 ± 0.3
L12A	79 ± 5	0.6 ± 0.01	0.04 ± 0.001	2.9 ± 0.1	0.25 ± 0.02
A16G	139 ± 11	3.7 ± 0.1	0.08 ± 0.001	2.1 ± 0.1	1.0 ± 0.2
I17V	91 ± 9	0.4 ± 0.04	0.03 ± 0.001	3.2 ± 0.1	0.08 ± 0.01
T18S	43 ± 4	5.9 ± 0.2	0.04 ± 0.012	1.2 ± 0.1	2.0 ± 0.1
A23G	49 ± 3	4.4 ± 0.1	0.05 ± 0.001	1.4 ± 0.1	1.7 ± 0.3
A24G	56 ± 3	1.1 ± 0.1	0.04 ± 0.001	2.3 ± 0.1	0.82 ± 0.04
T25S	96 ± 8	0.5 ± 0.01	0.04 ± 0.001	3.1 ± 0.2	0.08 ± 0.02
L32A	86 ± 9	0.2 ± 0.01	0.02 ± 0.001	3.6 ± 0.3	0.48 ± 0.02
A34G	51 ± 5	12 ± 0.2	0.20 ± 0.001	0.9 ± 0.1	2.3 ± 0.3
A36G	130 ± 11	1.1 ± 0.1	0.05 ± 0.001	2.8 ± 0.1	0.35 ± 0.02
T38S	137 ± 12	0.1 ± 0.01	0.01 ± 0.001	4.2 ± 0.4	-1.1 ± 0.3
V39A	85 ± 6	3.8 ± 0.1	0.18 ± 0.003	1.8 ± 0.1	1.3 ± 0.3
V42A	76 ± 7	1.0 ± 0.1	0.03 ± 0.001	2.5 ± 0.1	0.60 ± 0.02
T44S	84 ± 7	0.5 ± 0.01	0.03 ± 0.001	3.0 ± 0.3	0.17 ± 0.01
T49S	54 ± 7	0.8 ± 0.01	0.03 ± 0.001	2.5 ± 0.1	0.63 ± 0.04
T51S	26 ± 2	4.0 ± 0.1	0.11 ± 0.001	1.1 ± 0.1	2.0 ± 0.3
T53S	61 ± 3	2.2 ± 0.1	0.07 ± 0.001	2.0 ± 0.1	1.2 ± 0.1
T55S	91 ± 1	1.4 ± 0.1	0.07 ± 0.001	2.5 ± 0.2	0.66 ± 0.02

A

	Φ_{TS1}	Φ_{TS2}
WT		
I6V	*	*
L7A	**0.36** ± 0.02	**0.84** ± 0.04
L12A	*	*
A16G	**0.15** ± 0.01	**0.35** ± 0.01
I17V	*	*
T18S	**0.27** ± 0.02	**0.33** ± 0.03
A23G	**0.26** ± 0.02	**0.37** ± 0.03
A24G	**0.46** ± 0.02	**0.66** ± 0.05
T25S	*	*
L32A	**0.27** ± 0.02	**0.24** ± 0.03
A34G	**0.19** ± 0.01	**0.65** ±
A36G	**0.34** ± 0.03	**0.26** ± 0.03
T38S	**0.13** ± 0.01	**0.81** ± 0.06
V39A	**0.10** ± 0.01	**0.86** ± 0.05
V42A	**0.34** ± 0.03	**0.33** ± 0.02
T44S	*	*
T49S	**0.63** ± 0.04	**0.54** ± 0.04
T51S	**0.41** ± 0.03	**0.74** ± 0.04
T53S	**0.28** ± 0.03	**0.62** ± 0.06
T55S	**0.14** ± 0.01	**0.79** ± 0.04

B

*Table 4.7. Kinetic and thermodynamic parameters of wild-type (WT) and mutant G_B77 proteins. The chevron plots were fitted globally to a three-state model. k_{DI} is the rate constant for the formation of the intermediate from the denatured state; k_{NI} is the rate constant for the unfolding of the native state to the intermediate state; K_{part} is the partitioning factor k_{ID}/k_{IN}. The Tanford β value for the two transition states were $\beta_{TS1} = 0.69 \pm 0.04$ and $\beta_{TS2} = 0.96 \pm 0.04$. B. Experimentally determined Φ-values for the G_B77 mutants calculated for the first (Φ1) and the second (Φ1) transition state. * Mutants with $\Delta\Delta G_{D-N}$ too low (<0.35 kcal mol^{-1}) to calculate a reliable Φ-value.*

	k_F (s^{-1})	k_U (s^{-1})	ΔG_{D-N} (kcal mol^{-1})	$\Delta\Delta G_{D-TS}$ (kcal mol^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})
WT	96 ± 5	4.3 ± 0.1	1.8 ± 0.1		
I6V	61 ± 7	6.9 ± 0.7	1.3 ± 0.1	0.26 ± 0.01	0.54 ± 0.01
L9A	47 ± 1	21 ± 2	0.5 ± 0.01	0.42 ± 0.02	1.4 ± 0.1
A12G	220 ± 20	1.8 ± 0.1	2.8 ± 0.1	-0.49 ± 0.01	-1.00 ± 0.1
L20A	55 ± 1	0.2 ± 0.03	3.3 ± 0.1	0.33 ± 0.01	-1.4 ± 0.1
A24G	16 ± 1	12 ± 0.1	0.2 ± 0.02	1.07 ± 0.04	1.7 ± 0.1
T25S	115 ± 10	6.0 ± 0.3	1.7 ± 0.1	-0.11 ± 0.01	0.09 ± 0.01
Y33F	106 ± 10	8.0 ± 0.1	1.5 ± 0.1	-0.06 ± 0.01	0.35 ± 0.01
A34G	96 ± 6	13 ± 1	1.2 ± 0.1	0.00 ± 0.01	0.67 ± 0.01
A36G	86 ± 5	7.8 ± 0.3	1.4 ± 0.1	0.06 ± 0.01	0.41 ± 0.01
T38S	125 ± 10	1.6 ± 0.1	2.6 ± 0.2	-0.16 ± 0.01	-0.74 ± 0.01
V39A	79 ± 6	8.7 ± 0.4	1.3 ± 0.1	0.11 ± 0.01	0.53 ± 0.01
V42A	62 ± 6	10 ± 1	1.1 ± 0.1	0.25 ± 0.01	0.75 ± 0.01
T44S	95 ± 9	1.7 ± 0.1	2.4 ± 0.2	0.00 ± 0.01	-0.55 ± 0.01
T49S	29 ± 1	7.2 ± 0.2	0.8 ± 0.01	0.70 ± 0.03	1.0 ± 0.1
T51S	25 ± 1	18 ± 1	0.2 ± 0.02	0.80 ± 0.03	1.6 ± 0.1
T53S	40 ± 2	14 ± 1	0.6 ± 0.01	0.52 ± 0.02	1.2 ± 0.1
V54A	101 ± 9	17 ± 1	1.1 ± 0.1	-0.03 ± 0.02	0.78 ± 0.01
T55S	75 ± 6	8.5 ± 0.1	1.3 ± 0.1	0.15 ± 0.01	0.55 ± 0.01

A

	Φ
WT	
I6V	**0.48** ± 0.01
L9A	**0.31** ± 0.04
A12G	**0.49** ± 0.01
L20A	**-0.23** ± 0.01
A24G	**0.64** ± 0.02
T25S	*
Y33F	**-0.19** ± 0.01
A34G	**0.00** ± 0.01
A36G	**0.15** ± 0.01
T38S	**0.21** ± 0.01
V39A	**0.21** ± 0.01
V42A	**0.34** ± 0.01
T44S	**0.00** ± 0.01
T49S	**0.70** ± 0.05
T51S	**0.49** ± 0.10
T53S	**0.42** ± 0.03
V54A	**-0.04** ± 0.01
T55S	**0.26** ± 0.02

B

Table 4.8 A. Kinetic and thermodynamic parameters of wild-type (WT) and mutant G_B88 proteins, calculated from chevron plot analysis. The chevron plots were fitted globally to a two-state model. The Tanford β value for the transition state was $\beta_{TS} = 0.82 \pm 0.04$.

*B. Experimentally determined **Φ**-values for the G_B88 mutants.*

** Mutants with $\Delta\Delta G_{D-N}$ too low (<0.35 kcal mol^{-1}) to calculate reliable **Φ**–values.*

DISCUSSION

5.1 G_A88 and G_B88

It is well known that proteins with a significant similarity in their amino acid sequence are expected to have the same fold. In fact, analysis of the protein data bank (PDB) reveals that a sequence similarity of approximately 40% generally leads to a conserved three-dimensional structure and function (Wilson, et al. 2000). As reported in the Introduction, an interesting question is which physical-chemical properties of the amino acid sequence dictate a particular fold. This conundrum originated, in 1994, the famous "Paracelsus Challenge" set forth by Rose and Creamer, which was to convert one protein fold into another by changing less than 50% of the original sequence (Rose and Creamer 1994). Amazingly, this goal was fully achieved in only 3 years, when Dalal and co-workers (Dalal, et al. 1997) designed a sequence that, in spite of being 50% identical to a mostly β-sheet protein, folded into a four-helix bundle. Since then, several other scientists have achieved similarly impressive feats of design (Davidson 2008).

Interestingly, in 2008 it was demonstrated by Roessler and co-workers (Roessler et a., 2008) that a pair of natural homologue proteins, displaying 40% identity, exhibits markedly different folds. These proteins, named Pfl 6 and Xfaso 1, are both repressors of the Cro family and were identified in prophage sequences present in the genomes of the bacterial species, *Pseudomonas fluorescens* (Pfl 6) and *Xylella fastidiosa* (Xfaso 1). The three-dimensional structures of these two proteins reveal a similar N-terminal helix–turn–helix but widely diverging C-terminal region: Xfaso 1 displays an all-helical monomeric fold, whereas the Pfl 6 folds into an intertwined β-sheet dimer. These unexpected observations definitely overturn the basic assumption that proteins with significant similarity in their amino acid sequence share an identical fold.

In this perspective, the work by Bryan and co-workers led to the design of a pair of proteins with an extraordinarily high degree of sequence identity (88%) but different structure and function (Alexander, et al. 2007; He, et al. 2008). The successful design of these two heteromorphic proteins, called

G_A88 and G_B88, provides the opportunity to address the protein folding problem from a novel perspective. In fact, one may ask definite questions, such as: What are the key residues that determine a given structure? At which stages of folding the polypeptide chain commits to its native topology?

We have extensively characterized the mechanism and pathway of folding and unfolding of G_A88 and G_B88 at different pH values. Our results show that both engineered proteins appear to fold via a two-state mechanism, and protein topology is committed very early along the folding pathway. In particular, we detected clear differences in the denatured state properties of these two proteins. In fact, while just 7 of the 56 amino acids are different in between G_A88 and G_B88, only the latter displays a detectable residual structure in its denatured state. The extent of this residual structure may be tuned by changing pH, as reflected by analysis of $m_{D\text{-}N}$ values as a function of pH (pKa circa 5). Inspection of Figure 4.5 B shows that the $m_{D\text{-}N}$ decreases with decreasing pH, suggesting that the denatured state of G_B88 becomes more compact at acidic pH. Importantly, however, even at neutral pH, the observed $m_{D\text{-}N}$ is lower than that calculated from equilibrium experiments (i.e. 0.93 ± 0.05 kcal mol^{-1} M^{-1}), suggesting the presence of residual structure in the denatured state under more "physiological" conditions. Surprisingly, the 7 residues that are different in between the two proteins (see Figure 1.9) do not include amino acids titrating below neutral pH, suggesting that the observed compaction of the denatured state of G_B88 originates from long-distance effects.

Additional details regarding the structural features of the denatured states of G_A88 and G_B88 come from the molecular dynamics simulations performed in collaboration with Valerie Daggett and reported in the experimental Section. It appears that the denatured state of G_B88 displays some structural features that are reminiscent of its native topology, with detectable residual structure in the central α-helix (see Figure 4.6). This observation is also consistent with what predicted by the algorithm AGADIR (Munoz and Serrano 1997) that indicates a high helical propensity for the only α-helix of G_B88 as shown in Figure 5.1.

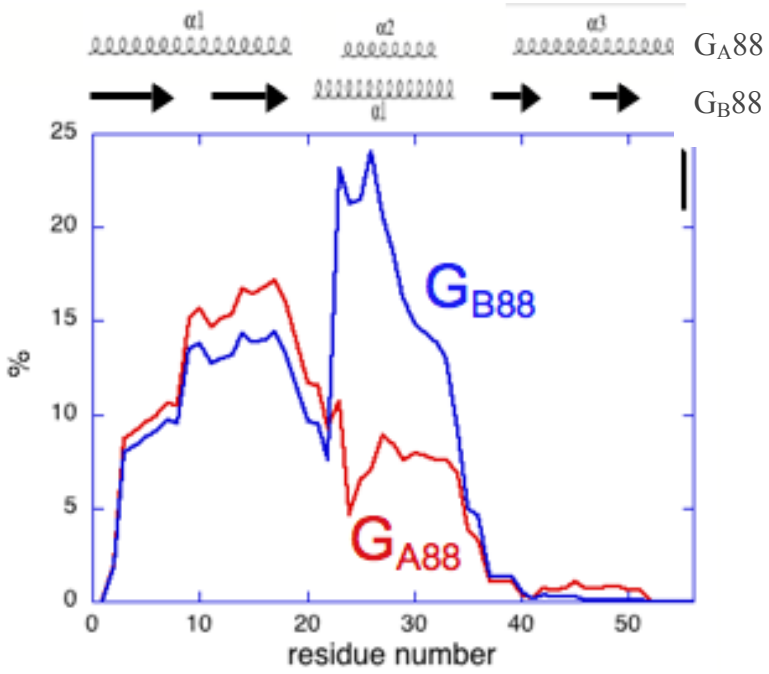


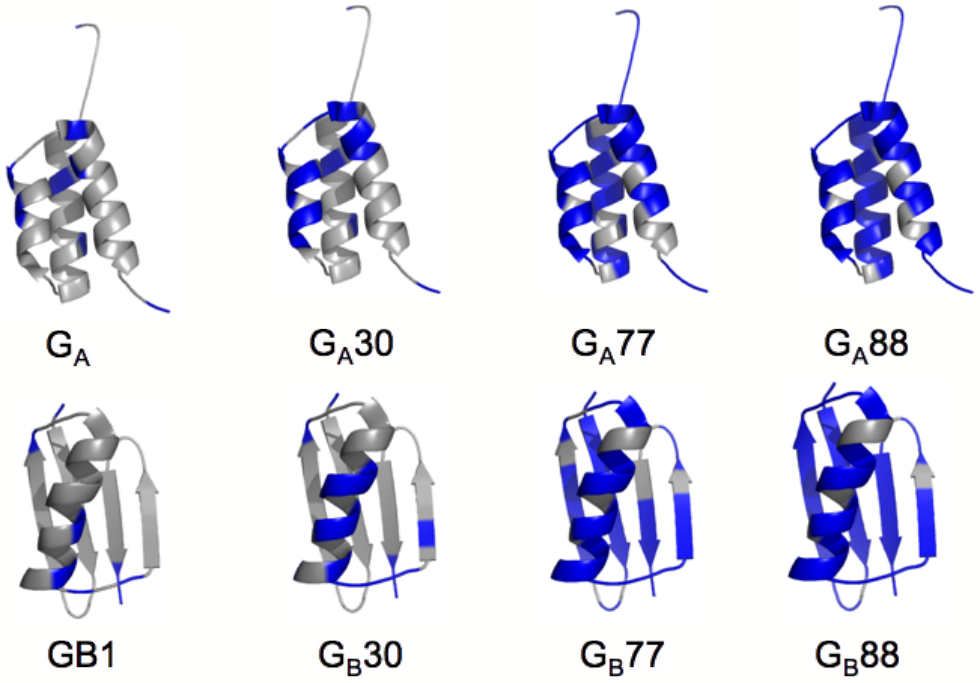
Figure 5.1. AGADIR profile of percentage helix probability based on local propensities for G_A88 (blue line) and G_B88 (red line). Secondary structure of both the proteins is shown on top of the AGADIR profile.

A plausible mechanism, emerging from both experiments and simulation, implies that the presence of a long central helix in G_B88 prevents the formation of the loop connecting α1 and α2 in G_A88. Ultimately this leads to a mixed α+β structure rather than the three-helix bundle topology of G_A88. Interestingly, it was recently shown that a population switch between the G_A and G_B structures may be induced even by a single amino-acid substitution (Alexander, et al. 2009). Thus, under conditions where the G_A fold is ~90% populated, mutation of Leu 45 into Tyr shifts the population to ~90% of the alternative G_B fold. This represents a clear-cut effect of conformational shift by mutational perturbation. Surprisingly, position 45 is not in the loop where G_A88 and G_B88 display a different helical propensity (Figure 1.9), providing additional clues for the hypothesis that the more pronounced helical content of the denatured state of G_B88 is affected by long-range interactions.
Overall, the comparison of the folding of G_A88 and G_B88 highlights a conundrum: based on the folding mechanism no single residue appears to act as a unique gatekeeper in the selection of protein topology, although only a

few residues (or possibly a single one) are responsible for the selective stabilization of either one of the two topologies. Thus it may be concluded that the folding of G_A88 and G_B88 suggests that native topology might be pre-sculpted in the denatured state, where incipient nuclei are present. Stabilization of such nuclei is affected by long-range interactions, and commitment to the native fold occurs by selective stabilization of these incipient nuclei, rather than by actively blocking alternative pathways.

5.2 The folding pathway of GB1: detecting an unexpected folding intermediate

The main task of the work is focussed on the comparison of the folding pathways of G_A88 and G_B88, with particular emphasis on the identification of common intermediates (if any). The surprising finding that the folding of these proteins is committed early, already in the denatured state, prompted us to carry out a systematic and very extensive analysis of each of the heteromorphic variants designed by Philip N. Bryan (Alexander, et al. 2007; He, et al. 2008), including the natural G_A and G_B domains. In the case of the natural G_A domain, we noted that this protein does not contain any intrinsic fluorescence probes (i.e. no Trp) and differs by only three amino acids from the variant G_A30. Therefore we omitted its characterization from our analysis and decided to focus on the folding pathway of GB1, a popular system for protein folding studies, as well as the set of heteromorphic pairs G_A30 and G_B30, G_A77 and G_B77, G_A88 and G_B88 (Figure 5.2).
The extensive experimental results obtained for the different domains are briefly discussed below.
The folding pathway of the small α/β protein GB1 has been extensively studied during the past two decades using both theoretical and experimental approaches (Alexander, et al. 1992; Clarke, et al. 1999b; Krantz, et al. 2002; Ding, et al. 2004; Chung, et al. 2010). Most of these studies provided a consensus view that the protein folds in a two-state manner.




```
GA    MEAVDANSLA QAKKAAIKEL KQYGIGDYTI KLINNAKTVE GVESLKNEIL KALPTE
GB1   MTYKLILNGK TLKGETTTEA VDAATAEKYF KQYANDNGVD GEWTYDDATK TFTVTE

GA30 MEAVDANSLA QAKEAAIKEL KQYGIGEKYI KLINNAKTVE GVWSLKNEIL KALPTE
GB30 MTYKLILNGK TLKGETTTEA VDAATAEKYF KLYANDKTVE GEWTYDDATK TFTVTE

GA77 TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIL KATVTE
GB77 TTYKLILNGK QLKEEAITEA VDAATAEKYF KLYANAKTVE GVWTYKDETK TFTVTE

GA88 TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIL TFTVTE
GB88 TTYKLILNLK QAKEEAIKEL VDAATAEKYF KLYANAKTVE GVWTYKDETK TFTVTE
```

Figure 5.2. Structures and sequence alignments of the different G_A and G_B variants. All engineered G_A and G_B proteins, designed by Bryan and co-workers, display structure and function similar to their respective natural wild-type domains G_A and GB1. For each protein, amino acid identities are shown in blue and nonidentities in in gray.

