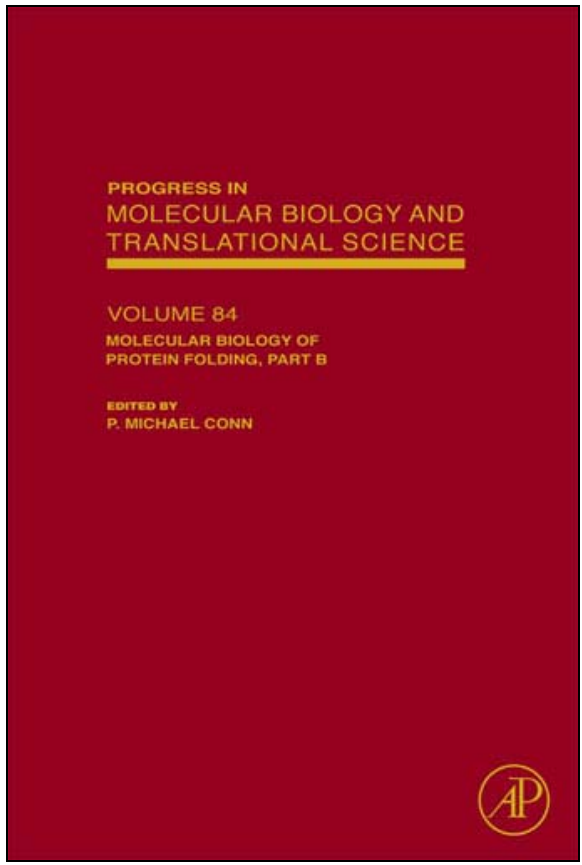


From: Anna L. Mallam and Sophie E. Jackson, Use of Protein Engineering Techniques to Elucidate Protein Folding Pathways. In P. Michael Conn, editor: Progress in Molecular Biology and Translational Science, Vol. 84, Burlington: Academic Press, 2008, pp. 57-113.
ISBN: 978-0-12-374595-8

Use of Protein Engineering Techniques to Elucidate Protein Folding Pathways

Anna L. Mallam and Sophie E. Jackson

Department of Chemistry, Cambridge, CB2 1EW, United Kingdom

Since the development and application of protein engineering techniques in the early 1980s, these methods have revolutionized many areas of biological science. Their use and impact in the field of protein folding has been enormous. This review describes many of the different approaches to studying protein folding pathways that have employed protein engineering techniques. The review covers two decades of science, describing the early work on tryptophan synthase and dihydrofolate reductase, through the development and widespread application of Φ-value analysis to the folding of small "model" systems, to more recent work on larger proteins with complex topologies. In addition, the application of protein engineering methods to study specific processes linked with protein folding such as proline isomerization and disulphide bond formation are also discussed. The review ends with a brief section on how these techniques are currently being used to engineer probes and proteins for single-molecule experiments.

Progress in Molecular Biology and Translational Science, Vol. 84
DOI: 10.1016/S0079-6603(08)00403-0

0079-6603/08 $35.00

I. Introduction

As with almost every other area of biological science, the most powerful tool that has been developed and which has revolutionized the study of protein folding pathways has been protein engineering. It is now possible to engineer proteins in many different ways: single-site mutations can be made which probe the role of a single side chain, multiple mutations can be introduced into a protein for many different purposes. These can be specifically and rationally designed, for example, to increase or decrease the intrinsic secondary structural propensity or to switch the surface and core of two similarly structured proteins, or a library of multiple mutants can be created from which proteins with specific properties can be selected. Protein engineering techniques can be used to generate small, single domains from large multi-domain proteins, these domains often representing the simplest of folding systems. The structures of proteins can also be manipulated in more complex ways—circular permutants, in which the original N and C- termini are fused and new N- and C-termini made, can be produced. In addition, chimeric proteins consisting of regions of sequence from different, usually related, proteins can be constructed. Novel fusion proteins can be made for a wide variety of uses: in some cases, these may be fusions of identical protein domains or repeat motifs, for example, for the study of the mechanical stability of proteins or the folding pathways of repeat proteins. In other cases, completely different proteins may be fused together to study how one folds in the presence of the other.

In this review, we describe how protein engineering techniques have been used over the past 25 years to study the folding pathways of many proteins with very different structures and folding mechanisms. The review starts with a brief account of the early work in the field and the first folding studies of mutant proteins. A large section on single-site mutants and Φ-value analysis provides a detailed account of the considerable number of studies that have used this approach to study the folding pathways of small proteins or protein domains. Here, we have tried to be as comprehensive as possible; however, the focus is on proteins where many mutations have been analyzed providing a detailed picture of the folding energy landscape for these proteins. Alternative approaches that have been developed but not as extensively applied as Φ-value analysis are discussed in the section on Ψ-value analysis, a novel approach which uses engineered metal binding sites to study folding. Recent work on the folding of repeat proteins, which contain two or more repeated structural units or motifs, is also described. Here, several different protein engineering strategies have been employed to elucidate the folding pathways of these unusual structures. The use of circular permutants to probe folding processes and, in particular, the role of local versus long-range interactions is covered in

Section VII. In addition, the application of protein engineering methods to study specific processes linked with protein folding such as proline isomerization and disulphide bond formation are also discussed. The review ends with a brief section on how protein engineering techniques are currently being used to engineer probes and proteins for single-molecule folding experiments.

II. Early Protein Engineering Studies of Folding Pathways

It was in the early and mid 1980s that the first studies were published, which used single point mutants of proteins to investigate the factors that govern protein stability and determine folding pathways. The group of C. Robert Matthews was the first to use the powerful and relatively new technique of protein engineering on the α subunit of tryptophan synthase (αTS). In 1983, a study was published on the effects of a single point mutant (Gly→Glu211) on both the stability and the kinetic behavior of αTS (*1*), thereby establishing what was going to become, and possibly still is, the most important tool for the experimental study of protein folding pathways. This initial work was rapidly followed with a series of studies on different sets of mutants of αTS, including a set of single point mutants testing the domain unfolding model (*2*), single and double mutants that established synergism within the protein structure (*3*), and multiple mutants at a single position (*4*). This body of work, the principles behind this approach, and the possible effects that mutations could have on stability curves, kinetic chevron plots, and energy levels were reviewed in three excellent articles published in 1987 (*5–7*).

Following on from their work on αTS, the Matthews group applied the protein engineering approach to studying folding pathways of another protein—dihydrofolate reductase (DHFR). In this case, a high-resolution crystal structure was available from the outset, such that mutations could be rationally designed to probe specific regions of the protein, Fig. 1. The role of secondary structure in the folding mechanism of DHFR was probed by studying a set of mutants in two of the α-helices and two of the β-strands (*8*). The effects of the mutations on stability were rationalized from the interactions that the mutated side chains made in the crystal structure. Significant differential effects of mutants in different elements of secondary structure were observed and established that different regions of secondary structure form at different times along the folding pathway. Multiple mutants at a single position in DHFR were made in order to determine the relative importance of position and specific character of the residue to folding (*9*). In this case, one set of mutants behaved in a consistent manner, whereas a tyrosine substitution showed strikingly different behavior, showing that position and specific amino acid type are important in determining how the protein folds. The influence of long-range