In the course of our characterization of the folding of GB1, we noted that this protein, when challenged over a wide range of denaturant concentrations, appeared to display a previously undetected complex behavior (namely a 'roll-over' effect, highlighted in Figure 4.9). Because this roll-over effect was inconsistent with a simple two-state behavior (as discussed in the

Introduction), we decided to study the folding pathway of GB1 under a variety of different experimental conditions, i.e. carrying out equilibrium and kinetic experiments over a wide range of pH values (from 2.0 to 9.6).

The existence of an intermediate in the folding of GB1 was previously proposed by Roder and coworkers (Park, et al. 1997; Park, et al. 1999), who suggested the presence of a collapsed state, accumulating in the *ms* time-range, as observed by continuous-flow ultra rapid mixing experiments. This low-energy intermediate was later questioned by Sosnick and co-workers (Krantz, et al. 2002). Importantly, however, the partially folded state identified in this work, is distinct from that previously suggested, being a high energy species that never accumulates and whose presence suggests that GB1 folds *via* a complex and rough energy barrier with at least two discrete major transition states.

When folding is characterized by a complex chevron plot, the deviation from linearity observed may have different origins: i) the curvature may be due to movement of the position of the transition state along a single broad barrier (Oliveberg 1998; Oliveberg, et al. 1998); or ii) the curvature may reflect a change in the rate-limiting step suggestive of a multi-state process (Bachmann and Kiefhaber 2001; Sauder J.M. 1996; Walkenhorst, et al. 1997; Wildegger and Kiefhaber 1997). In the case of GB1, the curved chevron plots are observed only above pH 6 (Figure 4.8), making the broad transition state model less likely since it would imply that it is possible i) to distort drastically the folding free-energy profile and ii) to switch between a narrow energy maximum (linear chevron plot) to a broad energy maximum (curved chevron plot) by changing pH. Furthermore, in analogy to what has been observed previously for other proteins (Bachmann and Kiefhaber 2001; Gianni, et al. 2009; Gianni, et al. 2005), detection of a roll-over effect only under some solvent conditions seems more consistent with a multi-step folding pathway. The observed chevron plots of GB1, obtained at different pH conditions, were fitted globally. The excellent statistical parameters of the global fit suggest that the two transition states are relatively robust and maintain their overall structural features when solvent conditions are varied. This observation is in stark contrast with previous experimental work, which suggested the unfolding *m*-value to depend strongly on experimental conditions (McCallister, et al. 2000; Park, et al. 1997; Park, et al. 1999). We conclude that both transition states *TS1* and *TS2* display a robust structure that is by-and-large maintained when solvent and/or sequence composition is altered.

It is of interest to analyze the dependence of the unfolding rate constants measured for the two transition states at different pH values. In fact, both energy barriers appear to display a monotonic transition at acidic and alkaline

pH values. Both profiles are consistent with a model involving the protonation of two different groups with apparent pK_a of ~ 4 and ~ 8 (Figure 4.10). While the alkaline transition displays approximately the same change in activation free energy for both *TS1* and *TS2*, the acidic transition is more pronounced in *TS1*, suggesting the presence of a charged interaction that is weak or not formed in *TS1* and is consolidated in *TS2*. Inspection of the three-dimensional structure of GB1 suggests that such an interaction may be either Lys4-Glu15 (located at the N-terminal β-hairpin) or Lys10-Glu56 (between the N-terminal turn and the C-terminus of the protein). These considerations led to the idea that the structures of the transition states appear robust to changes in pH and may be characterized by an extended nucleus, which is stabilized by both the N- and C-terminal beta-hairpins, as well as by contacts between the N- and C-terminal strands.

5.3 Comparing the folding pathway of the G_A and G_B variants at nearly atomic resolution: Φ-value analyses.

We have characterized, by extensive Φ-value analysis, the complete folding pathway of G_A30, G_A77, G_A88, as well as G_B30, G_B77 and G_B88. Whilst all the G_A proteins were consistent with a two state folding transition, in the case of G_B, we could detect an intermediate similarly to what observed for the naturally occurring GB1. Such an intermediate was more evident in the case of G_B30 and G_B77, whose chevron plots were obtained as a function of guanidine concentration. In the case of G_B88, we could not perform the kinetic experiments in presence of guanidine because the stability of most of the mutants was far too low. Thus the Φ-value analysis of G_B88 was carried out with urea as denaturant agent, allowing us to measure only the Φ-values corresponding to first transition state. In an effort to test whether the nature of the denaturant affects the experimentally determined Φ-values, we obtained the chevron plot of some selected variant using both denaturing agents. Importantly, we did not observe any significant variation in Φ-values when urea instead of guanidine was used.

When mapped onto the native structures, the experimentally determined Φ-values revealed a conserved transition state among all the three G_A variants (G_A30, G_A77, G_A88). In particular, the medium and high values of Φ, hallmarks of native interactions in the transition state, clustered at the interface between the helices α1 and α2 (Figure 4.13). These observations strongly suggests that the structure of the transition state of G_A is robust to

changes in amino acid composition and that folding to the G_A topology involves the formation of a conserved nucleus.

In the case of G_B, we observed a different behavior. In particular, the distribution of the measured Φ-values for the first transition state (Φ_1), plotted on the native structure of G_B, shows considerable differences among G_B30, G_B77 and G_B88. In fact, considering G_B30 and G_B88 it is possible to observe, a shift of the medium-high Φ-values from the first β-hairpin to the second, with G_B77 displaying an intermediate behavior (Figure 4.14). These results indicate that alternative folding nuclei, located at the hairpins between β1-β2 and β3-β4, drive the folding to the G_B-topology. These nuclei may be selectively stabilized depending on amino acid compositions.

Recently, it has been proposed that the number of accessible pathways for folding is determined by the different nucleation motifs contained within a given native topology (Figure 5.3). For example, the structure of ribosomal protein S6 seems to be composed of two different nucleation patterns, which act as indipendent cooperative units. It has been demostrated that each of these motifs constitutes a separate entry to parallel folding trajectories (Lindberg and Oliveberg 2007). Accordingly, it may be suggested that, in the case of G_B, the symmetrical organization of its three-dimensional topology implies the presence of multiple nucleation motifs that permit alternative folding pathways.

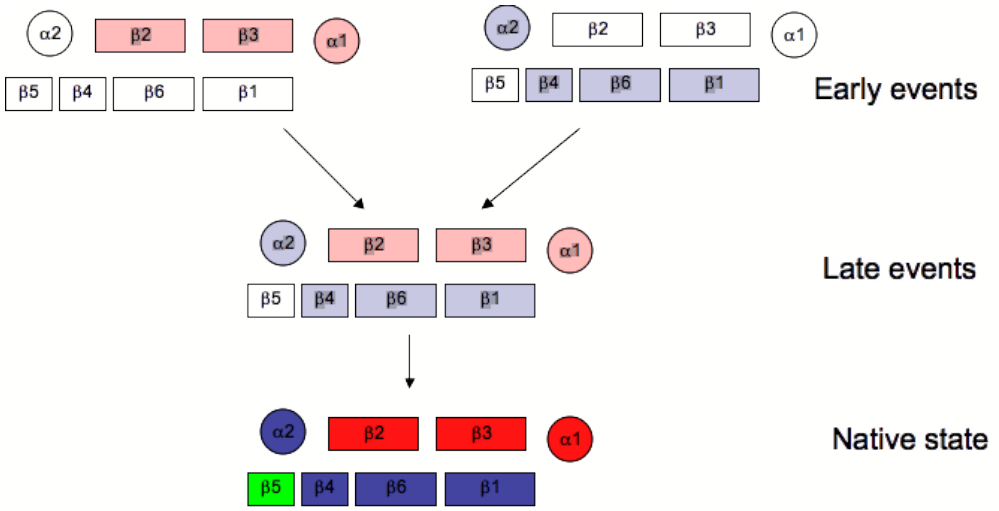


Figure 5.3. Schematic representation of the proposed model for explaining converging folding pathways as a consequence of the presence of alternative folding nuclei. The representation was designed based on the topology of the PDZ domain fold (S.Gianni personal communication). Two distinct nucleation motifs (indicated in pink and violet respectively) are formed in the early events of folding. The partitioning between these events may lead to alternative pathways, as for example observed in the G_B proteins. These two nuclei act as independent units that can be selectively stabilized by altering the sequence composition. At the late stages of folding, structure formation at the level of both nuclei results in apparently converging pathways. An extensive discussion of this scenario is reported in (Calosci et al., 2008).

REFERENCES

Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich
1994 Specific nucleus as the transition state for protein folding: evidence from the lattice model. Biochemistry 33(33):10026-36.
Alexander, P. A., et al.
2007 The design and characterization of two proteins with 88% sequence identity but different structure and function. Proc Natl Acad Sci U S A 104(29):11963-8.
Alexander, P. A., et al.
2009 A minimal sequence code for switching protein structure and function. Proc Natl Acad Sci U S A 106(50):21149-54.
Alexander, P., J. Orban, and P. Bryan
1992 Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. Biochemistry 31(32):7243-8.
Alonso, D. O., and V. Daggett
1998 Molecular dynamics simulations of hydrophobic collapse of ubiquitin. Protein Sci 7(4):860-74.
Anfinsen, C. B., et al.
1961 The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A 47:1309-14.
Bachmann, A., and T. Kiefhaber
2001 Apparent two-state tendamistat folding is a sequential process along a defined route. J Mol Biol 306(2):375-86.
Bai, Y., et al.
1995 Protein folding intermediates: native-state hydrogen exchange. Science 269(5221):192-7.
Baker, D.
2000 A surprising simplicity to protein folding. Nature 405(6782):39-42.
Baldwin, C. T., A. M. Reginato, and D. J. Prockop
1989 A new epidermal growth factor-like domain in the human core protein for the large cartilage-specific proteoglycan. Evidence for alternative splicing of the domain. J Biol Chem 264(27):15747-50.
Baldwin, R. L.

1989 How does protein folding get started? Trends Biochem Sci 14(7):291-4.

Banachewicz, W., et al.

2011 Malleability of folding intermediates in the homeodomain superfamily. Proc Natl Acad Sci U S A 108(14):5596-601.

Bond, C. J., et al.

1997 Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. Proc Natl Acad Sci U S A 94(25):13409-13.

Bowler, B. E.

2007 Thermodynamics of protein denatured states. Mol Biosyst 3(2):88-99.

Bowler, B. E.

2011 Residual structure in unfolded proteins. Curr Opin Struct Biol.

Brockwell, D. J., and S. E. Radford

2007 Intermediates: ubiquitous species on folding energy landscapes? Curr Opin Struct Biol 17(1):30-7.

Brunori, M., et al.

2003 Cytochrome c(551) as a model system for protein folding. Biophys Chem 100(1-3):409-19.

Brunori, M., et al.

2011 Morphogenesis of a protein: folding pathways and the energy landscape. Biochemical Society Transactions In Press.

Bryngelson, J. D., et al.

1995 Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21(3):167-95.

Buchner J. and Kiefhaber J.

2005 Protein Folding Handbook, Part I. Wiley Weinheim.

Calosci, N., et al.

2008 Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. Proc Natl Acad Sci U S A 105(49):19241-6.

Capaldi, A. P., C. Kleanthous, and S. E. Radford

2002 Im7 folding mechanism: misfolding on a path to the native state. Nat Struct Biol 9(3):209-16.

Capaldi, A. P., et al.

2001 Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. Nat Struct Biol 8(1):68-72.

Cellmer, T., et al.

2007 Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration. J Am Chem Soc 129(47):14564-5.

Chattopadhyay, K., E. L. Elson, and C. Frieden

2005 The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods. Proc Natl Acad Sci U S A 102(7):2385-9.

Chi, C.N., et al.

2007 A conserved folding mechanism for PDZ domains. FEBS Lett. 581:1109-1113.

Chiti, F., et al.

1999a Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat Struct Biol 6(11):1005-9.

Chiti, F., et al.

1999b Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat. Struct. Biol. 6:1005-1009.

Cho, J. H., and D. P. Raleigh

2009 Experimental characterization of the denatured state ensemble of proteins. Methods Mol Biol 490:339-51.

Chung, H. S., J. M. Louis, and W. A. Eaton

2009 Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. Proc Natl Acad Sci U S A 106(29):11837-44.

Chung, H. S., J. M. Louis, and W. A. Eaton

2010 Distinguishing between protein dynamics and dye photophysics in single-molecule FRET experiments. Biophys J 98(4):696-706.

Clarke, J., et al.

1999a Folding studies of Ig-like beta-sandwich proteins suggest they share a common folding pathway. Structure 7:1145-1153.

Clarke, J., et al.

1999b Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. Structure 7(9):1145-53.

Colon, W., et al.

1996 Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome c folding. Biochemistry 35(17):5538-49.

Daggett, V., and A. R. Fersht

2003 Is there a unifying mechanism for protein folding? Trends Biochem Sci 28(1):18-25.

Dalal, S., S. Balasubramanian, and L. Regan
1997 Protein alchemy: changing beta-sheet into alpha-helix. Nat. Struct. Biol. 4:548-552.

Davidson, A.R.
2008 A folding space odyssey. Proc. Natl. Acad. Sci. U S A 105:2759-2760.

Day, R., and V. Daggett
2003 All-atom simulations of protein folding and unfolding. Adv Protein Chem 66:373-403.

Dill, K. A.
1985 Theory for the folding and stability of globular proteins. Biochemistry 24(6):1501-9.

Dill, K. A., et al.
1995 Principles of protein folding--a perspective from simple exact models. Protein Sci 4(4):561-602.

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

Ding, K., J. M. Louis, and A. M. Gronenborn
2004 Insights into conformation and dynamics of protein GB1 during folding and unfolding by NMR. J Mol Biol 335(5):1299-307.

Dyson, H. J., and P. E. Wright
2002 Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Adv Protein Chem 62:311-40.

Dyson, H. J., and P. E. Wright
2004 Unfolded proteins and protein folding studied by NMR. Chem Rev 104(8):3607-22.

Elove, G. A., A. K. Bhuyan, and H. Roder
1994 Kinetic mechanism of cytochrome c folding: involvement of the heme and its ligands. Biochemistry 33(22):6925-35.

Eyring, H.
1935 J. Chem. Phys. 3:107.

Fanning, A. S., and J. M. Anderson
1999 PDZ domains: fundamental building blocks in the organization of protein complexes at the plasma membrane. J Clin Invest 103(6):767-72.

Ferguson, N., et al.
1999 Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. J Mol Biol 286(5):1597-608.

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

Fersht, A. R.
1995 Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. Proc Natl Acad Sci U S A 92(24):10869-73.

Fersht, A. R.
1997 Nucleation mechanisms in protein folding. Curr Opin Struct Biol 7(1):3-9.

Fersht, A.
1999 Structure and Mechanism in Protein Science. Freeman W.H. and Co New York.

Fersht, A. R., and V. Daggett
2002 Protein folding and unfolding at atomic resolution. Cell 108(4):573-82.

Fersht, A. R. and Sato, S.
2004 Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. USA* 101, 7976–7981

Friel, C.T., A.P. Capaldi, and S.E. Radford
2003 Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. J. Mol. Biol. (326):293-305.

Frauenfelder H., Sligar S.G., Wolynes P.G.
1991 The energy landscapes and motions of proteins. Science. Dec 13;254(5038):1598-603.

Geierhaas, C.D., et al.
2007 BPPred: a Web-based computational tool for predicting biophysical parameters of proteins. Protein Sci. 16:125-134.

Gianni , Mayor , and Fersht
2003 Structural insights in the folding of small single-domain proteins. The Italian Journal of Biochemistry 52(4).

Gianni, S
2012 THE Φ VALUE ANALYSIS. Encyclopedia of Biophysics in press.

Gianni, S., et al.
2009 Distinguishing between smooth and rough free energy barriers in protein folding. Biochemistry 48(49):11825-30.

Gianni, S., M. Brunori, and C. Travaglini-Allocatelli

2001a Refolding kinetics of cytochrome c(551) reveals a mechanistic difference between urea and guanidine. Protein Sci 10(8):1685-8.

Gianni, S., et al.

2005 Kinetic folding mechanism of PDZ2 from PTP-BL. Protein Eng Des Sel 18(8):389-95.

Gianni, S., et al.

2007a A PDZ domain recapitulates a unifying mechanism for protein folding. Proc Natl Acad Sci U S A 104(1):128-33.

Gianni, S., et al.

2003 Unifying features in protein-folding mechanisms. Proc Natl Acad Sci U S A 100(23):13286-91.

Gianni, S., et al.

2010 Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. Nat Struct Mol Biol 17(12):1431-7.

Gianni, S., et al.

2007b Identification and characterization of protein folding intermediates. Biophys Chem 128(2-3):105-13.

Gianni, S., et al.

2001b Snapshots of protein folding. A study on the multiple transition state pathway of cytochrome c(551) from Pseudomonas aeruginosa. J Mol Biol 309(5):1177-87.

Haglund E et al.

2008 Changes of protein-folding pathways by circular permutation: overlapping nuclei promote global cooperativity. J. Biol. Chem. 283:27904-27915.

Harrison, S. C., and R. Durbin

1985 Is there a single pathway for the folding of a polypeptide chain? Proc Natl Acad Sci U S A 82(12):4028-30.

He, Y., et al.

2008 NMR structures of two designed proteins with high sequence identity but different fold and function. Proc. Natl. Acad. Sci. U S A 105:14412-14417.

Hyeon, C., and D. Thirumalai

2003 Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? Proc Natl Acad Sci U S A 100(18):10249-53.

I.E. Sanchez, T. Kiefhaber

2003 J. Mol. Biol. 327:867–884.

Itzhaki, L. S., D. E. Otzen, and A. R. Fersht

1995 The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J Mol Biol 254(2):260-88.

Ivarsson, Y., et al.

2008 Folding and misfolding in a naturally occurring circularly permuted PDZ domain. J Biol Chem 283(14):8954-60.

J. N. Onuchic, N. D. Socci, Z. Luthey-Schulten and P. G. Wolynes,

1996 Protein Folding Funnels: The Nature of the Transition State Ensemble. Folding & Design 1(6):441-450.

J.N. Onuchic, P.G. Wolynes

2004 Theory of Protein Folding. Curr. Opinion Structural Biology 14:70-75.

Jackson, S. E., and A. R. Fersht

1991 Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. Biochemistry 30(43):10436-43.

Jackson, S.E, Fersht, A.R.

1991 Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30:10428-10435.

Jackson, S.E.

1998 How do small single-domain proteins fold? Fold. Des. 3:R81-91.

Jemth, P., et al.

2004 Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. Proc Natl Acad Sci U S A 101(17):6450-5.

Karplus, M., and D. L. Weaver

1976 Protein-folding dynamics. Nature 260(5550):404-6.

Karplus, M. and McCammon J.A.

2002 Molecular dynamics simulations of biomolecules. Nat Struct Biol. 9(9):646-52.

Kazmirski, S. L., and V. Daggett

1998a Non-native interactions in protein folding intermediates: molecular dynamics simulations of hen lysozyme. J Mol Biol 284(3):793-806.

Kazmirski, S. L., and V. Daggett

1998b Simulations of the structural and dynamical properties of denatured proteins: the "molten coil" state of bovine pancreatic trypsin inhibitor. J Mol Biol 277(2):487-506.