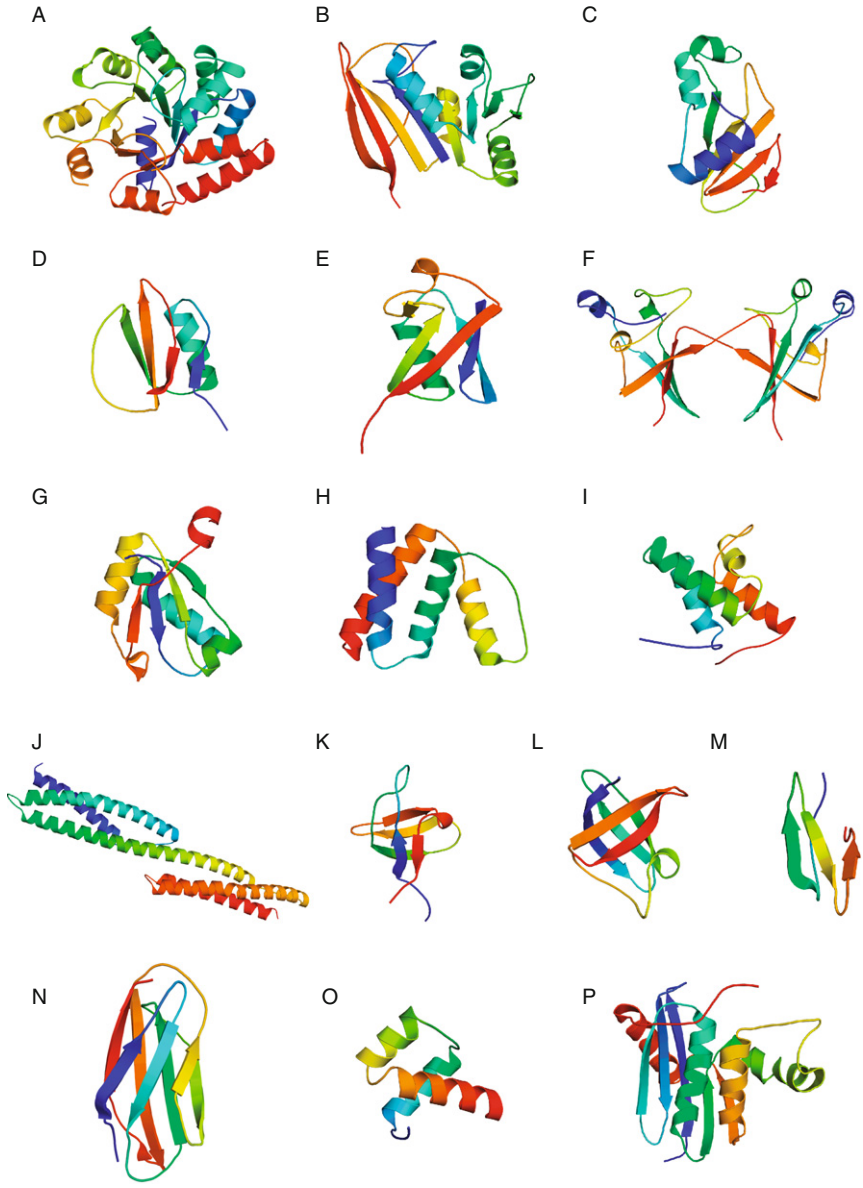


FIG. 1. Structures of some of the proteins whose folding pathways have been characterized using protein engineering techniques and Φ-value analysis. (A) The α subunit of tryptophan synthase (αTS) (PDB ID 1A50), (B) Dihydrofolate reductase (DHFR) (PDB ID 3DFR), (C) Barnase (PBD ID 1BAN), (D) CI2 (PDB ID 2CI2), (E) Ubiquitin (PDB ID 1UBQ), (F) Suc1 (PDB ID 1SCE), (G) U1A (PDB ID 1NU4), (H) Acyl co-enzyme, a binding protein (ACBP)

electrostatic interactions on folding was investigated in another study (*10*). In this case, a double mutant in which two positively charged side chains were engineered in the protein (Arg28Lys139) showed that long-range electrostatic interactions could affect stability and folding. This was the first example where mutations were shown to change the apparent mechanism from two-state to multistate.

αTS was chosen for folding studies as it was a single polypeptide chain with no prosthetic groups, and many missense mutants had been isolated by the Yanofsky group (*1*). However, αTS is a large (29 kDa) protein (Fig. 1) and is a complex system with which to study protein folding pathways. For αTS, both equilibrium and kinetic experiments showed evidence for multiple states and transitions (*11, 12*). Although wild-type DHFR followed a simple two-state model under equilibrium conditions, it too had complex kinetics (*13*).

III. Single Point Mutations and Φ-Value Analysis

At the beginning of the 1990s, Alan Fersht's group published a significant number of papers in which protein engineering techniques were used extensively on the small ribonuclease barnase to study both protein stability and folding. Barnase, a protein only a 110 residues in length, was a simpler system than either αTS or DHFR. Not only was it relatively small, but it showed robust two-state behavior under equilibrium conditions and evidence of a single intermediate state on the folding pathway (*14*). The effects of a mutation on protein stability and folding were formalized by the use of Φ values, which gave a quantitative measure of the effect of a mutation on the rate-limiting transition state relative to its effect on the energy of the native structure (*15*). This approach, together with the characterization of a significant number of mutants, enabled the structure and energetics of a folding transition state to be characterized in some detail for the first time (*15, 16*), and established a technique and analysis method that is now widely used by the protein folding community.

Figure 2 illustrates the theory behind Φ-value analysis and the interpretation of Φ values, which is described in detail in two classic papers published by the Fersht group in 1991 and 1992 (*17, 18*). The latter paper, a must-read for

(PDB ID 1HB8), (I) The immunity (Im) protein Im9 (PDB ID 1IMQ), (J) Repeats 15 and 16 of chicken brain α-spectrin (PDB ID 1U5P), (K) The SH3 domain from fyn (PDB ID 1FYN), (L) Cold shock protein B (CspB) (PDB ID 1MJC), (M) The WW domain (PDB ID 1K9Q), (N) A fibronetin type III domain from tenascin (PDB ID 1TEN), (O) *D. melanogaster* engrailed homeodomain (PDB ID 1ENH), (P) RNaseH (PDB ID 1RDD).