Kennedy, D., and C. Norman
2005 What don't we know? Science 309(5731):75.
Khorasanizadeh, S., I. D. Peters, and H. Roder
1996 Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nat Struct Biol 3(2):193-205.
Kmiecik, S., and A. Kolinski
2008 Folding pathway of the b1 domain of protein G explored by multiscale modeling. Biophys J 94(3):726-36.
Krantz, B. A., et al.
2002 Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. J Mol Biol 324(2):359-71.
Levinthal, C.
1968 Are there pathways for protein folding? Journal de Chimie Physique et de Physico-Chimie Biologique 65:44-45.
Lindberg, M. O., and M. Oliveberg
2007 Malleability of protein folding pathways: a simple reason for complex behaviour. Curr Opin Struct Biol 17(1):21-9.
Martinez, J. C., M. T. Pisabarro, and L. Serrano
1998 Obligatory steps in protein folding and the conformational diversity of the transition state. Nat Struct Biol 5(8):721-9.
Martínez, J.C., and L. Serrano
1999 The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nat. Struct. Biol. 6:1010-1016.
Matouschek, A., et al.
1990 Transient folding intermediates characterized by protein engineering. Nature 346(6283):440-5.
Matouschek, A., et al.
1989 Mapping the transition state and pathway of protein folding by protein engineering. Nature 340(6229):122-6.
Matthews , CR
1993 Pathways of protein folding. Annu Rev Biochem 62:653-683.
Mayor, U., et al.
2003 The complete folding pathway of a protein from nanoseconds to microseconds. Nature 421:863-867.
McCallister, E. L., E. Alm, and D. Baker
2000 Critical role of beta-hairpin formation in protein G folding. Nat Struct Biol 7(8):669-73.
Meng, W., et al.

2009 Native like structure in the unfolded state of the villin headpiece helical subdomain, an ultrafast folding protein. Protein Sci 18(8):1692-701.

Mirny, L. A., V. I. Abkevich, and E. I. Shakhnovich

1998 How evolution makes proteins fold quickly. Proc Natl Acad Sci U S A 95(9):4976-81.

Morrone, A. , et al.

2011a Gb1 Is not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate. . Biophysical Journal 101:1-8.

Morrone, A., et al.

2011b The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. J Biol Chem 286(5):3863-72.

Munoz, V., and L. Serrano

1997 Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. Biopolymers 41(5):495-509.

Myers, J. K., C. N. Pace, and J. M. Scholtz

1995 Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci 4(10):2138-48.

Neuweiler, H., C. M. Johnson, and A. R. Fersht

2009 Direct observation of ultrafast folding and denatured state dynamics in single protein molecules. Proc Natl Acad Sci U S A 106(44):18569-74.

Oliveberg, M.

1998 Alternative explanations for multi-state kinetics in protein folding:

transient aggregation and changing transition-state ensambles. Acc. Chem. Res. 31:765-772.

Oliveberg, M., and A. R. Fersht

1996 Formation of electrostatic interactions on the protein-folding pathway. Biochemistry 35(8):2726-37.

Oliveberg, M., et al.

1998 The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. J Mol Biol 277(4):933-43.

Otzen, D. E., and A. R. Fersht

1998 Folding of circular and permuted chymotrypsin inhibitor 2: retention of the folding nucleus. Biochemistry 37(22):8139-46.

Otzen, D. E., et al.

1999 Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. Biochemistry 38(20):6499-511.

Ozkan, S. B., I. Bahar, and K. A. Dill

2001 Transition states and the meaning of Phi-values in protein folding kinetics. Nat Struct Biol 8(9):765-9.

Pace, C. N.

1986 Determination and analysis of urea and guanidine hydrochloride denaturation curve. Methods Enzymol 131:266-280.

Pace, C.N.

1990 Measuring and increasing protein stability. . Trends Biotechnol. 8:93–98.

Park, S. H., K. T. O'Neil, and H. Roder

1997 An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. Biochemistry 36(47):14277-83.

Park, S. H., M. C. Shastry, and H. Roder

1999 Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. Nat Struct Biol 6(10):943-7.

Parker, M.J., J. Spencer, and A.R. Clarke

1995 An integrated kinetic analysis of intermediates and transition states in protein folding reactions. J. Mol. Biol. 253(5):771-86.

Pletneva, E. V., H. B. Gray, and J. R. Winkler

2005 Many faces of the unfolded state: conformational heterogeneity in denatured yeast cytochrome C. J Mol Biol 345(4):855-67.

Religa, T.L., et al.

2005 Solution structure of a protein denatured state and folding intermediate. Nature 437:1053-1056.

Riddle, D.S., et al.

1999 Experiment and theory highlight role of native state topology in SH3 folding. Nat. Struct. Biol. 6:1016-1024.

Rose, G.D., and T.P. Creamer

1994 Protein folding: predicting predicting. Proteins. Struct. Funct. Bioinf. 19:1-3.

Ruan, B., et al.

2004 Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification. Biochemistry 43:14539-14546.

Sanchez, I.E., and T. Kiefhaber

2003 Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. J. Mol. Biol. 325(2):367-376.

Santoro, M. M., and D. W. Bolen

1988 Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry 27(21):8063-8.

Sauder J.M., MacKenzie N.E. and Roder H.

1996 Kinetic Mechanism of Folding and Unfolding of Rhodobacter capsulatus

Cytochrome c2. Biochemistry 35:16852-16862.

Scaloni, F., et al.

2010 Deciphering the folding transition state structure and denatured state properties of nucleophosmin C-terminal domain. Proc Natl Acad Sci U S A 107(12):5447-52.

Scaloni, F., et al.

2009 Folding mechanism of the C-terminal domain of nucleophosmin: residual structure in the denatured state and its pathophysiological significance. FASEB J 23(8):2360-5.

Schellman, J. A.

1978 Solvent denaturation Biopolymers 17:1305–1322.

Scott, K. A., and J. Clarke

2005 Spectrin R16: broad energy barrier or sequential transition states? Protein Sci 14(6):1617-29.

Scott, K. A., L. G. Randles, and J. Clarke

2004 The folding of spectrin domains II: phi-value analysis of R16. J Mol Biol 344(1):207-21.

Shastry, M. C., S. D. Luck, and H. Roder

1998 A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. Biophys J 74(5):2714-21.

Shea, J. E., J. N. Onuchic, and C. L. Brooks, 3rd

2002 Probing the folding free energy landscape of the Src-SH3 protein domain. Proc Natl Acad Sci U S A 99(25):16064-8.

Shortle, D.

1995 Staphylococcal nuclease: a showcase of m-value effects. Adv Protein Chem 46:217-47.

Shortle, D., and M. S. Ackerman

2001 Persistence of native-like topology in a denatured protein in 8 M urea. Science 293(5529):487-9.

Shortle, D., and A. K. Meeker

1986 Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. Proteins 1(1):81-9.

Silow, M., and M. Oliveberg

1997 Transient aggregates in protein folding are easily mistaken for folding intermediates. Proc Natl Acad Sci U S A 94(12):6084-6.

Szilagyi, A. Kardos, J. Osvath, S. Barna, L. Zavodszky P.

2007 Protein Folding (Handbook of neurochemistry and molecular neurobiology, Cap 10). Springer-Verlag Berlin Heidelberg.

Tanford, C.

1968 Protein denaturation. Adv Protein Chem 23:121-282.

Ternstrom, T., et al.

1999 From snapshot to movie: phi analysis of protein folding transition states taken one step further. Proc Natl Acad Sci U S A 96(26):14854-9.

Timasheff, S.N.

1993 The control of protein stability and association by weak interactions with water: how do solvents affect these processes? Annu Rev Biophys Biomol Struct. 22:67-97.

Tirado-Rives, J., and W. L. Jorgensen

1993 Molecular dynamics simulations of the unfolding of apomyoglobin in water. Biochemistry 32(16):4175-84.

Travaglini-Allocatelli, C., et al.

1999 Folding mechanism of Pseudomonas aeruginosa cytochrome c551: role of electrostatic interactions on the hydrophobic collapse and transition state properties. J Mol Biol 289(5):1459-67.

Travaglini-Allocatelli, C., S. Gianni, and M. Brunori

2004 A common folding mechanism in the cytochrome c family. Trends Biochem Sci 29(10):535-41.

Travaglini-Allocatelli, C., et al.

2005 An obligatory intermediate in the folding pathway of cytochrome c552 from Hydrogenobacter thermophilus. J. Biol. Chem. 280:25729-25734.

Travaglini-Allocatelli, C., et al.

2003 Exploring the cytochrome c folding mechanism: cytochrome c552 from thermus thermophilus folds through an on-pathway intermediate. J Biol Chem 278(42):41136-40.

Tsai, J., M. Levitt, and D. Baker

1999 Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. J Mol Biol 291(1):215-25.

van Gunsteren, W. F., et al.

2001 The Key to Solving the Protein-Folding Problem Lies in an Accurate Description of the Denatured State Financial support from the Schweizerischer Nationalfonds (Project no. 21-50929.97) is gratefully acknowledged. Angew Chem Int Ed Engl 40(2):351-355.

Vendruscolo, M., et al.

2001 Three key residues form a critical contact network in a protein folding transition state. Nature 409(6820):641-5.

Walkenhorst, W.F., S.M. Green, and H. Roder

1997 Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. Biochemistry 36:5795-5805.

Wetlaufer, D. B.

1973 Nucleation, rapid folding, and globular intrachain regions in proteins. Proc Natl Acad Sci U S A 70(3):697-701.

White, G. W., et al.

2005 Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. J Mol Biol 350(4):757-75.

Wildegger, G., and T. Kiefhaber

1997 Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. Journal of Molecular Biology 270(2):294-304.

Wilson, C.A., J. Kreychman, and M. Gerstein

2000 Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. J. Mol. Biol. 297:233-249.

Wong, K. B., et al.

2000 Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. J Mol Biol 296(5):1257-82.

Wright, C. F., et al.

2003 Parallel protein-unfolding pathways revealed and mapped. Nat Struct Biol 10(8):658-62.

Zarrine-Afsar, A., S. M. Larson, and A. R. Davidson

2005 The family feud: do proteins with similar structures fold via the same pathway? Curr Opin Struct Biol 15(1):42-9.

Zarrine-Afsar, A., S. L. Lin, and P. Neudecker

2010 Mutational investigation of protein folding transition states by Phi-value analysis and beyond: lessons from SH3 domain folding. Biochem Cell Biol 88(2):231-8.

ATTACHMENTS

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 286, NO. 5, pp. 3863–3872, February 4, 2011

The Denatured State Dictates the Topology of Two Proteins with Almost Identical Sequence but Different Native Structure and Function*[S]

Received for publication, June 18, 2010, and in revised form, November 19, 2010 Published, JBC Papers in Press, November 29, 2010, DOI 10.1074/jbc.M110.155911

Angela Morrone[‡], Michelle E. McCully[§], Philip N. Bryan[¶], Maurizio Brunori[‡], Valerie Daggett[§1], Stefano Gianni[‡2], and Carlo Travaglini-Allocatelli[‡]

From the [‡]Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli," Università di Roma "La Sapienza", 5 00185 Rome, Italy, the [§]Biomolecular Structure and Design Program and Department of Bioengineering, University of Washington, Seattle, Washington 98195, and the [¶]Institute for Bioscience and Biotechnology Research/Department of Bioengineering, University of Maryland, Rockville, Maryland 20850

The protein folding problem is often studied by comparing the mechanisms of proteins sharing the same structure but different sequence. The recent design of the two proteins G_A88 and G_B88, displaying different structures and functions while sharing 88% sequence identity (49 out of 56 amino acids), allows the unique opportunity for a complementary approach. At which stage of its folding pathway does a protein commit to a given topology? Which residues are crucial in directing folding mechanisms to a given structure? By using a combination of biophysical and computational techniques, we have characterized the folding of both G_A88 and G_B88. We show that, contrary to expectation, G_B88, characterized by a native $\alpha+\beta$ fold, displays in the denatured state a content of native-like helical structure greater than G_A88, which is all-α in its native state. Both experiments and simulations indicate that such residual structure may be tuned by changing pH. Thus, despite the high sequence identity, the folding pathways for these two proteins appear to diverge as early as in the denatured state. Our results suggest a mechanism whereby protein topology is committed very early along the folding pathway, being imprinted in the residual structure of the denatured state.

Understanding the rules that govern the folding of proteins is one of the main unsolved problems in modern science (1). Current knowledge on the protein folding reaction has been achieved by extensively characterizing the folding mechanisms of simple globular proteins (2), and a comprehension of the folding pathways of larger multidomain systems is still far from being achieved. Given the diversity of protein structures and amino acid compositions, it is extremely difficult to draw general rules by studying folding kinetics of individual proteins. In fact, when considering the folding of different proteins, a comparison may be jeopardized by the variability in amino acid sequence and in the three-dimensional structure of the native and denatured states.

A powerful approach to elucidate relationships between sequence information and folding mechanism is to study proteins that differ in sequence but share the same overall fold (3–11). This strategy assumes that general correlations between amino acid sequences and folding pathways may be extrapolated by comparing folding processes of different members of a given protein family.

Generally, proteins with significant sequence similarity are expected to have a similar fold. In fact, analysis of the Protein Data Bank (PDB) reveals that a sequence similarity of 40% nearly always leads to a similar fold (12). This observation provoked Rose and Creamer (13) in 1994 to issue the "Paracelsus Challenge," whereby the protein folding community was confronted with the task of designing two proteins that are at least 50% identical but possessed different folds. Amazingly, this goal was fully achieved in only 3 years, when Regan and co-workers (14) designed a sequence that, despite being 50% identical to a mostly β-sheet protein, folded into a four-helix bundle. Since then, others have achieved similarly impressive results (15). Recently, ambitious work by Bryan and co-workers (16, 17) led to the design of pairs of proteins with an extraordinarily high degree of sequence identity but different folds and different functions. In particular, the sequences of two domains from streptococcal protein G were subjected to an iterative design of heteromorphic pairs, leading the authors to produce two protein G variants, called G_A88, which is mostly α-helical (the 3-helix bundle protein A fold), and G_B88, displaying the $\alpha+\beta$ protein G fold (Fig. 1). These two proteins share 88% sequence identity (49 out of 56 amino acids), yet they display two different structures and functions that are similar to the respective wild-type proteins. In parallel with studies on protein families, this protein engineering achievement offers the unique opportunity for a complementary study on protein folding mechanism addressing two key questions. 1) At which stage of its folding pathway does a pro-

* This work was partially supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (Grants 2007B57EAB_004, 20074TJ3ZB_005, and RBRN07BMCT_007). This work was supported, in whole or in part, by National Institutes of Health Grants GM50789 (to V. D.) and GM062154 (to P. N. B.). This work was also supported by a grant from the Department of Defense through the National Defense Science and Engineering Graduate Fellowship Program (to M. E. M.) for the MD studies.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S3.

[1] To whom correspondence may be addressed. Tel.: 206-543-9305; E-mail: daggett@u.washington.edu.

[2] To whom correspondence may be addressed. Dipartimento di Scienze Biochimiche "A. Rossi Fanelli," Università di Roma "La Sapienza," P.le A. Moro, 5 00185 Rome, Italy. Tel.: 39-06-49910548; Fax: 39-06-4440062; E-mail: stefano.gianni@uniroma1.it.

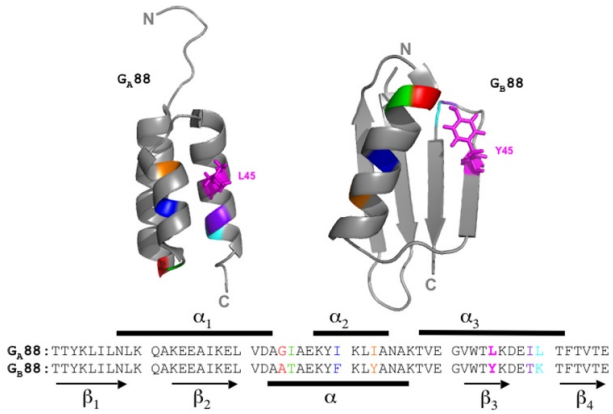


FIGURE 1. **G_A88 and G_B88 structures.** Sequence alignment and secondary structure are shown below. The 7 residues that differ between the two proteins are shown in different colors: 24 (*red*), 25 (*green*), 30 (*blue*), 33 (*orange*), 45 (*magenta*), 49 (*violet*), and 50 (cyan). As discussed under "Discussion", under conditions where the G_A fold is >90% populated, mutation of Leu^{45} into Tyr (highlighted in *sticks* on the structures and in *bold* in the sequence alignment) shifts the population to >90% of the G_B fold (38).

tein commit to a given topology? 2) Which residues are crucial in directing folding to a given native structure?

Here we present for the first time an extensive characterization of the folding mechanisms of G_A88 and G_B88 by experiment and molecular dynamics (MD)[3] simulation. The results obtained under a variety of solvent conditions suggest the presence in the denatured state of G_B88 of pH-sensitive residual structure, as indicated by a pH dependence of its m_{D-N} value, which is not observed for G_A88. The MD simulations are consistent with these findings, showing that for G_B88 the non-polar solvent-accessible surface area decreases markedly at low pH. Interestingly, in analogy with earlier observations on a similar heteromorphic protein A/protein G pair sharing 59% sequence identity (*i.e.* A219 and G311) (18), the extent of native-like helical structure in the denatured state of G_B88 (the protein G $\alpha+\beta$ fold) is greater than that of denatured G_A88, which is all-α in its native conformation. Both our current and our earlier studies of this system suggest that protein topology is committed very early along the folding pathway being "imprinted" in the residual structure of the denatured state; this weak, loosely defined topology is sufficient to dictate the pathway of folding. The significance of these observations from the perspective of previous work on the folding of proteins with the same topology but very different sequences is discussed.

EXPERIMENTAL PROCEDURES

The buffers used were 50 mM sodium phosphate from pH 8.0 to 6.3, 50 mM sodium acetate from pH 5.5 to 3.8, 50 mM sodium formate from pH 3.4 to 3.0, and 50 mM sodium phosphate/phosphoric acid from pH 2.8 to 2.0. In an effort to check whether the buffer composition had any effect on the folding experiments, we performed additional control experiments in the presence of 50 mM Tris at pH 7.2 and 8.0, 50 mM Mes at pH 6.2 and 5.5, and 50 mM sodium phosphate (first protonation) at pH 2.0. Both the folding and the unfolding reactions were essentially independent from buffer composition. All reagents were analytical grade.

Protein Expression and Purification

Bryan and co-workers (16) cloned G_A88 and G_B88 genes into the vector pG58, which encodes an engineered subtilisin pro-sequence as the N terminus of the fusion protein. Proteins were purified employing an affinity chromatography previously developed (19). Soluble cell extract of pro-domain fusion proteins was injected on a 5-ml Bio-Scale^{TM} Mini Profinity eXact cartridge at 5 ml/min to allow binding and then washed with 10 column volumes of 100 mM $NaPO_4$ (pH 7.2) to remove impurities. To cleave and elute the purified target proteins, 15 ml of 100 mM NaF in the presence of 100 mM $NaPO_4$ (pH 7.2) were injected at 5 ml/min. After the first 10 ml, the flow was stopped, and the column was incubated for 30 min to allow complete cleavage. After elution, the purified proteins were then dialyzed to remove potassium fluoride.

Equilibrium Unfolding

Equilibrium denaturations were followed on a JASCO circular dichroism (CD) spectropolarimeter (JASCO, Inc., Easton, MD), in a 1-cm quartz cuvette (Hellma). CD spectra of G_A88 and G_B88 were recorded between 250 and 200 nm. Protein concentrations were typically 6 μM.

[3] The abbreviations used are: MD, molecular dynamics; TS, transition state; r.m.s.d., root mean square deviation; DSSP, Dictionary of Protein Secondary Structure; SASA, solvent-accessible surface area.

Stopped-flow Measurements

Single mixing kinetic folding experiments were carried out on a Pi-star stopped-flow instrument (Applied Photophysics, Leatherhead, UK); the excitation wavelength was 280 nm, and the fluorescence emission was measured using a 320 nm cut-off glass filter. In all experiments, performed at 25 and 10 °C, refolding and unfolding were initiated by an 11-fold dilution of the denatured or the native protein with the appropriate buffer. Final protein concentrations were typically 1 μM. The observed kinetics were always independent of protein concentration (from 0.5 to 5 μM), as expected from monomolecular reactions without effects due to transient aggregation (20).

Data Analysis

Equilibrium experiments—Assuming a standard two-state model, the urea-induced denaturation transitions were fit to the equation

$$\Delta G_d = m_{D-N}(D - D_{50}) \quad \text{(Eq. 1)}$$

where m_{D-N} is the slope of the transition (proportional to the increase in solvent-accessible surface area ongoing from the native to the denatured state) and D_{50} is the midpoint of the denaturation transition. An equation that takes into account the pre- and post-transition baselines was used to fit the observed unfolding transition (21).

Kinetic Experiments—Analysis was performed by non-linear least squares fitting of single exponential phases using the fitting procedures provided in the Applied Photophysics software. The chevron plots were fitted, using the Kaleidagraph software package, by numerical analysis based on a two-state model following the equation

$$k_{obs} = k_F + k_U \quad \text{(Eq. 2)}$$

where k_F and k_U represent the folding and unfolding rate constants, respectively. The logarithm of each microscopic rate constant was assumed to vary linearly with denaturant concentration (22).

Molecular Dynamics Simulations

We performed 11 all-atom, explicit solvent MD simulations for each protein, G_A88 and G_B88: 298 K (30 ns), 498 K neutral and low pH (3×50 ns, 2×5 ns), for a total of 700 ns (0.7 μs) of simulation time. The starting structures were taken from the published NMR ensembles (17) (G_A88, PDB id 2jws, Model 1; G_B88, PDB id 2jwu, Model 3 (298 K), Model 1 (498 K)). Low pH systems were created by protonating all aspartate and glutamate residues (neither protein contains a histidine). The simulations were all performed using our in-house MD software, *in lucem* molecular mechanics (*il*mm),[4] with the Levitt *et al.* (24) all-atom force field and the microcanonical ensemble (N, constant number of particles, system volume, and total energy). Non-bonded terms were treated with an 8 (at 498 K) or 12 Å (298 K) force-shifted cutoff. The proteins were minimized and solvated with explicit F3C flexible waters (25) using our standard protocol (26). Briefly, the protein was treated *in vacuo* for 1000 steps of steepest descent minimization. Pre-equilibrated F3C water was added within 1.8 Å of the protein to a box extending at least 10–12 Å from the protein on all sides. The water was then minimized for 1000 steps and then subjected to 500 ps of dynamics with 2-fs time steps and an additional 500 steps of steepest descent minimization. Finally, the protein was minimized for 500 steps.

The denatured state ensemble was defined as the final 30 ns of the three longer (50 ns) 498 K simulations. The Cα root mean square deviation (r.m.s.d.) was calculated over all 56 residues as well as for just the core residues. The core was defined as the consecutive residues beginning with the N-terminal residue of the first secondary structure element as defined by He *et al.* (17) to the C terminus of the final element (G_A88, residues 9–51; G_B88, residues 1–55). The percentages of helix and solvent-accessible surface area (SASA) were calculated using our in-house implementations of the Dictionary of Protein Secondary Structure (DSSP) (27) and Lee and Richards (28) algorithms, respectively. The percentage of helix or β-structure was reported over the total 56 residues as the number of structured residues as defined by DSSP, and nonpolar SASA was reported as the sum of the SASA for all nonpolar residues (Ala, Val, Phe, Pro, Met, Ile, Leu, Trp, and Gly). When values were reported relative to the native state, the average value over the 30-ns 298 K simulation was used as the native value.

Multidimensional scaling of the all-against-all Cα r.m.s.d. matrix was performed using the R statistical package to assign conformational ensembles. The exit from the native-like cluster, or point of no return, in the three-dimensional projection of the multidimensional scaling was defined as the transition state (TS), and the preceding 5 ps was defined as the transition state ensemble, as described previously (29, 30). The average Cα r.m.s.d. to the starting structure was reported for each TS structure, as was the average pairwise Cα r.m.s.d. between all TS structures in the ensemble.

The fraction of time that residues were in contact was calculated for the native ensembles (all 30 ns of the native state), the TS ensembles (5 ps from each 498 K unfolding simulation), and the denatured state (the last 30 ns of the three long 498 K simulations). Two residues were considered in contact if they contained carbon atoms that were $<$5.4 Å apart or any other two non-hydrogen atoms that were $<$4.6 Å apart. Hydrogen bonds were measured in the denatured state (last 30 ns of the three 50-ns 498 K simulations) for specific residue pairs using the following criteria: 1) the distance between the donor hydrogen and acceptor atom was $\leq$2.6 Å; 2) the donor-hydrogen-acceptor angle was within 45° of linearity; and 3) the charges on the donor and acceptor atoms were <-0.3, and the charge on the hydrogen was $>+0.3$.

RESULTS

Equilibrium Unfolding of G_A88 and G_B88—To study the folding mechanism of G_A88 and G_B88, we carried out both equilibrium and kinetic (un)folding experiments. The results of urea-induced equilibrium denaturation of G_A88 and G_B88 measured at 10 °C, pH 7.2 in 50 mM sodium phosphate buffer,

[4] D. A. C. Beck, D. O. V. Alonso, and V. Daggett, in lucem molecular mechanics (*il*mm), University of Washington, Seattle, WA.

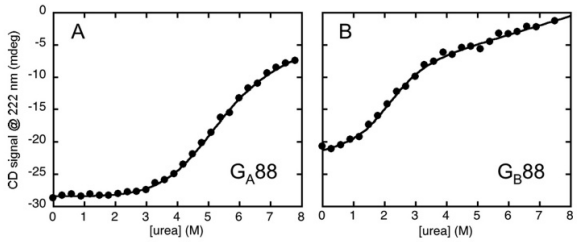


FIGURE 2. **Equilibrium denaturation of G_A88 (*A*) and G_B88 (*B*) monitored by CD in 50 mM sodium phosphate buffer at pH 7.2 and 10 °C.** Lines are the best fit to a two state unfolding mechanism. *mdeg*, millidegrees.

as monitored by far-UV circular dichroism (CD) spectroscopy, are provided in Fig. 2 (representative spectra for native and denatured G_A88 and G_B88 are also provided in supplemental Fig. S1). The observed transitions follow a simple two-state behavior, suggesting the absence of stable equilibrium intermediates for both proteins and indicating that these designed variants are capable of cooperative (un)folding reactions. Furthermore, the reaction was fully reversible under all conditions explored.

At physiological pH and in the absence of denaturant, the unfolding free energy of G_A88 derived from a two-state analysis and a global fit of 60 wavelengths (from 250 to 220 nm) is $\Delta G_{D-N} = 3.00 \pm 0.18$ kcal/mol, with a m_{D-N} value of 0.62 ± 0.04 kcal mol^{-1} M^{-1}. In the case of G_B88, $\Delta G_{D-N} = 2.35 \pm 0.30$ kcal/mol, and the m_{D-N} value is 1.10 ± 0.08 kcal mol^{-1} M^{-1}. Considering that the G_B88 construct contains ~10 structured residues more than G_A88, both m_{D-N} values are consistent with those expected for proteins of this size, according to the BPPred database (31). Hence, the seeming difference in cooperativity, as reflected by the different m_{D-N} values, can be accounted for by the difference in the number of structured residues between the two proteins.

Folding and Unfolding Kinetics—We carried out extensive fluorescence kinetic experiments on both proteins under a variety of different experimental conditions. In particular, the folding and unfolding kinetics were investigated at several pH values, ranging from 10 to 2. In the case of G_A88, it was not possible to measure reliable folding and unfolding rate constants at 25 °C over a wide range of denaturant concentrations because the rates were too fast for our stopped-flow apparatus. Thus, kinetic folding data for the two proteins were recorded at 10 °C to slow the process for G_A88 and to obtain values under the same conditions for comparison in the case of G_B88. In all cases, the folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final denaturant concentration (representative folding and unfolding time courses are reported as supplemental Fig. S2). Semi-logarithmic plots of the observed folding/unfolding rate constants of G_A88 and G_B88 *versus* denaturant concentration (*i.e.* chevron plots) at pH 7.2 are presented in Fig. 3. Both proteins displayed a V-shaped chevron plot, a hallmark of two-state folding (22). In the case of G_A88, there was excellent agreement between the thermodynamic parameters obtained

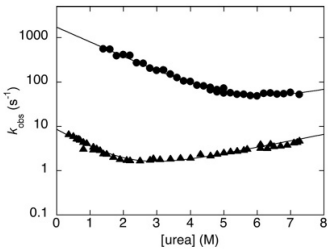


FIGURE 3. **Semi-logarithmic plot of the observed rate constant for folding and unfolding of G_A88 (●) and G_B88 (▲) *versus* [urea] at pH 7.2 in 50 mM NaP_i obtained at 10 °C, monitored by fluorescence emission.**

by equilibrium and kinetic data; however, in the case of G_B88, there was a minor deviation, the m_{D-N} value being 1.10 ± 0.08 kcal mol^{-1} M^{-1} from equilibrium experiments and 0.90 ± 0.05 kcal mol^{-1} M^{-1} from chevron plot analysis. As observed previously for other small single domain proteins (32–34), a significant deviation of m_{D-N} from equilibrium and kinetic data suggests that there are residual structure and/or changes in the exposure of non-polar residues in the denatured state of the protein. Because the small deviation observed for G_B88 is at the limit of experimental detection, we performed additional experiments under various conditions as well as MD simulations to further investigate these options, as described below.

A powerful method to address the global properties of folding transition and denatured states is the analysis of chevron plots recorded under different experimental conditions or on various site-directed variants (35). In fact, because the *m*-values (slopes of the unfolding and refolding arms of the chevron plots) reflect the change in accessible surface area upon (un)folding, analysis of their dependence on reaction conditions may be of diagnostic value to identify transition state movements along the reaction coordinate, as well as denatured state collapse or residual structure. In this study, we compared the folding kinetics of G_A88 and G_B88 at various pH values, ranging from 10 to 2. Inspection of Fig. 4 reveals that although both the stability and the *m*-values of G_A88 are insensitive to pH and this protein is fully native even at pH 2.0

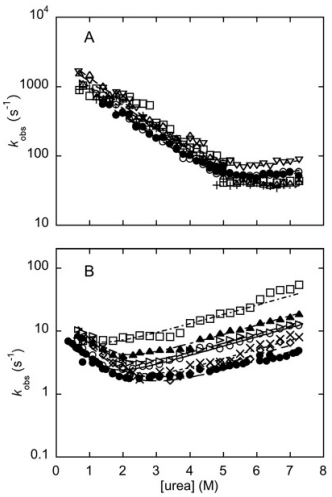


FIGURE 4. **Chevron plots of G_A88 and G_B88 at different pH values.** *A*, semi-logarithmic plot of the observed rate constant for folding and unfolding of G_A88 *versus* [urea] measured at different pH values and 10 °C (▽ pH 10.0, ○ pH 8.0, ● pH 7.2, □ pH 4.0, + pH 3.0, △ pH 2.0). *B*, semi-logarithmic plot of the observed rate constant for folding and unfolding of G_B88 *versus* [urea] measured at different pH values and 10 °C (○ pH 8.0, ● pH 7.2, ◇ pH 6.2, X pH 5.5, ▷ pH 4.7, ▲ pH 4.3, □ pH 4.0). The lines are the best fit to a two-state model.

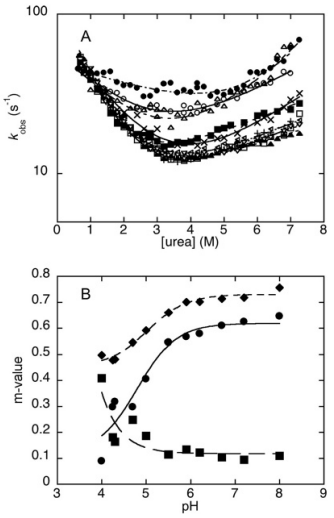


FIGURE 5. *A*, semi-logarithmic plot of the observed rate constants for folding and unfolding of G_B88 *versus* [urea] measured at different pH values in presence of 0.4 M sodium sulfate at 25 °C (■ pH 8.0, ▲ pH 7.2, ▽ pH 6.7, □ pH 6.2, + pH 5.9, ◁ pH 5.5, X pH 5.0, △ pH 4.7, ○ pH 4.3, ● pH 4.0); lines are the best fit to a two-state model. *B*, dependence of m_F (●), m_U (■), and m_{D-N} (◆) on pH; lines are the best fit to an equation implying a single protonation site with $pK_a \sim 5$.

(supplemental Fig. S3), G_B88 is destabilized at pH <5. However, the low stability of the latter protein and the poor definition of the observed refolding arms prevented a quantitative analysis of the *m*-values. Therefore as detailed below, we resorted to investigating the folding of G_B88 under stabilizing conditions.

Certain inorganic salts, such as phosphates and sulfates, favor compact protein conformations because of the preferential exclusion of solvent from the protein surface (36); this makes them potent stabilizers of both the native and the partially folded states. The chevron plots of G_B88 measured at different pH values and in the presence of 0.4 M sodium sulfate are provided in Fig. 5*A*. As expected, the stabilizing salt allows for a better definition of the refolding arms of the chevron plots. Consequently, we carried out a quantitative analysis of folding parameters over a wide range of pH conditions in the presence of salt. Fig. 5*B* shows the dependence on pH of calculated m_{D-N}, m_F, and m_U values for G_B88. The data fit to the protonation of a single titratable group with an apparent $pK_a \sim 5$. Interestingly, the m_{D-N} decreases with decreasing pH values, suggesting that the denatured state of this small single domain protein becomes more compact at acidic conditions. Importantly, however, even at neutral pH, the observed m_{D-N} is lower than that calculated from equilibrium experiments (*i.e.* 0.93 ± 0.05 kcal mol^{-1} M^{-1}), suggesting the presence of residual structure in the denatured state under physiological conditions.

Molecular Dynamics Simulations—To further investigate the differences between G_A88 and G_B88, MD simulations were conducted. Five independent thermal unfolding simulations were performed for each protein at 498 K at both neutral and low pH, in addition to a simulation at room temperature (298 K) for each protein as a control. Snapshots from the thermal unfolding at neutral pH are presented in Fig. 6. The sequence positions where the two proteins display different amino acids are highlighted as *balls*. Interestingly, in G_B88, the first hairpin ($\beta1/\beta2$) had a tendency to be loosely maintained in the denatured state, whereas the second hairpin ($\beta3/\beta4$) was more extended. In G_A88, the C-terminal region tended to collapse down more than in G_B88, leading to slightly more interactions involving $\beta3$ and $\beta4$ (residues 42–55). Specifically, these residues had 21.6 ± 5.5 internal residue-residue contacts in G_A88 *versus* 18.8 ± 4.5 contacts for G_B88. Finally, the central helix (Fig. 6, displayed in *green*) was fairly well preserved in the denatured state of G_B88.

A quantitative comparison of the unfolding simulations is provided in Table 1, where we report the average properties over the three independent unfolding simulations for each protein. The Cα r.m.s.d. relative to the starting structure reached over 11 Å in all cases. Although there was little

change in the Cα r.m.s.d. upon lowering the pH for G_A88, in the case of G_B88, the Cα r.m.s.d. increased at low pH. The overall residual helical content in the denatured state at neutral pH was similar for G_A88 and G_B88. The helix content of both proteins increased when the pH was lowered. Interestingly, the helical content in the denatured state of G_B88 at low pH was surprisingly high, being 104% of the native extent of helix content. In contrast, G_A88 contained 39% of its native helix content. Although we could not directly address the helical content of the denatured states, we experimentally observed the m_{D-N} value of G_B88 to decrease with decreasing pH (Fig. 5*B*). This observation is consistent with the MD simulations, suggesting the denatured state of G_B88 to be more structured at acidic than at neutral pH. Importantly, such a dependence was not observed in G_A88, whose m_{D-N} value was found insensitive to pH.

The relative compaction of G_B88 as compared with G_A88 can also be seen in Fig. 6. This effect is reflected in the non-polar SASAs of the two denatured states relative to their control native states (Table 1). The non-polar SASA at neutral pH was ~1000 Å^2 lower in the starting structures (1330 ± 83 and 1185 ± 72 Å^2 for G_A88 and G_B88, respectively). Relative to the native state, the non-polar SASA of the denatured state at neutral pH increased by 182% in the case of G_A88 and by 174% for G_B88. Although both proteins had some reduction of non-polar SASAs when the pH was lowered, G_B88 was more sensitive to acidification, and the increase in non-polar exposure upon unfolding was reduced relative to neutral pH. This finding parallels the experimental observations, which clearly indicate the denatured state of G_B88 to be more compact at acidic conditions, as reflected by a decreased m_{D-N} value (Fig. 5*B*).

TS ensembles were identified for each of the simulations described above, as well as the two shorter simulations for each protein at both pH conditions (Fig. 7). The representative TS structures for G_A88 show that their helical contents and overall size were quite similar at both pHs. On the other hand, the G_B88 TS ensemble was very sensitive to pH, and the structures at neutral pH were much more native-like than those at low pH. The β1/β2 hairpin was more robust than the β3/β4 hairpin in the TS. Furthermore, the two hairpins did not appear to interact directly and instead behaved as different entities physically separated by the central helix.

Interestingly, the heterogeneity of the G_A88 and G_B88 TS ensembles at neutral pH were similar, particularly when the effect of the unstructured N terminus of G_A88 was accounted for (compare the core Cα r.m.s.d. with the "To self" values in Table 2). Both ensembles are ~5 Å Cα r.m.s.d. from their respective starting structures. The pH sensitivity of the G_B88 TS ensemble was dramatic; the Cα r.m.s.d. to the starting structure increased by 1.5 Å, and the average Cα r.m.s.d. between structures within the TS ensemble also increased by 1.4 Å. The core Cα r.m.s.d. was 6.2 Å for the G_B88 TS ensemble at low pH, and it was quite distorted with a Cα r.m.s.d. of 7.0 Å

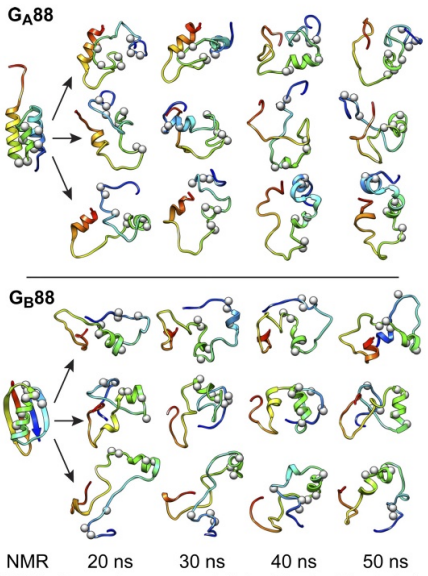


FIGURE 6. **Denatured state at neutral pH.** G_A88 (above) and G_B88 (below) are colored from *red* to *blue* from the N terminus to the C terminus. The Cα atoms of differing residues (24, 25, 30, 33, 45, 49, and 50) are shown as *gray balls*. The NMR structures of the native states are shown on the *left*, and the calculated structures at 20, 30, 40, and 50 ns for each of the three simulations at 498 K (*top*, 1; *middle*, 2; *bottom*, 3) are depicted.