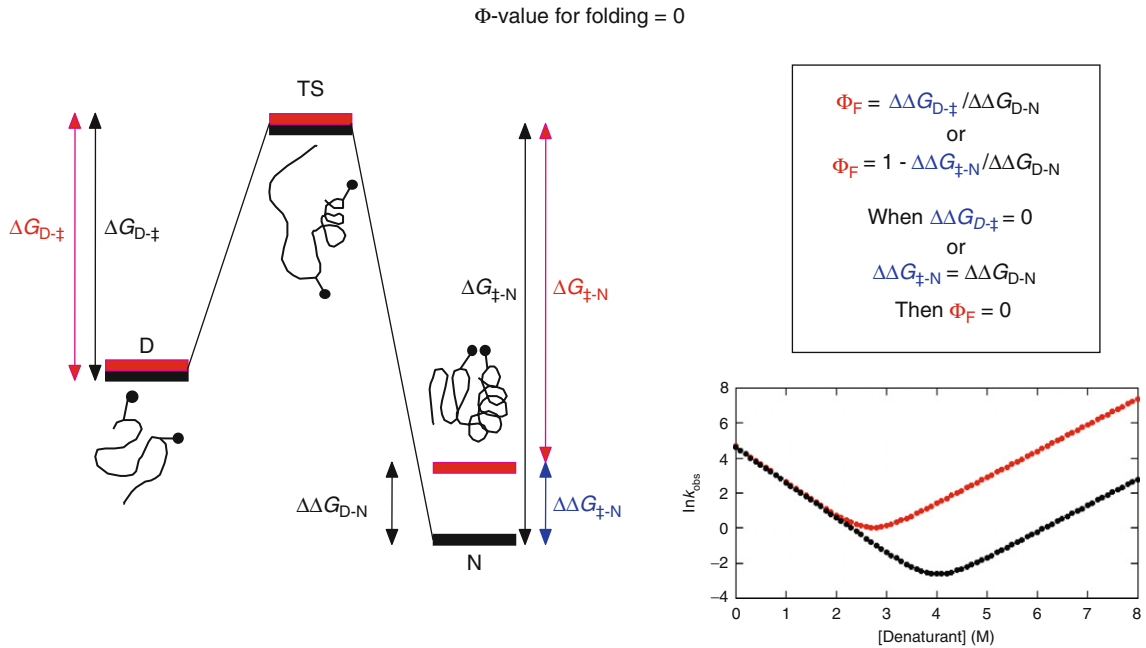


FIG. 2. (*Continued*)

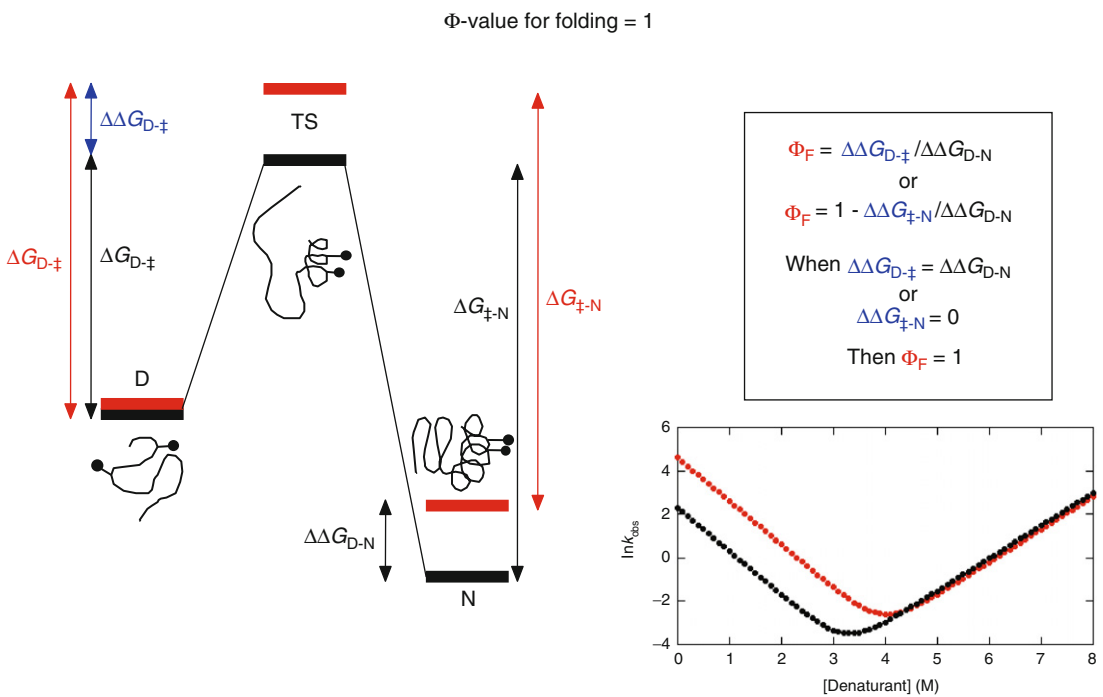


FIG. 2. (*Continued*)