TABLE 1

Properties of the native and denatured state ensembles from MD simulations at neutral (N) and low (L) pH

	pH and temperature	Cα r.m.s.d.[a]	% of α-helix[b]	% of native α-helix[c]	NP SASA[d]	% of native NP SASA[e]
		Å			Å^2	
G_A88	N, 298	3.4 (0.4)	67 (3)	100	1330 (83)	100
	N, 498	11.4 (1.3)	15.8 (10.7)	24 (16)	2424 (204)	182 (19)
	L, 498	11.8 (2.5)	26.4 (11.8)	40 (18)	2309 (224)	174 (20)
G_B88	N, 298	2.5 (0.4)	26 (1)	100	1185 (72)	100
	N, 498	11.3 (1.5)	15.5 (7.6)	59 (29)	2062 (198)	174 (20)
	L, 498	12.7 (1.2)	27.4 (13.8)	104 (52)	1867 (191)	158 (19)

[a] All properties were averaged over the final 30 ns of the three long 498 K simulations, and the S.D. is indicated in parentheses. The values for the 298 K simulations are defined as 100% native.
[b] α-Helix was defined using DSSP (27).
[c] Percentage of native α-helix was calculated relative to the percent of α-helix in the 30-ns 298 K native simulation.
[d] Nonpolar SASA (NP SASA) was defined as the sum of the SASA for all hydrophobic residues (see "Experimental Procedures").
[e] Percentage of native nonpolar SASA (NP SASA) was calculated relative to the nonpolar SASA in the 30-ns 298 K native simulation.

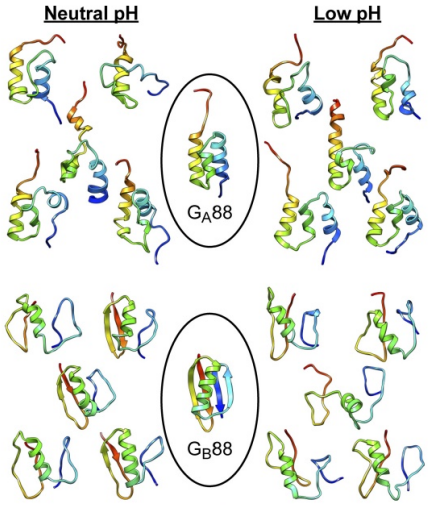


FIGURE 7. **Structures of the transition state at neutral and low pH.** G_A88 (above) and G_B88 (below) are colored from *red* to *blue* from the N terminus to the C terminus. The NMR structure of the native states is shown in a *ellipse* in the middle for each of the two proteins. For each set, runs 1–5 are ordered clockwise from top left, to top right, to middle, to bottom left, and to bottom right. Transition state times for G_A88 at neutral pH are at the following simulation times: 486, 483, 249, 331, and 225 ps for runs 1–5, respectively. Times for G_A88 at low pH are at: 362, 227, 842, 77, and 165 ps. Times for G_B88 at neutral pH are at: 422, 449, 248, 276, and 338 ps. Times for G_B88 at low pH are at: 159, 102, 164, 305, and 147 ps. Transition states were selected as the first cluster exit in the three-dimensional multidimensional scaling of the all-against-all Cα r.m.s.d. matrix. For more details, see "Experimental Procedures."

TABLE 2

Cα r.m.s.d. of the transition state from neutral and low pH simulations at 498 K

		Cα r.m.s.d.		Core[a] Cα r.m.s.d.	
	pH	To start[b]	To self[c]	To start	To self
		Å		Å	
G_A88	N	5.5 (0.7)	6.7 (1.0)	4.2 (0.5)	4.7 (0.4)
	L	4.9 (0.5)	5.5 (0.9)	3.5 (0.6)	3.7 (0.5)
G_B88	N	5.5 (1.1)	4.8 (0.9)	5.4 (1.1)	4.7 (0.9)
	L	7.0 (1.6)	6.2 (1.4)	6.9 (1.5)	6.2 (1.4)

[a] Core is residues 9–51 for G_A88 and 1–55 for G_B88. The core value removes the effect of the unstructured N terminus in G_A88. Also note that G_A88 low *versus* neutral pH is 4 Å and that the value comparing the two pH regions for G_B88 is 6 Å.

[b] The Cα r.m.s.d. to the starting structure was reported for all five transition state structures ($n = 5$).

[c] The pairwise Cα r.m.s.d. between all five transition state structures was reported ($n = 10$).

from its starting structure (Fig. 7). In contrast, the G_A88 TS remained within ~4 Å of its starting structure, and the spread within a TS ensemble was ~4.5 Å (excluding the N terminus of G_A88).

The overall folding pathways from representative simulations (run 1 in each case) of G_A88 and G_B88 are presented in Fig. 8 as the reverse of the simulated unfolding process. It appears that the topology, residual structure, and interactions in the denatured state direct whether the protein will fold into the helical G_A88 structure or the mixed α/β G_B88 structure. In G_A88, the protein has some dynamic residual structure, whereas the main chain is fairly fluid with different main chain interactions occurring over time within the collapsed state; the productive interactions with respect to folding are local along the sequence, with folding of kernels of helical structure that then dock together and consolidate in the TS ensemble. In contrast, in the case of G_B88, the approximate topology of the native state was already apparent in the denatured state, with segregation of the two hairpin regions by the central helix. This helix was more stable in G_B88 than in G_A88 due to improved packing interactions between residues 30 and 33, which in G_A88 are both Ile, whereas in the case of G_B88, they are Phe and Tyr. Moreover, substitution of Gly to Ala at residue 24 and Ile to Thr at residue 25 in G_B88 also increased the helical propensity of the region.

Interestingly, although the sequence of the N-terminal region is identical in the two proteins, there was a tendency for the region to form a helix in G_A88 and a loose hairpin in G_B88 (Figs. 7 and 8). This difference was, in large part, due to a hydrogen bond between Thr^1 and Glu^{19}, which was present in 99.8% of the G_B88 denatured state structures at neutral pH. At low pH, however, the loose $\beta1/\beta2$ hairpin was not present because Glu^{19} was protonated. Although positions 1 and 19 are identical in G_A88 and G_B88, in the case of the latter, we did not observe this hydrogen bond in the denatured state at neutral pH. Instead, Glu^{19} tended to interact with the solvent, and Thr^1 either interacted with the solvent or formed hydrogen bonds with Asp^{47} and Glu^{48} (over 67 and 60% of the denatured state, respectively). The backbone ϕ/ψ angles of residues 47 and 48 when interacting with Thr^1 were compatible with forming $\alpha3$ in G_A88. These observations indicate that long range interactions play a critical role in the residual structure in the denatured state of these proteins.

The $\beta3/\beta4$ turn was also fractionally present in the denatured state of G_B88 due to the presence of two side chain hydrogen bonds: Asp^{47}–Lys^{50} and Asp^{47}–Tyr^{45} (57 and 14% of the time at neutral pH, respectively). Of note, these hydrogen bonds were not present in the low pH denaturing simulations of G_B88 due to the protonation of Asp^{47}, nor were they present in the denatured state of G_A88 where both Lys^{50} and Tyr^{45} are mutated to Leu. This interaction appears to stabilize the $\beta3/\beta4$ turn in G_B88, thus preventing these residues from assuming an α-helical structure, as they do in G_A88.

DISCUSSION

Critical insights on many problems in biology have been classically achieved using simplified model systems. Although a comprehensive understanding of the folding of large multidomain proteins is still an aspiration, the successful design of two heteromorphic proteins sharing 88% sequence identity, called G_A88 and G_B88 (16), provides a unique opportunity to unveil the mechanism whereby a few key residues commit the polypeptide chain to its characteristic and functionally competent native topology. Here we have characterized the folding and unfolding kinetics of G_A88 and G_B88 by experiment

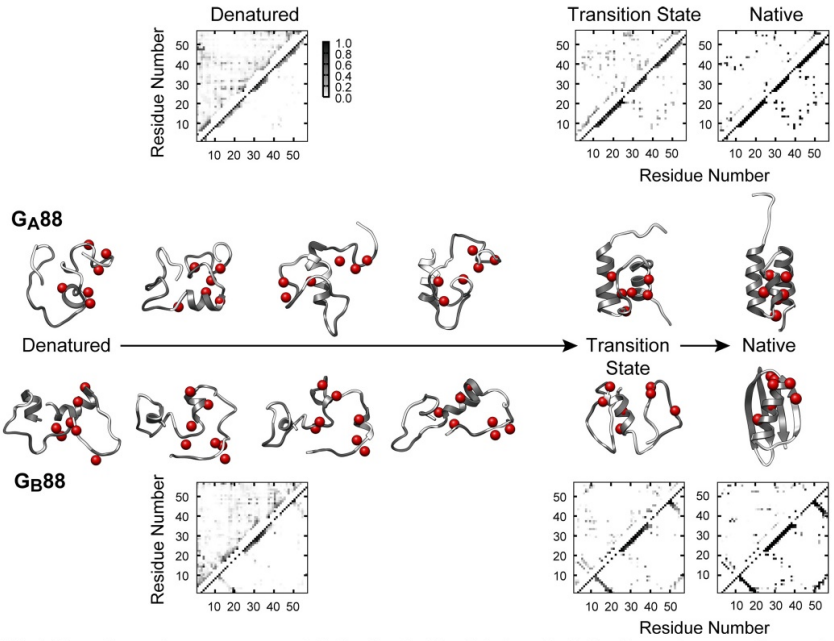


FIGURE 8. **Folding pathway and contact maps at neutral pH.** G_A88 (above) and G_B88 (below) are colored *white* with secondary structure elements colored *dark gray* (G_A88, residues 9–23, 27–34, and 39–51; G_B88, residues 1–8, 13–20, 23–36, 42–46, and 51–55). Structures are from 50, 40, 30, and 20 ns. The transition state (G_A88, 0.486 ns; G_B88, 0.422 ns) of run 1 is at 498 K. The NMR structures are also shown to the *right*. Cα atoms of differing residues (24, 25, 30, 33, 45, 49, and 50) are shown as *red balls*. The fraction of time in contact for G_A88 (above) and G_B88 (below) were plotted with non-native contacts in the *top-left triangle* of each panel and with native contacts (present in the starting structure) in the *bottom-right* of each panel. Native contacts are reported for the full 30-ns 298 K simulation (n = 30,000). Transition state contacts are reported for all transition state ensembles (n = 30); denatured state contacts are plotted for the last 30 ns of the three long 498 K simulations (n = 90,000). Two residues were considered in contact if they contained carbon atoms that were ≤5.4 Å apart or any other non-hydrogen atoms that were ≤4.6 Å apart. Contacts were colored from *white* (never occurred) to *black* (present 100% of the time).

and simulation. The key findings show that both engineered proteins appear to fold via a two-state mechanism, and protein topology is committed very early along the folding pathway.

Understanding the Mechanism by Which Proteins Commit to Their Topology Highlights the Role of Long Range Interactions in Denatured States—For the purpose of this study, it is essential to understand the mechanism whereby a few key residues univocally determine native topology, *i.e.* which structural determinants preclude the sequence of G_A88 from adopting the G_B88 topology and vice versa? Experimentally, we clearly detected a difference in denatured state properties of the two proteins. In fact, we observed the m_{D-N} value of G_B88 to decrease with decreasing pH (Fig. 5*B*). This observation is consistent with the MD simulations suggesting the denatured state of G_B88 to be more structured at acidic than neutral pH, as mirrored both by its native helical content and by its solvent-accessible surface area (Table 1). Importantly, such a dependence was not observed in G_A88, whose m_{D-N} value was found experimentally to be insensitive to pH. In summary, although only 7 of the 56 amino acids are different between G_A88 and G_B88, only the latter displays in its denatured state a detectable residual structure. Such a structure may be tuned by changing pH, as reflected both by analysis of m_{D-N} values as a function of pH (apparent pK_a ~5) and by comparison of the MD simulations at neutral and low pH. Surprisingly, the residues that are different between the two proteins (Fig. 1) do not include amino acids titrating below neutral pH, suggesting that the observed compaction of the denatured state of G_B88 originates from non-local effects. Indeed, the MD simulations of G_B88 highlight the presence of side chain hydrogen bonds in the $\beta3/\beta4$ hairpin turn, in the denatured state at neutral pH. These hydrogen bonds involve residue Asp^{47}, which is protonated in the low pH simulations and does not form hydrogen bonds with either Tyr^{45} or Lys^{50} as it does at neutral pH. Interestingly, residues 45 and 50 are

both mutated to Leu in G_A88 and are part of the $\alpha3$ helix. However, Asp^{47} along with Glu^{48} tends to interact with Thr^1 in the denatured state of G_A88. This interaction favored backbone dihedral angles, which were compatible with α-helix rather than the $\beta3/\beta4$ turn, as in G_B88.

According to AGADIR (37) and our MD simulations, the 7-residue difference in sequence between G_A88 and G_B88 yields an increased helical propensity for G_B88. In particular, the 4 residues forming the loop connecting $\alpha1$ and $\alpha2$ in G_A88 adopt a helical conformation in the structure of native G_B88. Furthermore, MD reveals that substitution of Gly to Ala at residue 24 and Ile to Thr at residue 25 increases the helical propensity of the region. Thus, experiments and simulations converge in supporting the hypothesis that a longer, more stable α-helix in G_B88 prevents the latter sequence from folding to the G_A88 structure. Interestingly, it was recently shown that a switch between the G_A and G_B structures may be obtained even with a single amino acid substitution (38). Under conditions where the G_A fold is >90% populated, mutation of Leu^{45} into Tyr shifts the population to >90% of the G_B fold. Surprisingly, position 45 is not in the loop where G_A88 and G_B88 display a different helical propensity (Fig. 1), indicating that the greater helical content of the denatured state of G_B88 is affected by long range interactions.

Overall, comparison of the folding of G_A88 and G_B88 highlights a conundrum; although only a few residues (or even a single one) are responsible for the selective stabilization of the two alternative topologies, information on the folding mechanism indicates that no single residue appears to act as a unique gatekeeper in the selection of protein topology. Both experiments and simulations on the folding of G_A88 and G_B88 suggest that native topology might be presculpted in the denatured state, where incipient nuclei are present. Stabilization of such nuclei is affected by long range interactions, and commitment to the native fold occurs by selective stabilization of these incipient nuclei rather than by actively blocking alternative pathways.

Do Engineered Proteins Display Cooperative Folding?—An intriguing general question is whether folding is under evolutionary pressure. This issue was recently discussed by Baker and co-workers (39) in a study on the folding of a *de novo* designed protein, Top7, characterized by a novel non-natural topology. Although most small naturally occurring proteins fold in a cooperative mode (40), Top7 displays a non-cooperative folding mechanism, suggesting that cooperativity may be a result of natural selection. In this context, it may seem somewhat puzzling that both G_A88 and G_B88 fold in a cooperative manner, displaying single exponential folding and unfolding kinetics as well as V-shaped chevron plots, a hallmark of two-state folding (22). However, a possible explanation to reconcile these apparently conflicting observations is that although the structure of Top7 is non-natural (41), both G_A88 and G_B88 have been engineered starting from naturally occurring frameworks (16).

It was reported recently that malleability of protein folding pathways stems from the existence of multiple nuclei within a given protein structure (42). Implicit in this view, a natural topology would contain one or more nucleation motifs, representing the minimal units to encode for the final structure (43). Nucleation of the motif is the basis for cooperative folding, and the number of accessible pathways is related to the number of nucleation motifs within a protein (44). On the basis of the folding pathways observed for Top7, G_A88, and G_B88, it is tempting to speculate that because naturally occurring topologies contain nucleation motifs, productive folding of these substructures would result in cooperative transitions. In contrast, non-natural proteins may not contain stable nucleation motifs and, thus, they appear to fold non-cooperatively. We may conclude that evolution does not *directly* select for cooperative folding, but rather it selects for topologies that can fold in a cooperative manner.

Comparison with Studies on Protein Families—The study of homologous proteins has represented a powerful approach to obtain insight into protein folding (3–11, 34, 45, 46), especially when combined with structural information on intermediate or transition states. In a recent study, we addressed the structural features of the early and late transition states of two homologous three-state proteins, PSD-95 PDZ3 and PTP-BL PDZ2 (3). For different PDZ domains, we observed that the late folding transition states (*TS2*) are more similar to each other than the early transition states (*TS1*). This observation would suggest that although native topology defines the late stages of folding in a unique way, significant freedom in creating structural contacts is observed for the early events. In this perspective, it is of interest to compare the cases of G_A88/G_B88, whose folding appears to diverge early in the denatured state, with that of the PDZ family, where a strong native bias is seen only at the late stages of folding. A plausible scenario to reconcile these apparently contrasting results would imply the presence of multiple nucleation motifs at the early stages of PDZ folding. In fact, the apparent structural divergence of *TS1* for PDZ2 and PDZ3 most likely arises from selective stabilization of alternative nuclei in the two denatured proteins, which may then appear to explore distinct early folding pathways. When folding proceeds to the native state, the alternative nuclei all consolidate in a native-like conformation, and folding pathways appear to converge. Support for this hypothesis comes from circular permutation experiments on PDZ2, whereby alternative nuclei may be selectively stabilized and alter the early events of folding without affecting the late ones (47, 48). On the other hand, G_A88 and G_B88 do not appear to contain alternative nuclei; moreover, because they display two completely different structures, their respective nucleation motifs may be completely independent such that their folding pathways diverge from the very early stages. Future work based on protein engineering experiments, MD simulations, and ϕ-value analysis (23) will further address structural determinants in the folding of this heteromorphic protein pair in an effort to identify crucial residues for the stabilization of these two alternative topologies and their respective folding motifs and to extend the experimental and theoretical work to more complex multidomain systems.