anyone undertaking Φ-value analysis, discusses the assumptions in the method, the optimal types of mutation to make in order to interpret Φ values with a high degree of accuracy, and the limitations of the approach. The initial study on barnase used measurements on the unfolding kinetics of wild type and mutants to calculate the energetics and structure in the rate-limiting transition state. This approach had to be taken for barnase as folding kinetics could not be used directly due to the transient population of an intermediate state during folding. The principle of microscopic reversibility was used to justify the use of measurements made under strongly denaturing conditions to describe a folding transition state in native-like conditions (*19*). Subsequently, this was tested by measurements on an even simpler, kinetically two-state system, CI2 (*20*). In this case, Φ values were calculated from unfolding data acquired at high concentrations of denaturant, in a similar manner to that used for barnase, but in addition, Φ values were also calculated directly from folding data measured by pH-jump experiments in water (*21*). Both sets of Φ values were within experimental error, thus validating the method.

FIG. 2. Free energy diagrams illustrating the principles of Φ-value analysis. The free energy of the denatured (D), transition state (TS) and the native (N) state are shown for both the wild-type protein (black) and for a destabilizing mutant (red). The degree to which the mutation destabilizes the native state is given by $\Delta\Delta G_{D\text{-}N}$, the difference in the free energy of unfolding between wild type and mutant. The effect of the mutation on the energy of the transition state is given by $\Delta\Delta G_{D\text{-}\ddagger}$ which for a two-state system can be calculated directly from folding data. For a non-two-state system, then $\Delta\Delta G_{\ddagger\text{-}N}$ can be calculated from unfolding data and also used to calculate a Φ value. The Φ value is simply the ratio of $\Delta\Delta G_{D\text{-}\ddagger}$ to $\Delta\Delta G_{D\text{-}N}$. The structure of a model protein is shown in the figure to illustrate how a mutated residue and the interactions it might make with other groups in the protein changes as folding proceeds.(A) In this case, the mutation has a destabilizing effect on the native state of the protein but no effect on the transition state, such that $\Delta\Delta G_{D\text{-}\ddagger}$ and therefore Φ are both zero. This is the case when the side chain of the mutated residue makes no interactions in the transition state, thus, the protein is unstructured in this region in the transition state. In this case, the effect of the mutation is observed in an increase in the unfolding rate of the protein, with no effect on the folding kinetics.(B) In this case, the mutation has a destabilizing effect on the native state of the protein and also has an equally destabilizing effect on the transition state, such that $\Delta\Delta G_{D\text{-}\ddagger}$ is the same as $\Delta\Delta G_{D\text{-}N}$ and the Φ value is one. This is the case when the side chain of the mutated residue makes as many interactions in the transition state as it does in the native state, thus, the protein is highly structured in this region in the transition state. In this case, the effect of the mutation is observed in a decrease in the folding rate of the protein, with no effect on the unfolding kinetics.Fractional Φ values (not illustrated here) are indicative of the side chain of the mutated residue making some, but not all, of its native-like interactions in the transition state and therefore this region of the protein being partially structured in the transition state. For a comprehensive discussion on the theory of Φ−value analysis, see references (17) and (18).

Extensive protein engineering and Φ-value analysis studies on barnase and CI2 provided the most detailed characterizations of folding transition states to date (*15–17, 21–23*). The studies on CI2, in particular, were used to develop a new mechanism for the folding of small, monomeric proteins—the nucleation–condensation mechanism (*23*).

The engineering of single point mutations into proteins and Φ-value analysis has now been applied to a large number of proteins with different structures (Fig. 1), stabilities, and folding pathways. Most of these proteins are relatively small with comparatively simple folding kinetics. The following sections describe in detail the results from these studies.

A. Mixed α/β Proteins

1. Ubiquitin-Like β-Grasp Folds

Proteins with a ubiquitin-like β-grasp fold have a β-hairpin–α-helix–β-hairpin topology which results in the α-helix packing against a mixed β-sheet to form the core of the protein. The first protein engineering/protein folding study on ubiquitin focused on a single residue which lies in the core of the protein at the α-helix/β-sheet interface. Single mutations at this site were shown to change the folding kinetics of the protein, although the few mutants characterized were insufficient for a full Φ-value analysis (*24*). Despite the interest in ubiquitin as a model system for studying protein folding pathways (*25*), a comprehensive mutational analysis of the folding pathway was not published until 2005 (*26*). This study revealed that the folding nucleus of ubiquitin is highly polarized with structure residing almost entirely in the N-terminal region of the protein comprising the N-terminal β-hairpin and α-helix, with little or no structure in the C-terminal region of the β-sheet (*26*). In a very recent study, the folding of another member of the β-grasp fold family of proteins, the ras-binding domain (RBD) from a raf kinase, was also characterized extensively using Φ-value analysis (*27*). Despite very low sequence similarity between the two proteins, the RBD was also shown to have a structurally polarized transition state similar to that found for ubiquitin, with particularly native-like structure in the N-terminal β-hairpin. All residues constituting the inner core of the protein were to some degree involved in structure in the transition state. Interestingly, an analysis of the effect of mutations on the energy of the transition state showed that it was structurally more diffuse than as assessed by Φ values (*27*).