REFERENCES

1. Editorial (2005) *Science* **309,** 78–102
2. Fersht, A. R. (2008) *Nat. Rev. Mol. Cell Biol.* **9,** 650–654

3. Calosci, N., Chi, C. N., Richter, B., Camilloni, C., Engström, A, Eklund, L., Travaglini-Allocatelli, C., Gianni, S., Vendruscolo, M., and Jemth, P. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 19241–19246
4. Chi, C. N., Gianni, S., Calosci, N., Travaglini-Allocatelli, C., Engström, K., and Jemth, P. (2007) *FEBS Lett.* **581,** 1109–1113
5. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M., and Dobson, C. M. (1999) *Nat. Struct. Biol.* **6,** 1005–1009
6. Clarke, J., Cota, E., Fowler, S. B., and Hamill, S. J. (1999) *Structure* **7,** 1145–1153
7. Friel, C. T., Capaldi, A. P., and Radford, S. E. (2003) *J. Mol. Biol.* 293–305
8. Martínez, J. C., and Serrano, L. (1999) *Nat. Struct. Biol.* **6,** 1010–1016
9. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999) *Nat. Struct. Biol.* **6,** 1016–1024
10. Travaglini-Allocatelli, C., Gianni, S., Dubey, V. K., Borgia, A., Di Matteo, A., Bonivento, D., Cutruzzolà, F., Bren, K. L., and Brunori, M. (2005) *J. Biol. Chem.* **280,** 25729–25734
11. Travaglini-Allocatelli, C., Gianni, S., Morea, V., Tramontano, A., Soulimane, T., and Brunori, M. (2003) *J. Biol. Chem.* **278,** 41136–41140
12. Wilson, C. A., Kreychman, J., and Gerstein, M. (2000) *J. Mol. Biol.* **297,** 233–249
13. Rose, G. D., and Creamer, T. P. (1994) *Proteins* **19,** 1–3
14. Dalal, S., Balasubramanian, S., and Regan, L. (1997) *Nat. Struct. Biol.* **4,** 548–552
15. Davidson, A. R. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 2759–2760
16. Alexander, P. A., He, Y., Chen, Y., Orban, J., and Bryan, P. N. (2007) *Proc. Natl. Acad. Sci. U.S.A.* **104,** 11963–11968
17. He, Y., Chen, Y., Alexander, P., Bryan, P. N., and Orban, J. (2008) *Proc. Natl. Acad. Sci. U.S.A.* **105,** 14412–14417
18. Scott, K. A., and Daggett., V. (2007) *Biochemistry* **46,** 1545–1556
19. Ruan, B., Fisher, K. E., Alexander, P. A., Doroshko, V., and Bryan, P. N. (2004) *Biochemistry* **43,** 14539–14546
20. Silow, M., and Oliveberg, M. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94,** 6084–6086
21. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry* **27,** 8063–8068
22. Jackson, S. E., Fersht, A. R. (1991) *Biochemistry* **30,** 10428–10435
23. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) *J. Mol. Biol.* **224,** 771–782
24. Levitt, M., Hirshberg, M., Sharon, R., and Daggett, V. (1995) *Comput. Phys. Commun.* **91,** 215–231
25. Levitt, M., Hirshberg, M., Sharon, R., Laidig, K. E., and Daggett, V. (1997) *J. Phys. Chem. B* **101,** 5051–5061
26. Beck, D. A., and Daggett, V. (2004) *Methods* **34,** 112–120
27. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22,** 2577–2637
28. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400
29. Li, A., and Daggett, V. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91,** 10430–10434
30. Li, A., and Daggett, V. (1996) *J. Mol. Biol.* **257,** 412–429
31. Geierhaas, C. D., Nickson, A. A., Lindorff-Larsen, K., Clarke, J., and Vendruscolo, M. (2007) *Protein Sci.* **16,** 125–134
32. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M., Alonso, D. O., Daggett, V., and Fersht., A. R. (2003) *Nature* **421,** 863–867
33. Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M., and Fersht, A. R. (2005) *Nature* **437,** 1053–1056
34. White, G. W., Gianni, S., Grossmann, J. G., Jemth, P., Fersht, A. R., and Daggett, V. (2005) *J. Mol. Biol.* **350,** 757–775
35. Sánchez, I. E., and Kiefhaber, T. (2003) *J. Mol. Biol.* **325,** 367–376
36. Timasheff, S. N. (1993) *Annu. Rev. Biophys. Biomol. Struct.* **22,** 67–97
37. Muñoz, V., and Serrano, L. (1997) *Biopolymers* **41,** 495–509
38. Alexander, P. A., He, Y., Chen, Y., Orban, J., and Bryan, P. N. (2009) *Proc. Natl. Acad. Sci. U.S.A.* **106,** 21149–21154
39. Watters, A. L., Deka, P., Corrent, C., Callender, D., Varani, G., Sosnick, T., and Baker, D. (2007) *Cell* **128,** 613–624
40. Tanford, C. (1970) *Adv. Protein Chem.* **24,** 1–95
41. Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., and Baker, D. (2003) *Science* **302,** 1364–1368
42. Lindberg, M. O., and Oliveberg, M. (2007) *Curr. Opin. Struct. Biol.* **17,** 21–29
43. Hubner, I. A., Lindberg, M., Haglund, E., Oliveberg, M., and Shakhnovich, E. I. (2006) *J. Mol. Biol.* **359,** 1075–1085
44. Haglund, E., Lindberg, M. O., and Oliveberg, M. (2008) *J. Biol. Chem.* **283,** 27904–27915
45. McCallister, E. L., Alm, E., and Baker, D. (2000) *Nat. Struct. Biol.* **7,** 669–673
46. Wensley, B. G., Gärtner, M., Choo, W. X., Batey, S., and Clarke, J. (2009) *J. Mol. Biol.* **390,** 1074–1085
47. Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M., and Gianni, S. (2009) *J. Am. Chem. Soc.* **131,** 11727–11733
48. Ivarsson, Y., Travaglini-Allocatelli, C., Morea, V., Brunori, M., and Gianni, S. (2008) *Protein Eng. Des. Sel.* **21,** 155–160

Please cite this article in press as: Morrone et al., GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate, Biophysical Journal (2011), doi:10.1016/j.bpj.2011.09.013

Biophysical Journal Volume 101 October 2011 1–8

GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate

Angela Morrone,[†△] Rajanish Giri,[†△] Rudesh D. Toofanny,[‡] Carlo Travaglini-Allocatelli,[†] Maurizio Brunori,[†] Valerie Daggett,[‡*] and Stefano Gianni[†*]

[†]Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università di Roma "La Sapienza", Rome, Italy; and [‡]Department of Bioengineering, University of Washington, Seattle, Washington

ABSTRACT The folding pathway of the small α/β protein GB1 has been extensively studied during the past two decades using both theoretical and experimental approaches. These studies provided a consensus view that the protein folds in a two-state manner. Here, we reassessed the folding of GB1, both by experiments and simulations, and detected the presence of an on-pathway intermediate. This intermediate has eluded earlier experimental characterization and is distinct from the collapsed state previously identified using ultrarapid mixing. Failure to identify the presence of an intermediate affects some of the conclusions that have been drawn for GB1, a popular model for protein folding studies.

INTRODUCTION

Understanding the role and structure of partially folded intermediates is of fundamental mechanistic importance for protein folding studies. Earlier work suggested that the folding of small single domain proteins generally conforms to an all-or-none behavior (1), involving simultaneous formation of secondary and tertiary structure (2). Following this view, folding occurs in a two-state fashion, via condensation around a marginally stable nucleus, and discrete intermediates tend to be avoided (3). When the inherent stability of folding nuclei is increased, however, even very simple protein systems appear to fold in a more complex fashion (4), with population of partially folded intermediates, which may either transiently accumulate leading to multiphasic kinetics, or be a high energy species en route to the native state (5). The presence of such local minima in the landscape is very difficult to address experimentally (6) and intermediates may sometimes escape detection.

The B1 IgG-binding domain of streptococcal protein G, generally called GB1, has played a central role in protein folding studies being the system of choice in more than 200 works carried out using a wide variety of experimental and theoretical approaches; see for example (7–19). Because of its small size and its simple and highly symmetrical topology, this small protein domain has represented an ideal candidate for a vast number of different studies. Over-and-above contrasting views on the presence of a low energy early collapsed state as detected by ultrarapid mixing (13,15,16), experimental work has been taken to indicate that GB1 folds in a canonical two-state process (20), via a polarized folding transition state, with native-like structure localized in the C-terminal β-hairpin (14). On the other hand, despite the evidence arising from experiments, some theoreticians have predicted the presence of intermediates and heterogeneous pathways for the folding of GB1 (19), raising some doubts about the applicability of a bona fide two-state mechanism. In this study, we have undertaken an extensive characterization of the folding of GB1 by experiments and simulations that provide evidence for a folding intermediate. This partially structured state is an on-pathway species that, despite two decades of studies, escaped detection and characterization.

MATERIALS AND METHODS

The buffers used were 50 mM Glycine/NaOH from pH 9.6 to 9.0, 50 mM Tris/HCl from pH 9.0 to 7.2, 50 mM Bis-Tris/HCl from pH 7.0 to 6.0, 50 mM sodium acetate from pH 5.5 to 3.8, 50 mM sodium formate from pH 3.4 to 3.0, and 50 mM sodium phosphate/phosphoric acid from pH 2.8 to 2.0. All reagents were of analytical grade.

Protein expression and purification

GB1 gene was cloned into the vector pG58, which encodes an engineered subtilisin prosequence as the N-terminus of the fusion protein. Proteins were purified employing an affinity chromatography previously developed (21). Soluble cell extract of prodomain fusion proteins was injected onto a 5-ml Bio-Scale Mini Profinity eXact cartridge (Bio-Rad, Hercules, CA) at 5 ml/min to allow binding and then washed with 10-column volumes of 100 mM $NaPO_4$ (pH 7.2) to remove impurities. To cleave and elute the purified target proteins, 15 ml of 100 mM NaF in the presence of 100 mM $NaPO_4$ (pH 7.2) were injected at 5 ml/min. After the first 10 ml, the flow was stopped and the column incubated for 30 min to allow complete cleavage. After elution, the purified proteins were then dialyzed to remove sodium fluoride.

Equilibrium unfolding

Equilibrium denaturations were carried out on a Fluoromax single photon counting spectrofluorometer (Jobin-Yvon, Edison, NJ). Tryptophan

Submitted July 22, 2011, and accepted for publication August 19, 2011.

△Angela Morrone and Rajanish Giri contributed equally to this work.

*Correspondence: daggett@uw.edu or stefano.gianni@uniroma1.it

This article is dedicated to the late Professor Quentin H. Gibson, the inventor of the stopped flow spectrophotometer in 1954, who passed away on March 16, 2011.

Editor: Doug Barrick.

0006-3495/11/10/0001/8 $2.00

doi: 10.1016/j.bpj.2011.09.013

Please cite this article in press as: Morrone et al., GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate, Biophysical Journal (2011), doi:10.1016/j.bpj.2011.09.013

fluorescence emission spectra were recorded in a cuvette (1-cm light path) between 300 and 400 nm. The excitation wavelength was 280 nm. Protein concentrations were typically 3 μM.

Stopped-flow measurements

Single mixing kinetic folding experiments were carried out on a Pi-star or on an SX-18 stopped-flow instrument (Applied Photophysics, Leatherhead, UK); the excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. In all experiments, performed at 25°C, refolding and unfolding were initiated by an 11-fold dilution of the denatured or the native protein with the appropriate buffer. Final protein concentrations were typically 1 μM. The observed kinetics were always independent of protein concentration (from 0.5 to 5 μM after mixing protein concentration), as expected from monomolecular reactions without effects due to transient aggregation (22).

Data analysis

Equilibrium experiments

Assuming a standard two-state model, the guanidine-induced denaturation transitions were fitted to the following equation:

$$\Delta G_d = m_{D-N}(D - D_{50}),$$

where m_{D-N} is the slope of the transition (proportional to the increase in solvent-accessible surface area on going from the native to the denatured state) and D_{50} is the midpoint of the denaturation transition. An equation, which takes into account the pre- and post-transition baselines, was used to fit the observed unfolding transition (23). Equilibrium folding stabilities calculated at different pH conditions are listed in Table 1.

Kinetic experiments

Analysis was performed by nonlinear least-squares fitting of single exponential phases using the fitting procedures provided in the Applied Photophysics software. The chevron plots were fitted globally by numerical analysis based on a three-state model as discussed in the Results section. The global fit of chevron plots at different pH was obtained with Prism software (Graphpad).

Molecular dynamics simulation methods

All-atom explicit solvent molecular dynamics (MD) simulations were performed for protein GB1 by starting with the NMR structure (PDB 3GB1 (24)). Simulations included one native state simulation at 25°C (21 ns) and five unfolding simulations at 225°C (1 × 41 ns, 1 × 38, 3 × 2 ns) resulting in a total of 106 ns of simulation time. All simulations were performed using our in-house MD software, *in lucem* molecular mechanics (25) with the Levitt et al. (26) all-atom force field. The microcanonical ensemble was used, where the number of atoms, unit cell volume, total energy, and linear momentum were conserved. The protein was solvated in a periodic box of flexible F3C water molecules (27). For a more in-depth discussion of the simulation protocols please see (28).

One-dimensional reaction coordinate

To identify the transition and intermediate states, we calculated a multidimensional property space derived from 15 physical properties of the protein and then calculated a one-dimensional (1D) reaction coordinate based on these 15 properties (29). The 15 properties were native contacts, nonnative contacts, radius of gyration, end-to-end distance, main-chain solvent accessible surface area (SASA) (30) side-chain SASA, polar SASA, nonpolar SASA, main-chain polar SASA, main-chain nonpolar SASA, side-chain polar, side-chain nonpolar SASA, total SASA, fraction of helix, and fraction of β-structure. The distance between two structures in property space is calculated as the average Euclidean distance between the 15-dimensional points. The mean distance in property space was calculated for each time point in a simulation of interest to the native state reference, which contained all the structures of the native state simulation excluding the first nanosecond. The 1D reaction coordinate was created from a histogram of the mean distance to reference for all structures. To compare with

TABLE 1 Folding parameters of GB1 as a function of pH

pH	k_{DI} (s^{-1})	$k_{NI}/(1 + K_{part})$ (s^{-1})	K_{part}	$\Delta G_{D\text{-}N}$* (kcal mol^{-1})	$\Delta G_{D\text{-}N}^{\dagger}$ (kcal mol^{-1})
2.0	1000 ± 90	11.0 ± 3.0	1.7 ± 0.2	2.7 ± 0.3	3.9 ± 0.8
2.5	1000 ± 90	19.2 ± 4.8	3.0 ± 0.4	2.3 ± 0.2	3.0 ± 0.7
3.0	1050 ± 95	5.2 ±1.3	0.9 ±0.1	3.1 ± 0.4	3.8 ± 0.4
3.5	1200 ± 120	5.1 ±1.3	1.2 ±0.1	3.2 ± 0.2	3.9 ± 0.4
4.0	1400 ± 120	1.3 ± 0.3	0.42 ± 0.09	4.1± 0.2	4.5 ± 0.4
4.5	1800 ± 170	0.4 ± 0.1	0.20 ± 0.04	5.0 ± 0.2	4.2 ± 0.4
5.0	1300 ± 120	0.15 ± 0.04	0.09 ± 0.02	5.3 ± 0.2	4.9 ± 0.5
5.5	1370 ± 100	0.15 ± 0.04	0.13 ± 0.03	5.4 ± 0.1	4.6 ± 0.5
6.0	720 ± 70	0.12 ± 0.03	0.09 ± 0.02	5.1 ± 0.1	5.2 ± 0.5
6.5	830 ± 70	0.14 ± 0.04	0.10 ± 0.02	5.1 ± 0.2	4.7 ± 0.5
7.0	670 ± 70	0.19 ± 0.05	0.12 ± 0.02	4.8 ± 0.2	5.3 ± 0.5
7.5	630 ± 70	0.27 ± 0.07	0.11 ± 0.02	4.6 ± 0.2	4.3 ± 0.4
8.0	600 ± 60	0.4 ± 0.1	0.12 ± 0.03	4.3 ± 0.3	4.3 ± 0.4
8.5	500 ± 40	0.6 ± 0.2	0.11 ± 0.02	3.9 ± 0.2	3.8 ± 0.4
9.0	390 ± 40	0.9 ± 0.2	0.13 ± 0.03	3.6 ± 0.2	3.3 ± 0.3
9.6	220 ± 25	1.3 ± 0.3	0.11 ± 0.02	3.0 ± 0.2	3.0 ± 0.3

*Calculated from chevron plot analysis. The Chevron plots were fitted globally to a three-state model with shared *m*-values. k_{DI} is the microscopic rate constant for the formation of the intermediate from the denatured state; k_{NI} is the microscopic rate constant for the unfolding of the native state to the intermediate state; K_{part} is the partitioning factor k_{ID}/k_{IN} reflecting the difference between the activation barriers for the intermediate to revert to the reagents rather than proceeding to the products. The analysis returned a total $m_{D\text{-}N} = 1.95 \pm 0.2$ kcal mol^{-1} M^{-1}. The Tanford β-values for the two transition states were $\beta_{TS1} = 0.76 \pm 0.04$ and $\beta_{TS2} = 0.93 \pm 0.04$.

†Calculated from equilibrium denaturation. Equilibrium denaturations were fitted both individually and globally with shared $m_{D\text{-}N}$ value. The global analysis returned $m_{D\text{-}N} = 1.75 \pm 0.2$ kcal mol^{-1} M^{-1}. This value was consistent within error with the values obtained by fitting individually each independent equilibrium experiments, as well as with the value calculated from kinetic experiments.

Please cite this article in press as: Morrone et al., GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate, Biophysical Journal (2011), doi:10.1016/j.bpj.2011.09.013

experimental Tanford β-values, we calculated the ratio of the average total (SASA) for the transition state ensemble and the average total SASA for the native state simulation.

To investigate the unfolding pathway of GB1 we calculated contact matrices, for each state identified, based on the fraction of time that the residues were in contact. A pair of residues was considered in contact if it contained carbon atoms that were <5.4 Å apart or one carbon atom or any other atoms <4.6 Å.

RESULTS

Equilibrium denaturation of GB1

To study the folding mechanism of GB1, we carried out both equilibrium and kinetic experiments. Guanidine-induced equilibrium denaturation of GB1, monitored by fluorescence spectroscopy, was accomplished at 25°C exploring a wide range of pH values, from 2.0 to 9.6 (data not shown). Equilibrium denaturation curves were fitted both individually and globally with a shared $m_{D\text{-}N}$ value. Values obtained from the global analysis, listed in Table 1, were consistent within error with the values obtained by fitting each independent equilibrium experiment individually, as well as with the value calculated from kinetic experiments, confirming the two-state nature of the equilibrium unfolding transition of GB1. The denaturation profiles were all consistent with two-state unfolding and returned an m-value of 1.75 ± 0.2 kcal mol^{-1} M^{-1}.

Kinetic experiments

The folding and unfolding kinetics of GB1 were investigated at several pH values, ranging from 2.0 to 9.6. In all cases, folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final guanidine concentration. Each rate constant was obtained from the average of at least five independent shots in stopped-flow experiments. Semilogarithmic plots of the observed folding/unfolding rate constants of GB1 versus denaturant concentration (chevron plot) at the different pH values are presented in Fig. 1. Surprisingly, the unfolding arm of the chevron plots at pH values higher than 6.0 shows a deviation from linearity that becomes evident at high guanidine concentrations (rollover effect). This deviation is highlighted in Fig. 2 where the chevron plot of GB1 measured at pH 9.0 is reported together with the residuals of the fit, showing a clear systematic deviation from the expected values for a two-state behavior. This effect escaped previous studies probably because of the restricted range of experimental conditions, limited to [GdnHCl] < 5.5 M (14,15). Indeed, if we were to ignore the data we recorded for [GdnHCl] > 5.5 M, the unfolding arm of the chevron plots would appear essentially linear but would display a puzzling change in slope with the different experimental conditions.

Analysis of chevron plots is a common and powerful tool for detecting protein folding intermediates (31). In fact, if there is only one rate-limiting energy barrier, the logarithm

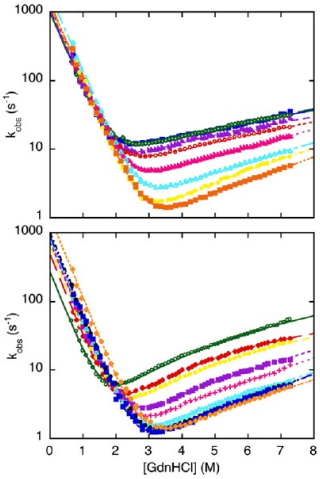


FIGURE 1 Folding kinetics of GB1. Top panel: Chevron plots measured from pH 2.0 to pH 5.5 at 25°C (*blue*, pH 2.0; *green*, pH 2.5; *violet*, pH 3.0; *red*, pH 3.5; *magenta*, pH 4.0; *cyan*, pH 4.5; *yellow*, pH 5.0; *orange*, pH 5.5). Bottom panel: Chevron plots measured from pH 5.5 to pH 9.6 at 25°C (*orange*, pH 5.5; *blue*, pH 6.0; *black*, pH 6.5; *cyan*, pH 7.0; *magenta*, pH 7.5; *violet*, pH 8.0; *yellow*, pH 8.5; *red*, pH 9.0; *green*, pH 9.6). Lines are the best global fit to a three-state equation with shared m-values. If data at [GdnHCl] > 5.5 M were to be ignored, a quasilinear unfolding arm with an apparent change in unfolding m-values would be seen.

Web 3C

of the observed folding and unfolding kinetics is expected to return a V-shaped dependence (32). Therefore, a deviation from linearity in either the folding or the unfolding branches of the curve may be considered of diagnostic value for the identification of intermediates (33,34). If a partially folded intermediate is present, the folding kinetics can be described by a three-state mechanism:

$$\mathrm{D} \underset{k_{ID}}{\overset{k_{DI}}{\rightleftharpoons}} \mathrm{I} \underset{k_{NI}}{\overset{k_{IN}}{\rightleftharpoons}} \mathrm{N}, \qquad \text{(Scheme 1)}$$

where k_{DI} is the microscopic rate constant for the formation of the intermediate from the denatured state, k_{ID} is the microscopic rate constant for the unfolding of the intermediate to the denatured state, k_{IN} is the microscopic rate constant for the formation of the native state from the intermediate, and k_{NI} is the microscopic rate constant for the unfolding of the native state to the intermediate state. Two approximations have been introduced to describe the folding pathway of three-state systems. The intermediate may be assumed to be in a fast preequilibrium with one of the

Please cite this article in press as: Morrone et al., GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate, Biophysical Journal (2011), doi:10.1016/j.bpj.2011.09.013

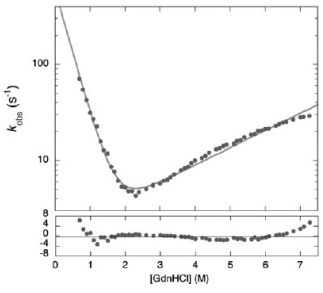


FIGURE 2 Chevron plot of wild-type GB1 measured at pH 9.0. The gray line is the best fit to a two-state equation; the residuals of the fit below show a clear systematic deviation from the expected values for a two-state behavior. Therefore, data at [GdnHCl] > 5.5 M are critical to detect the roll-over effect.

ground states, the curvature in the chevron plot being caused by accumulation of the intermediate (33). Alternatively, if the intermediate is assumed to be at steady state and at negligible concentration, the curvature in the observed (un) folding kinetics arises from a change in the rate-limiting step with changing denaturant concentration (34–38). Both approximations lead to nearly identical numerical solutions and are thus indistinguishable based only on the analysis of rate constants. In theory, however, accumulation of the intermediate should lead to multiphasic observed kinetics and more than one relaxation rate constant should be observed. In the case of GB1, because we observed single exponential folding kinetics under all conditions explored, we favored a model involving a high-energy on-pathway intermediate and the observed chevron plots were fitted to the following equation:

$$k_{obs} = k_{DI} + \frac{k_{NI}}{1 + K_{part}},$$

where K_{part} is the partition factor k_{ID}/k_{IN} proportional to the difference between the activation barriers for the intermediate state to revert to the reagents rather than proceeding to the products. The logarithm of each microscopic rate constant was assumed to vary linearly with denaturant concentration (the slope of each dependence yielding the corresponding *m*-value). The observed chevron plots were fitted globally with shared *m*-values. Parameters calculated from global analysis (listed in Table 1) allow the identification of the relative positions of the two activation barriers along the reaction coordinate in terms of their relative accessible surface area (Tanford β-value), resulting in a β_T-value of 0.76 ± 0.04 for the transition state TS1 and 0.93 ± 0.04 for the native-like activation barrier TS2. The excellent statistical parameters of the global analysis indicate that the two activation barriers are robust to changes in pH conditions and display a conserved solvent accessible surface area when pH and protein stability are altered.

The effect of pH on the folding kinetics of GB1

The analysis of the chevron plots reported in Fig. 1 allowed determination of the folding and unfolding rate constants of GB1 over a very wide range of pH. The folding rate constants display a negligible dependence on pH (data not shown). A plot of the logarithm of the apparent unfolding rate constants from the native state to TS2 (k_{NI}) and TS1 (formally equivalent to $k_{NI}/(1 + K_{part})$) as a function of pH are reported in Fig. 3. Interestingly, both TS1 and TS2 display sigmoidal transitions at acidic and alkaline pH consistent with protonation of at least two groups in the native state with apparent pK_a values of ~4 and ~8. Importantly, the acid transition for the unfolding rate constant of TS1 (changing by almost two orders of magnitude) is more pronounced than that for the unfolding rate constant of TS2 (changing by less than an order of magnitude), suggesting the contribution of a salt bridge that is weak or not formed in TS1, but is consolidated in TS2. As detailed in the Discussion section, inspection of the three-dimensional structure of GB1 and the unfolding profiles obtained by MD simulations suggest the acidic group to be either Glu-15 or Glu-56.

MD simulations

To further explore the putative intermediate state, all-atom explicit solvent MD simulations of GB1 were carried out. Multiple simulations were performed to model thermal unfolding (at 225°C) as well as native-state behavior at 25°C (control). The transition and intermediate states were

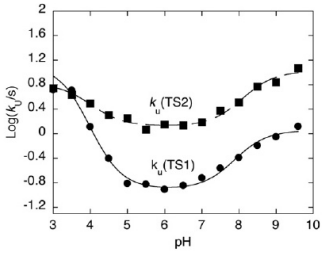


FIGURE 3 Unfolding rate constants versus pH. Logarithm of the calculated unfolding rate constants from the native state N to the first transition state k_u(TS1), and from the native state to the second transition state k_u(TS2) as a function of pH. The lines are the best fit to a model involving the protonation of two groups. In both cases, we obtained approximately the same pK_a of ~4 and ~8.

Please cite this article in press as: Morrone et al., GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate, Biophysical Journal (2011), doi:10.1016/j.bpj.2011.09.013

identified by calculating a multidimensional property space derived from 15 physical properties of the protein and then embedding this within a 1D reaction coordinate based on the 15 properties (as detailed in Materials and Methods) (Fig. 4). Five independent thermal unfolding simulations were performed; all simulations passed through the intermediate state (Fig. 5) however only two of them (runs 2 and 4) showed the buildup of an intermediate as defined by the 1D reaction coordinate in Fig. 4 and thus warranted further analysis. The native, transition, intermediate, and denatured states were identified for the two simulations displaying buildup of the intermediate. The calculated Tanford β_T-values based on simulations, in particular on total SASA, were 0.94 and 0.92 for TS2 in simulations 2 and 4, respectively, in agreement with the experimentally derived value of 0.93 $\pm$ 0.04. For TS1 the calculated β_T-values were 0.73 and 0.75 for runs 2 and 4, respectively, in agreement with the experimental value of 0.76 $\pm$ 0.04.

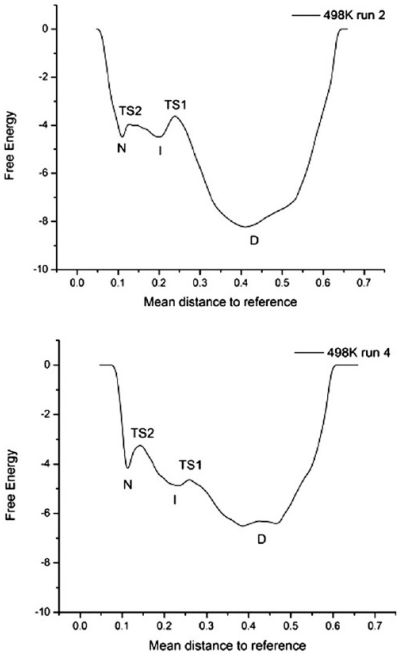


FIGURE 4 Free energy reaction coordinate for two unfolding simulations of GB1. A 1D reaction coordinate is created by calculating the mean distance to reference in a 15-dimensional property space for all structures in an unfolding simulation. The free energy reaction coordinates are calculated by taking the negative log of the counts in a mean distance to the reference histogram. Each 498 K simulation is shown with a black line. The location of unfolding states D is indicated on each free energy reaction coordinate. Top: Run 2 shows three-state unfolding with TS2 located between 0.12 and 0.13 mean distance to reference and TS1 located between 0.23 and 0.25 mean distance to reference. Bottom: Run 4 also shows three-state unfolding with TS2 located between 0.135 and 0.15 mean distance to reference and TS1 was located between 0.25 and 0.265 mean distance to reference.

Considering the unfolding runs in reverse, the simulated folding pathways were largely conserved in the different simulations (Fig. 6), the main difference originating from formation of nonnative contacts in the loop between strands A and B. Starting in the denatured state nonnative contacts dominated the contact matrix and there was residual helix, which gave rise to native contacts. Considering TS1, contacts began to form between strands A and B, strands A and D, and also strands C and D; the helix was nearly fully formed. The intermediate contained contacts between strand B and the helix. There were still native contacts between strands A and B. The N-terminal ends of strand A formed contacts with strand D, although many were short-term nonnative contacts. Contacts also formed between strand C and the helix as the latter began to move toward the protein core. Contacts between strands C and D were present and almost fully formed. From I to TS2 the protein gained native contacts between the helix and strand A that allowed the helix to dock to the core. Contacts between strands A and B, strands C and D, and strands A and D became almost fully native-like. In moving from TS2 to the native state, the helix docked to the core of the protein

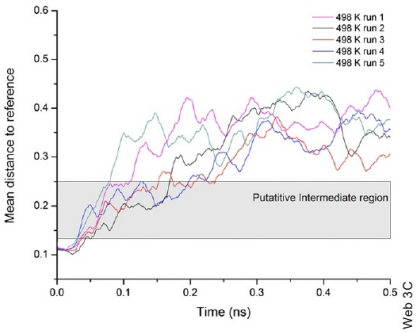


FIGURE 5 Mean distance to reference versus time for each unfolding simulation at 498 K. Lines have been smoothed using adjacent averaging to remove noise. The intermediate state region (0.135 to 0.15) in property space was determined from the positions of TS1 and TS2 in the free energy reaction coordinates from Fig. 4 for runs 2 and 4. It is clear that runs 2 and 4 populate this region significantly while the others pass through the intermediate more rapidly.

Please cite this article in press as: Morrone et al., GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate, Biophysical Journal (2011), doi:10.1016/j.bpj.2011.09.013

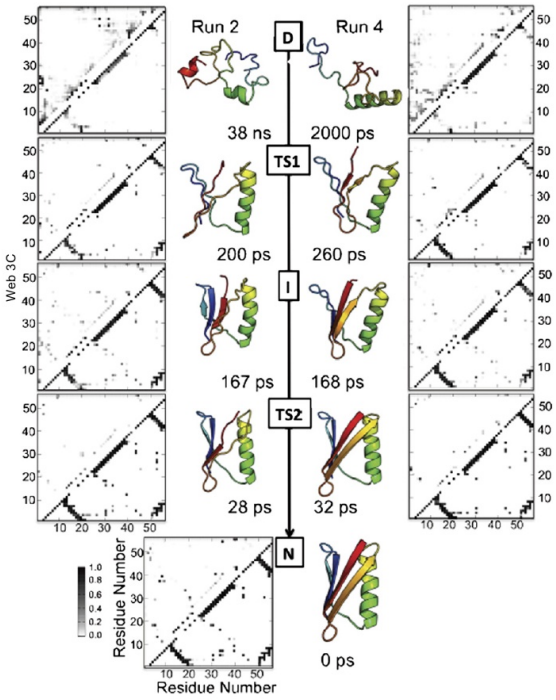


FIGURE 6 Folding pathway and contact maps for two unfolding simulations of GB1. Representative structures from each state are colored in rainbow from red to blue. The native structure of GB1 contains four β-strands (A = *red*, B = *orange*, C = *cyan*, and D = *blue*) and a helix . Contact maps show the fraction of time residues were in contact with nonnative interactions (*above the diagonal*) or with native contacts (*below*). Contacts are colored in a grayscale from white to black, where white indicates that the residues were never in contact and black those that were in contact 100% of the time. Native contacts are reported over the full 498 K simulation (n = 20,000). TS, I, and D contacts are reported for the time points constituting the ensemble of structures as determined by the free energy reaction coordinate (Fig. 5).

making contacts with strands A, B, and C. The contacts between strands A and B, strands C and D, and strands A and D became fully formed.

DISCUSSION

We have carried out an extensive characterization of the folding pathway of this protein by experiments and MD simulations. The kinetic experiments, carried out over a wide range of pH values (from 2.0 to 9.6), highlighted the presence of a rollover effect in the chevron plots, which is typically diagnostic of the presence of a folding intermediate. Such an intermediate was detected by unfolding MD simulations, which also allowed us to provide a structural characterization of the (un)folding pathway.

The existence of an intermediate in the folding of GB1 has been previously proposed by Roder and co-workers (15,16), who suggested the presence of a collapsed state accumulating in the μs time range, as observed by continuous-flow ultrarapid mixing experiments. This low-energy intermediate (which later was questioned by Sosnick and co-workers (13)) was also consistent with a detailed ab initio simulation reported by Kmiecik and Kolinski (39). Importantly, however, the partially folded state identified in our work is distinct from that previously suggested, being a high energy species that never accumulates; nevertheless, its existence suggests that GB1 folds via a complex and rough energy barrier with at least two discrete major transition states. Thus, while the equilibrium unfolding is consistent with a simple two-state model (7,8,20), folding and unfolding kinetics are more complex.

When folding is characterized by a complex chevron plot, the deviation from linearity observed in a chevron plot may have different origins: i), the curvature may be due to a movement of the position of the transition state along a single broad barrier (40,41); or ii), it may reflect a change

Please cite this article in press as: Morrone et al., GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate, Biophysical Journal (2011), doi:10.1016/j.bpj.2011.09.013

in the rate-limiting step suggestive of a multistate process (34,35,37,38). In the case of GB1, the curved chevron plots are observed only above pH 6 (Fig. 1), making the broad transition state model less likely because it would imply that it is possible to drastically distort the folding free energy profile and to switch between a narrow energy maximum (linear chevron plot) to a broad energy maximum (curved chevron plot) by changing pH. Furthermore, in analogy to what has been observed previously for other proteins (35,42,43), detection of a rollover effect only under some solvent conditions seems more consistent with a multistep folding pathway. The unbiased unfolding MD simulations reported in this work, which identified a folding intermediate, also supported this interpretation. The excellent statistical parameters of the global fit suggest that the two transition states surrounding the intermediate are relatively robust and maintain their overall structural features when solvent conditions are varied. This observation is in stark contrast with previous experimental work, which suggested the unfolding *m*-value to depend strongly on experimental conditions (14–16). We conclude that both TS1 and TS2 display a robust structure that is by and large maintained when solvent and/or sequence composition is altered.

It is of interest to analyze the dependence on pH of the unfolding rate constants measured for the two transition states (Fig. 3). Both energy barriers appear to display a monotonic transition at acidic and alkaline pH values. Both profiles are consistent with a model involving the protonation of two different groups with apparent pK_a values of ~4 and ~8 (Fig. 3). While in the alkaline region the total change in activation free energy is approximately the same for TS1 and TS2, the acidic transition is more pronounced in TS1, suggesting the presence of a charged interaction that is weak or not formed in TS1 but consolidated in TS2. Inspection of the three-dimensional structure of GB1 suggests that such a salt bridge may be either Lys-4-Glu-15 (located at the N-terminal β-hairpin, strands A and B) or Lys-10-Glu-56 (between the N-terminal turn and the C terminus of the protein). Both interactions would be consistent with MD simulations, suggesting that the contacts between strands A and B, strands C and D, and strands A and D are just incipient in TS1, whereas they are almost fully formed in TS2.

The thermal unfolding simulations are in good agreement with the experiments and reveal an on-pathway intermediate. The β_T-values for the transition states on either side of the intermediate ensemble are in excellent agreement with the experimentally determined ones. The structure of the intermediate is very native-like with reduced native contacts between β-strands and loss of contacts between the helix and the core of the protein, which facilitates undocking (Fig. 6). Importantly, although some studies have proposed that GB1 folds via a polarized transition state, with native-like structure located in the C-terminal hairpin (14,44), our simulations reported previously suggest folding to occur via a slightly different mechanism. In fact, structure formation appears to involve a more extended nucleus, which is stabilized by both the N- and C-terminal β-hairpins, as well as by contacts between the N- and C-terminal strands. This scenario suggests the presence of multiple overlapping folding nuclei (or a diffuse, rather than polarized, nucleus).

To conclude, we have reassessed the folding of GB1 by kinetic experiments and simulations and show that, while its equilibrium unfolding conforms to a two-state mechanism, its folding and unfolding kinetics are more complex and involve the presence of an energy barrier with at least two discrete transition states and one on-pathway intermediate. The structures of these transition states, which we have assessed by MD simulations, appear robust to changes in pH and are characterized by an extended nucleus, which is stabilized by both the N- and C-terminal β-hairpins, as well as by contacts between the N- and C-terminal strands.

This work was supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (RBRN07BMCT_007, to M.B.) and from the National Institutes of Health (GM50789 to V.D.).

REFERENCES

1. Fersht, A. R. 2008. From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat. Rev. Mol. Cell Biol.* 9:650–654.
2. Itzhaki, L. S., D. E. Otzen, and A. R. Fersht. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* 254:260–288.
3. Fersht, A. R. 1995. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA.* 92:10869–10873.
4. White, G. W., S. Gianni, …, V. Daggett. 2005. Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J. Mol. Biol.* 350:757–775.
5. Gianni, S., N. R. Guydosh, …, A. R. Fersht. 2003. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. USA.* 100:13286–13291.
6. Bryngelson, J. D., J. N. Onuchic, …, P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* 21:167–195.
7. Alexander, P., J. Orban, and P. Bryan. 1992. Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. *Biochemistry.* 31:7243–7248.
8. Chung, H. S., J. M. Louis, and W. A. Eaton. 2010. Distinguishing between protein dynamics and dye photophysics in single-molecule FRET experiments. *Biophys. J.* 98:696–706.
9. Ding, K., J. M. Louis, and A. M. Gronenborn. 2004. Insights into conformation and dynamics of protein GB1 during folding and unfolding by NMR. *J. Mol. Biol.* 335:1299–1307.
10. Frank, M. K., G. M. Clore, and A. M. Gronenborn. 1995. Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci.* 4:2605–2615.
11. Hubner, I. A., J. Shimada, and E. I. Shakhnovich. 2004. Commitment and nucleation in the protein G transition state. *J. Mol. Biol.* 336: 745–761.