The immunoglobulin binding domains of protein L and protein G also adopt a β-grasp structure similar to ubiquitin and RBD (*28*) and their folding pathways have been extensively studied using protein engineering techniques by the Baker group. For protein L, the first β-hairpin is highly structured in the transition state, whilst the second β-hairpin is not formed (*29*), and the α-helix

is also largely disrupted (*30*). Destabilizing surface hydrophobic core mutants of protein L have been studied and shown not to affect the folding rate, suggesting that nonnative hydrophobic interactions do not interfere with hydrophobic core assembly (*31*). Together, the folding of more than 70 mutants of protein L have been measured, making it one of the most fully characterized proteins. The results emphasize that secondary structure formation can play an equally important role as hydrophobic core formation (*32*). In contrast to the results obtained for protein L, a protein engineering study on the folding pathway of protein G showed a very different order of structure formation. For this protein, it is the second β-hairpin that is structured in the transition state with the first β-hairpin having little structure (*33*). The difference in folding mechanism is attributed to intrinsic differences in the relative stabilities of the two β-hairpins in protein L and protein G. To test this hypothesis, the Baker group redesigned the first β-hairpin of protein G to maximize its stability and redetermined the folding pathway by Φ-value analysis (*34*). As expected, the engineered protein folded in a manner similar to protein L, establishing that it is possible to switch the folding pathway of a protein by modulating the intrinsic stability of elements of secondary structure (*34*).

2. Cell Cycle Regulation Proteins

The folding of two cell cycle regulation proteins, suc1 and cks1, has been studied by extensive mutational analysis and in one case, by Φ-value analysis (*35–37*). These proteins adopt an α/β structure in which two β-strands are followed by two α-helices and then two more β-strands which pack to form a two-layer structure with antiparallel β-sheet. Φ-value analysis of 57 mutants of monomeric suc1 was used to generate a detailed picture of the structure of the transition state ensemble for folding for this protein (*35*). The highest Φ values were clustered around β-strands 2 and 4 which form the center of the β-sheet, this region of the structure forming the folding nucleus. Other regions of the protein had lower Φ values and are less structured in the transition state (*35*). Suc1 is known to form a domain-swapped dimer in which one molecule exchanges a β-strand with an identical partner. A Φ-value analysis on the dimer established that it has similar structure in the transition state to that found for the monomer; however, the Φ values were consistently higher in the dimer than the monomer indicating that there was greater structure (*36*). These results were interpreted in terms of enthalpy/entropy compensation—the dimer loses more entropy than the monomer on forming the transition state ensemble, and this loss of entropy is overcome by a greater enthalpic gain brought about by the formation of native-like structure (*36*). In contrast to monomeric suc1, a mutational analysis of the related cell cycle regulation protein cks1 was complex as large changes in the slope of the unfolding and refolding limb of chevron plots was observed on mutation (*37*). Despite this, a

novel analysis was employed which showed that cks1 folds sequential pairs of β-strands first—β1/β2 and β3/β4, these strands then pack onto each other and the α-helix to form the core (37). It is likely that the differences in folding behavior between suc1 and cks1 are due to the very different stabilities of the two proteins, and that the folding of suc1 is a much more cooperative process than cks1 because the entropic costs of forming long-range interactions in the hydrophobic core are adequately compensated for by the formation of a large number of favorable interactions (37).

3. Ferrodoxin-Like Folds

A large number of proteins are classified in the ferrodoxin-like fold superfamily (28). These proteins have an α and β sandwich structure with an antiparallel β-sheet, the topology is $\beta\alpha\beta\beta\alpha\beta$. This includes the proteins AcP, Ada2h, U1A, and S6.

Acyl phosphatase (AcP) and Ada2h have an $\alpha + \beta$ structure in which the two helices pack against the four-stranded antiparallel β-sheet. The folding pathways of both proteins have been studied by protein engineering and Φ-value analysis, and the structure of their transition states compared (38–40). For Ada2h, 20 single point mutations spread throughout the structure were used to probe the structure in the transition state. Some secondary structure was found to be formed in addition to the hydrophobic core, which is in the process of being consolidated in the transition state (39). A folding nucleus comprising of the packing of α-helix 2 and the two central β-strands was identified (39). In addition, stabilizing mutations in α-helix 2 were generated which resulted in a faster folding variant of Ada2h (38). A study on AcP showed that the transition state ensemble is an expanded form of the native structure where most interactions are at least partially established, and which was very similar to that observed for Ada2h suggesting that protein folding transition states are conserved (40).

Folding studies on the wild-type spliceosomal protein U1A showed interesting features with a symmetrically curved chevron plot, the folding transition state becoming more native-like at high concentrations of denaturant (41). A Φ-value analysis of the folding pathway of U1A showed a pattern of values consistent with a nucleation–condensation mechanism with a structurally diffuse nucleus. The first interactions formed during folding are between β-strands 2 and 3 and the first α-helix which forms part of the hydrophobic core. Subsequently, β-strand 1 becomes structured and then β-strand 4 and α-helix 2 pack against this structure to form the fully folded protein (41). A movement of the transition state with denaturant concentration enabled a fuller picture of the interactions and structure present in the transition state to

be elucidated. Both the formation of, and the subsequent growth of, the folding nucleus could be followed. The results showed that the folding of different regions of the protein is highly coupled (*41*).

The folding of S6, a ribosomal protein which also adopts the ferrodoxin-like fold has also been characterized by Φ-value analysis by the Oliveberg group (*42*). In contrast to U1A, wild-type S6 shows classic two-state folding kinetics with no curvature, however, many of the mutants change the kinetic *m*-values (the slopes of the chevron plot) leading to kinks and curves. This unusual behavior is linked to mutations which are clustered in a distinct region in the native structure and suggests a general plasticity of the energy landscape for folding. The results show that interactions in the hydrophobic core form early in folding, whilst entropically disfavored interactions between the N- and C-termini form very late after the protein has overcome the energetic barrier and transition (*42*).

4. Other α/β Proteins

The ribosomal protein L23 has a central four-stranded β-sheet that interacts with three α-helices to form a saddle-shaped hydrophobic core. Seventeen mutants of L23 have been made and their unfolding and folding kinetics determined (*43*). All the mutants had fractional Φ values with the highest being 0.44, indicating that no element of structure is fully formed in the transition state and that this protein has an extended folding nucleus. These results are consistent with a nucleation-condensation mechanism. The residues with the highest Φ values cluster in the centre of the core of the protein and link the central β-strand 4 with α-helix 2. All other residues have lower Φ values which showed a gradual decrease as one moves away from the nucleating position (*43*). Analysis of Hammond effects for the folding of mutants of L23 showed a set of primary interactions that are critical in maintaining the correct overall topology for nucleation which are formed on the uphill side of the barrier, and a secondary set of interactions that are formed as the protein traverses the transition state and which lead to downhill folding to the native state (*43*).

RNaseH is known to fold through a populated intermediate state and has been proposed to fold in a hierarchical manner (*44*). Protein engineering techniques have been used to acquire further evidence to support this hypothesis. Mutations within what had been identified as the folding core of the protein were observed to destabilize the kinetic intermediate and slow folding, whilst mutations outside the core had little effect on the stability of the intermediate but were observed to affect both transition and native states (*45*). Interactions that were formed in the intermediate state were found to persist in the transition state, consistent with a hierarchical model of folding (*45*). Mutation of three residues involved in a buried salt bridge in RNaseH created a protein in which the intermediate state accumulated even in the presence of

low concentrations of denaturant, leading to a system which was three-state under equilibrium conditions (*46*). In another study, some mutations at position 53, a hydrophobic residue located at the interface between two helices, resulted in destabilization of the intermediate and a switch from three- to two-state folding kinetics (*47*).

B. All-α-Helical Proteins

The folding pathways of many all-α-helical proteins have now been characterized using protein engineering techniques. This includes studies on proteins with very different sizes and helical topologies including the small three- and five-helical bundle structures of engrailed homeodomain and λ repressor, the larger four-helical bundle structure of ACBP, and the nonbundle structures of the Immunity proteins and spectrin domains.

1. Engrailed Homeodomain

The engrailed homeodomain (EnHD) from *Drosophila melanogaster* is an example of a small, fast folding protein domain that forms a three-helical bundle structure (*48*). Various aspects of the energy landscape for folding of this protein have been investigated experimentally by using protein engineering techniques, which in conjunction with molecular dynamics simulations have provided an exquisite picture of how this small protein domain folds, Fig. 3 (*49, 50*). A highly destabilized mutant of EnHD, L16A, which is predominantly unfolded under conditions where the wild-type protein is folded, was used to investigate the denatured state under native-like conditions (*49*). Extensive secondary structure was observed and the polypeptide chain was found to be remarkably compact and globular, although the side chains and backbone were highly mobile (*49*). A classic Φ-value analysis of EnHD revealed that there is also extensive secondary structure and helix formation in the transition state, with the helices being nearly formed and docking of the helices together in the process of taking place, consistent with a diffusion–collision model of folding (*51*). A truncated mutant of EnHD in which only helix-2 and -3 were present was folded and monomeric in solution (*50*). Kinetic experiments on this truncated mutant showed that it folded on the same time scale as the fast folding phase of the full-length EnHD, consistent with this region of the protein forming structure extremely fast on the folding pathway (*50*).