BPJ 3151

Please cite this article in press as: Morrone et al., GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate, Biophysical Journal (2011), doi:10.1016/j.bpj.2011.09.013

12. Islam, S. A., M. Karplus, and D. L. Weaver. 2004. The role of sequence and structure in protein folding kinetics; the diffusion-collision model applied to proteins L and G. *Structure.* 12:1833–1845.
13. Krantz, B. A., L. Mayne, …, T. R. Sosnick. 2002. Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J. Mol. Biol.* 324:359–371.
14. McCallister, E. L., E. Alm, and D. Baker. 2000. Critical role of beta-hairpin formation in protein G folding. *Nat. Struct. Biol.* 7:669–673.
15. Park, S. H., K. T. O'Neil, and H. Roder. 1997. An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. *Biochemistry.* 36:14277–14283.
16. Park, S. H., M. C. Shastry, and H. Roder. 1999. Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nat. Struct. Biol.* 6:943–947.
17. Sheinerman, F. B., and C. L. Brooks, 3rd. 1997. A molecular dynamics simulation study of segment B1 of protein G. *Proteins.* 29:193–202.
18. Shen, T., C. Zong, J. J. Portman, and P. G. Wolynes. 2008. Variationally determined free energy profiles for structural models of proteins: characteristic temperatures for folding and trapping. *J. Phys. Chem. B.* 112:6074–6082.
19. Shimada, J., and E. I. Shakhnovich. 2002. The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl. Acad. Sci. USA.* 99:11175–11180.
20. Chung, H. S., J. M. Louis, and W. A. Eaton. 2009. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc. Natl. Acad. Sci. USA.* 106:11837–11844.
21. Alexander, P. A., Y. He, …, P. N. Bryan. 2007. The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc. Natl. Acad. Sci. USA.* 104:11963–11968.
22. Silow, M., and M. Oliveberg. 1997. Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci. USA.* 94:6084–6086.
23. Santoro, M. M., and D. W. Bolen. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry.* 27:8063–8068.
24. Kuszewski, J., A. M. Gronenborn, and G. M. Clore. 1999. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* 121:2337–2338.
25. Beck, D. A. C., D. O. Alonso, and V. Daggett. 2000–2011. *in lucem* Molecular Mechanics (*il*mm). University of Washington, Seattle, WA.
26. Levitt, M., M. Hirshberg, …, V. Daggett. 1995. Potential-energy function and parameters for simulations of the molecular-dynamics of proteins and nucleic-acids in solution. *Comput. Phys. Commun.* 91:215–231.
27. Levitt, M., M. Hirshberg, …, V. Daggett. 1997. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J. Phys. Chem. B.* 101:5051–5061.
28. Beck, D. A. C., and V. Daggett. 2004. Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods.* 34: 112–120.
29. Toofanny, R. D., A. L. Jonsson, and V. Daggett. 2010. A comprehensive multidimensional-embedded, one-dimensional reaction coordinate for protein unfolding/folding. *Biophys. J.* 98:2671–2681.
30. Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.
31. Gianni, S., Y. Ivarsson, …, C. Travaglini-Allocatelli. 2007. Identification and characterization of protein folding intermediates. *Biophys. Chem.* 128:105–113.
32. Jackson, S. E., and A. R. Fersht. 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry.* 30: 10428–10435.
33. Parker, M. J., J. Spencer, and A. R. Clarke. 1995. An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J. Mol. Biol.* 253:771–786.
34. Wildegger, G., and T. Kiefhaber. 1997. Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* 270:294–304.
35. Bachmann, A., and T. Kiefhaber. 2001. Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* 306:375–386.
36. Sánchez, I. E., and T. Kiefhaber. 2003. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* 325:367–376.
37. Sauder, J. M., N. E. MacKenzie, and H. Roder. 1996. Kinetic mechanism of folding and unfolding of Rhodobacter capsulatus cytochrome c2. *Biochemistry.* 35:16852–16862.
38. Walkenhorst, W. F., S. M. Green, and H. Roder. 1997. Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. *Biochemistry.* 36:5795–5805.
39. Kmiecik, S., and A. Kolinski. 2008. Folding pathway of the b1 domain of protein G explored by multiscale modeling. *Biophys. J.* 94:726–736.
40. Oliveberg, M. 1998. Alternative explanations for multi-state kinetics in protein folding: transient aggregation and changing transition-state ensambles. *Acc. Chem. Res.* 31:765–772.
41. Oliveberg, M., Y. J. Tan, …, A. R. Fersht. 1998. The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. *J. Mol. Biol.* 277:933–943.
42. Gianni, S., M. Brunori, …, M. Zhang. 2009. Distinguishing between smooth and rough free energy barriers in protein folding. *Biochemistry.* 48:11825–11830.
43. Gianni, S., N. Calosci, …, C. Travaglini-Allocatelli. 2005. Kinetic folding mechanism of PDZ2 from PTP-BL. *Protein Eng. Des. Sel.* 18:389–395.
44. Karanicolas, J., and C. L. Brooks, 3rd. 2002. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 11:2351–2361.

Bioch. Soc. Trans. [in press]

Morphogenesis of a protein: folding pathways and the energy landscape*

Maurizio Brunori, Stefano Gianni, Rajanish Giri, Angela Morrone and Carlo Travaglini-Allocatelli
[1]Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università di Roma "La Sapienza", Rome, Italy

Correspondence should be addressed to
Maurizio Brunori
Dipartimento di Scienze Biochimiche "A. Rossi Fanelli"
Università di Roma "La Sapienza"
P.le A.Moro, 5,
00185 ROME, Italy
Phone: +39 06 4450291; FAX: +39 06 4440062; email: maurizio.brunori@uniroma1.it

*We are pleased to dedicate this paper to our friend Professor Michael T. Wilson (Colchester, UK) on the occasion of the Symposium in his honor.

Introduction

Acquisition of a well defined and functionally competent 3D structure from a disordered polypeptide is a complex morphogenetic event which is basic to the life of a cell. After synthesis at the ribosome, the nascent polypeptide chain is exposed to many different stimuli and potential attacks, demanding an efficient and precise mechanism whereby the native state is quickly populated. Because for small proteins folding occurs spontaneously, it was concluded fifty years ago that the native structure is the most stable accessible conformation, i.e. it corresponds to the lowest free energy state [1]. It is astonishing how rapidly a polypeptide can find the correct conformation starting from a myriad of random coil structures. Given that folding cannot occur *via* a stochastic search among all possible conformations, the existence of a preferential pathway in the transition from the denatured to the native conformation has been classically invoked to account for the efficiency of folding [2]. Therefore, the traditional approach to protein folding focussed on the role of intermediates thought to be obligatory sequential states in the overall process.
Since the process whereby a protein acquires its native 3D shape involves formation of many non-covalent weak bonds, a classical one-trajectory view of protein folding is likely to be an over-simplification of the underlying mechanism. An original viewpoint emerged when the concept of energy landscape was extended to folding, described as a process taking place on a rugged free-energy surface, shaped as a funnel and involving entropy-enthalpy compensation [3]. Whilst the presentation of the funnel model provided a novel outlook on protein folding, a detailed experimental description of such a complex scenario is still a challenge. Hereby we present some recent experimental work, partly carried out in Rome, and briefly discuss its significance with reference to the energy landscape theory.

The surprising simplicity of protein folding kinetics

When the unfolded state folds into the native conformation, very many non-covalent bonds are formed, the solvent is excluded from the core of the protein and conformational entropy drops as the folded state is approached. Surprisingly however, when a small globular protein is challenged with increasing concentrations of denaturant, very often the native structure melts following a simple two-state reaction, such that only the native and denatured states are populated [4]. This all-or-none behaviour implies that the interactions stabilizing the native conformation are energetically coupled and tend to break (and form) cooperatively. The

Bioch. Soc. Trans. [in press]

experimental signature of cooperativity in protein folding lies in the sigmoidal reversible equilibrium transition and the single exponential time course when both the folding and unfolding reactions are followed in a transient experiment [5].

The observed rate constants for small single domain proteins can span several orders of magnitudes, ranging from 0.2 s^{-1} in the case of Acyl-Phosphatase [6] to up to 10^5 s^{-1} in the case of the Engrailed Homeodomain [7]. Because folding occurs more slowly than a diffusion-limited process (the pre-exponential factor being in the order of 10^6 to 10^8 s^{-1} [8,9]), it may be concluded that it is a barrier limited process. Thus, a two-dimensional free-energy diagram for a two-state folder consistent with kinetic folding and unfolding experiments has a single barrier for the major transition state (TS). Often the reaction coordinate of such an energy diagrams is the Tanford β_T value, which reflects the non-accessible surface area of the transition state relative to the denatured and native states (defined respectively as 0 and 1). β_T can be calculated from the dependence of the folding and unfolding rate constants on denaturant concentration (the so called chevron plot). For a two-state folder, crucial information about the energetic and structural properties of the transition state can be inferred by a combination of transient kinetics, site-directed mutagenesis and computer simulations (as championed by Fersht and coworkers [10]). Reliable structural information on the folding transition state (*via* the so-called Φ value analysis) and its characteristic β_T value is essential to unveil the set of contacts crucial to yield the native state.

We shall briefly outline below some experiments relevant to a few basic questions, such as: Is the denatured state populated under "native" conditions involved in dictating the final fold? Is it possible to reveal the presence of multiple nuclei at the early stages of folding? Do the structural nuclei of partially folded species along the pathway relate to the structure of the native state?

Plasticity and convergence of folding pathways

Implicit in the view of the energy landscape perspective is that folding occurs *via* a statistical selection between alternative parallel folding pathways (Figure 1). Experimentally, however, it is not so easy to document alternative folding routes. In fact, over-and-above cases such as lysozime [11] and cytochrome c_{551} [12] where evidence for more than one path was inferred by double jump mixing experiments, parallel folding pathways may be hidden in the reaction kinetics since often only one relaxation is experimentally accessible. In these cases, the presence of parallel routes may emerge from a complex analysis of the dependence of the (un)folding rate constant on solvent composition [13]. Quite recently, the ribosomal protein S6 and the PDZ domain have been subjected to extensive protein engineering experiments that, as discussed below, suggest the presence of alternative nuclei dictating the folding pathway, an observation fully consistent with the energy landscape perspective.

Wild-type S6 is a small globular single domain protein that, disregarding chain connectivity, is topologically consistent with two specular halves. The folding pathway, characterized by protein engineering and fast kinetics, suggests that the native state is achieved *via* a diffused transition state that encompasses most of the structure; however a nucleation centre with a detectable content of native like structure could be identified at the interface between helix1 and the N-terminal strand β1 [14-16]. By systematically altering the sequence connectivity between the six secondary structure elements synthesizing different circularly permuted variants, Oliveberg and coworkers observed that S6 displays a minimal folding nucleus composed of a *"two-strands-one-helix"* motif. The exact components of this motif however depend on the linear organization of the secondary structure elements, suggesting that circular permutation tunes the balance between competing nuclei involving in all cases the same helix but docking on different strands. This was taken as evidence for pathway malleability in protein folding, with more than one nucleation-competent structural motif within the protein domain. Thus, changes in

Bioch. Soc. Trans. [in press]

sequence connectivity may alter the relative stability of the folding nuclei and bias the dominant pathway towards alternative routes.

The PDZ domains are a family of proteins that has attracted a lot of attention because of their crucial role in cellular protein-protein interactions and selectivity of binding [17]. Given the ever growing evidence for of the interconnection between binding and folding, PDZ domains have been extensively investigated by kinetics and protein engineering. Curiously the PDZ domains from bacteria and plants may be considered as natural circularly permuted variants of their metazoan counterparts [18]. Hence they are attractive objects to investigate the relationships between sequence connectivity and protein topology and, for the first time, offer an opportunity to compare engineered and naturally evolved sequences.

The folding of all the PDZ domains studied so far (6 canonical and 2 circular permutants) involves the presence of an intermediate and two sequential transition states (Fig. 2) [19-22]. The intermediate may be either high energy as detected for canonical PDZ domains, or populated during folding as demonstrated in circularly permuted variants [23-25]. It was particularly interesting to compare the folding pathway of the second PDZ domain from the protein tyrosine phosphatase BL (PDZ2) to its engineered circularly permuted variant (*cp*PDZ2). These two proteins display the same overall structure and sequence composition, but different chain connectivity. Analysis of the structural features of the early and late events in the folding of PDZ2 and cpPDZ2 reveals that while the late stages in the pathway are essentially unaffected by circular permutation, early stages are structurally different with stabilization of alternative nuclei [24,25]. Thus in analogy with the S6 proteins, PDZ domains family display alternative folding pathways involving different nuclei, that can be selectively stabilized *via* loop entropy perturbations, such as circular permutation. Over-stabilization of a nucleus may lead to frustration of the folding landscape, involving the segregation into local minima that compete for producing the native state. An example of such a scenario is represented by the D1pPDZ domain, a naturally occurring circularly permuted variant that displays an off-pathway kinetic trap characterized by a misfolded N-terminal hairpin incorrectly docked on an otherwise native-like structure [26].

The folding funnel model predicts that the late stages of protein folding are more robust to sequence variations, while the early stages are more malleable [3]. This prediction is difficult to address experimentally, requiring as it does structural information on the early and late stages of folding, as well as capturing the presence of parallel folding pathways. Once again the PDZ domains proved to be ideal candidates to test such a prediction. Different PDZ domains, such as PDZ2 and PDZ3 (30% sequence identity) both fold *via* a complex mechanism involving the presence of at least one high-energy folding intermediate, as shown in Figure 2. Early and late events can be addressed experimentally, by probing the structure of the denatured-like transition state TS1 and the native-like transition state TS2 through a combination of Φ value analysis and restrained molecular dynamics simulations, to obtain atomic-level structures. It was found that the late transition states of PDZ2 and PDZ3 are structurally similar (TS2 being nearly super-imposable in both cases), while the early ones are quite different [19]. Thus in the case of TS1, limited but detectable native-like structural features could be identified in two different regions i.e. the β1-β4-β6 interaction (which is present in both PDZ2 and PDZ3) and the β2-β3 hairpin (which is dominant in PDZ3). Remarkably, the alternative nucleus involving β2-β3 could also be detected in *cp*PDZ2 [24], suggesting that folding may be rerouted through the alternative pathway *via* either sequence composition or circular permutation.

Overall, these results suggest that the bias of the native structure is weak at early stages of the folding reaction, allowing for alternative early folding events eventually leading to an intermediate; while it is very pronounced in the late stages when the native topology essentially dictates the folding mechanism. In summary, the influence and control of native topology on the structure of intermediate and transition states is loose and permissive at the

early stages of folding, but becomes more stringent and effective as the system is approaching the native state, reflecting the funnelling nature of the energy landscape.

Acknowledgements
Partially supported by a grant from the Ministero dell'Istruzione dell'Università e della Ricerca of Italy (RBRN07BMCT_007, to M.B.).

References

1. Anfinsen, C. B., Haver, E., Sela, M., and White, F. H. J. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc. Natl. Acad. Sci. USA **47**, 1309-1314
2. Levinthal, C. (1968) Are there pathways for protein folding? J. Chem. Phys. **65**, 44-45
3. Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins **21**, 167-195
4. Jackson, S. E., Fersht, A.R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry **30**, 10428-10435
5. Jackson, S. E. (1998) How do small single-domain proteins fold? Fold. Des. **3**, R81-91
6. van Nuland, N. A., Chiti, F., Taddei, N., Raugei, G., Ramponi, G., and Dobson, C. M. (1998) Slow folding of muscle acylphosphatase in the absence of intermediates. J. Mol. Biol. **283**, 883-891
7. Gianni, S., Guydosh, N. R., Khan, F., Caldas, T. D., Mayor, U., White, G. W., DeMarco, M. L., Daggett, V., and Fersht, A. R. (2003) Unifying features in protein-folding mechanisms. Proc Natl Acad Sci U S A **100**, 13286-13291
8. Fierz, B., Satzger, H., Root, C., Gilch, P., Zinth, W., and Kiefhaber, T. (2007) Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. Proc Natl Acad Sci U S A **104**, 2163-2168
9. Kubelka, J., Hofrichter, J., and Eaton, W. A. (2004) The protein folding 'speed limit'. Curr. Opin. Struct. Biol. **14**, 76-88
10. Matouschek, A., Kellis, J.T. Jr, Serrano, L., Bycroft, M., Fersht, A.R. (1990) Transient folding intermediates characterized by protein engineering. Nature **346**, 440-445
11. Kiefhaber, T. (1995) Kinetic traps in lysozyme folding. Proc Natl Acad Sci U S A **92**, 9029-9033
12. Travaglini-Allocatelli, C., Gianni, S., Brunori, M. (2004). A common folding mechanism in the cytochrome c family. Trends Biochem. Sci. 29: 535-541.
13. Wright, C. F., Lindorff-Larsen, K., Randles, L. G., and Clarke, J. (2003) Parallel protein-unfolding pathways revealed and mapped. Nat. Struct. Biol. **10**, 658-662
14. Lindberg, M., Tangrot, J., and Oliveberg, M. (2002) Complete change of the protein folding transition state upon circular permutation Nat. Struct. Biol. **9**, 818-822
15. Hubner, I. A., Lindberg, M., Haglund, E., Oliveberg, M., and Shakhnovich, E. I. (2006) Common motifs and topological effects in the protein folding transition state J. Mol. Biol. **359**, 1075-1085
16. Lindberg, M. O., and Oliveberg, M. (2007) Malleability of protein folding pathways: a simple reason for complex behaviour. Curr. Opin. Struct. Biol. **17**, 21-29
17. Jemth, P., and Gianni, S. (2007) PDZ domains: folding and binding. Biochemistry **46**, 8701-8708
18. Liao, D. I., Qian, J., Chisholm, D. A., Jordan, D. B., and Diner, B. A. (2000) Crystal structures of the photosystem II D1 C-terminal processing protease Nat. Struct. Biol. **7**, 749-753
19. Calosci, N., Chi, C. N., Richter, B., Camilloni, C., Engstrom, Å., Eklund, L., Travaglini-Allocatelli, C., Gianni,

S., Vendruscolo, M., and Jemth, P. (2008) Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. Proc Natl Acad Sci U S A **105**, 19241-19246

20. Gianni, S., Calosci, N., Aelen, J. M., Vuister, G. W., Brunori, M., and Travaglini-Allocatelli, C. (2005) Kinetic folding mechanism of PDZ2 from PTP-BL. Prot. Eng. Des. Sel. **18**, 389-395
21. Gianni, S., Geierhaas, C. D., Calosci, N., Jemth, P., Vuister, G. W., Travaglini-Allocatelli, C., Vendruscolo, M., and Brunori, M. (2007) A PDZ domain recapitulates a unifying mechanism for protein folding. Proc Natl Acad Sci U S A **104**, 128-133
22. Haq, S. R., Jürgens, M. C., Chi, C. N., Koh, C. S., Elfström, L., Selmer, M., Gianni, S., and Jemth, P. (2010) The plastic energy landscape of protein folding: a triangular folding mechanism with an equilibrium intermediate for a small protein domain. J. Biol. Chem. **285**, 18051-18059
23. Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M., and Gianni, S. (2008) Folding and Misfolding in a naturally occurring circularly permuted PDZ domain J. Biol. Chem. **283**, 8954-8960
24. Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M., and Gianni, S. (2009) Engineered symmetric connectivity of secondary structure elements highlights malleability of protein folding pathways. J. Am. Chem. Soc. **131**, 11727-11733
25. Ivarsson, Y., Travaglini-Allocatelli, C., Morea, V., Brunori, M., and Gianni, S. (2008) The folding pathway of an engineered circularly permuted PDZ domain. Prot. Eng. Des. Sel. **21**, 155-160
26. Gianni, S., Ivarsson, Y., De Simone, A., Travaglini-Allocatelli, C., Brunori, M., and Vendruscolo, M. (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. Nat. Struct. Mol. Biol. **17**, 1431-1437

Bioch. Soc. Trans. [in press]

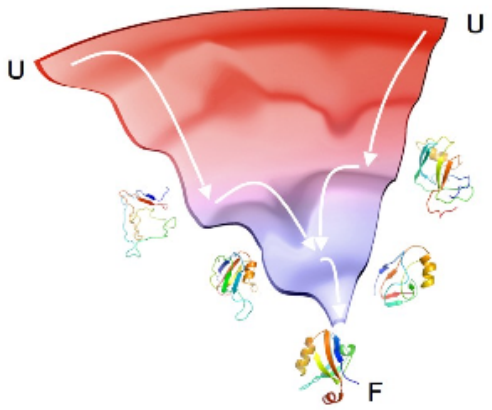


Figure 1. *Schematic representation of a folding funnel: the free energy drives the polypeptide towards the folded state, the most stable conformation, and the conformational entropy dramatically decreases as the native state is approached. Implicit in this view is that folding may occur via alternative parallel pathways. This concept is graphically depicted using the structures of the early and late folding transition states for different PDZ domains, as obtained in ref. [19].*

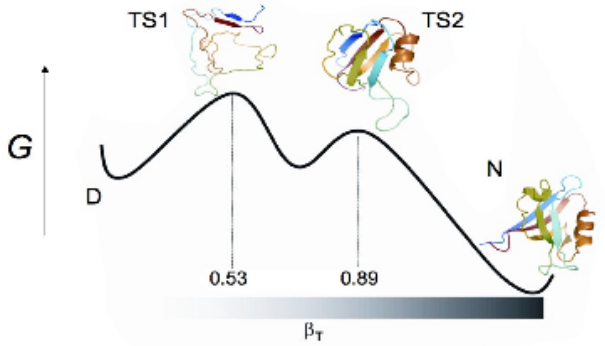


Figure 2. *Folding free-energy diagram of PDZ2. Representative structures of TS1 and TS2 are reported together with the native state (Protein Data Bank ID code 1GM1). Structures were calculated using a combination of Φ value analysis and molecular dynamics simulations, as described in ref. [21], modified.*