2. λ Repressor

A monomeric form of the λ repressor has been engineered by truncating the protein to residues 6–85. This construct has an all-α five-helical bundle structure which has been shown to unfold and fold extremely quickly (*52*). A thermostable variant of λ repressor (6–85), in which helix 3 was specifically stabilized by two Gly→Ala mutations, was shown to fold an order of magnitude

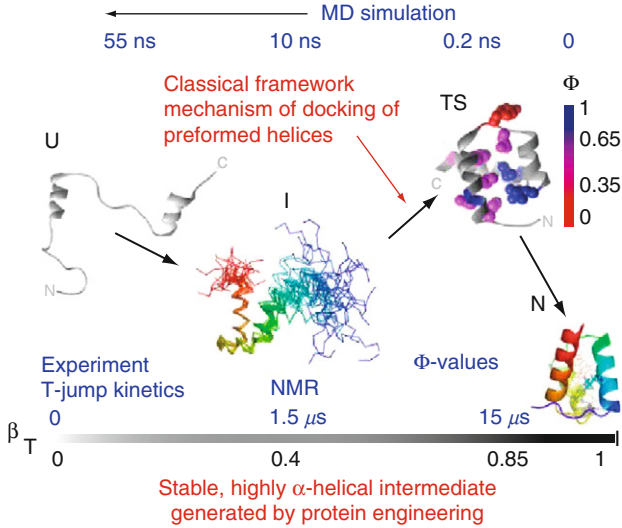


FIG. 3. The complete folding pathway of engrailed homeodomain from nanoseconds to microseconds. The Fersht group used protein engineering techniques in conjunction with Φ-value analysis to characterize the rate-determining transition state for folding of engrailed homeodomain. The Φ values are color coded such that high values (close to 1) are blue, whilst low values (close to zero) are red. Protein engineering techniques were also used to characterize the intermediate state by NMR. Fast reaction kinetics using temperature-jump apparatus were performed to observe and characterize the formation of the intermediate state from the denatured ensemble, and computational molecular dynamic (MD) simulations were performed to gain further structural detail on all states on the folding pathway. This figure is adapted from reference (51).

faster in less than 20 μs (*53, 54*). At the time, this was the fastest folding protein known. The results of these studies also showed that intrinsic helical stabilities play an important role in the folding of this type of structure and that the position of the transition state on the folding reaction coordinate is sensitive to mutation (*53, 54*). A diffusion-collision model of folding was used along with estimates of intrinsic α-helical propensity to develop a model for the folding of the monomeric form of the λ repressor, which accurately predicted not only folding rates but also the nature of the transition state ensemble (*55*). The model predicted that whereas the pseudo wild-type protein folds more slowly through a small number of pathways, the double Gly→Ala mutant folds faster through a large number of different pathways, as observed (*55*). Subsequent protein engineering studies on the folding pathway of monomeric λ repressor have shown that a buried hydrogen bond is not formed in the transition state (*56*), and that α-helix N- and C-capping motifs can have different effects on

folding kinetics depending upon their position (*57*). The Oas group have also used a similar approach on another ultra-fast folding small, all-helical protein—the B-domain from protein A (BdpA). Despite the fact that the wild-type 58-residue, three-helical bundle protein folded with super-fast kinetics, a stabilized Gly→Ala mutant was found to fold even faster on a time scale of about 3 μs (*58*). Thus, even for this super-fast folding protein, the folding rate has not been optimized by evolution.

3. ACBP

Acyl coenzyme A binding protein (ACBP) is a four-helical bundle protein which has been studied extensively by the Poulsen group. A comparison of ACBPs shows that there are 26 highly conserved positions in the 81-residue protein. These residues have all been mutated in bovine ACBP and the effect on folding kinetics measured (*59*). Mutants at 8 out of the 24 positions showed large effects on their folding rates indicating that these residues are critical for fast folding. The residues are all located in the hydrophobic core in the interface between the N- and C-terminal α-helices (*59*). In a more recent study, a set of mutations was made in yeast ACBP and Φ values calculated (*60*). Out of the set of 16 equivalent mutations in yeast and bovine ACBP, only 5 showed consistently high Φ values in both proteins. In yeast ACBP, there were a further three mutants with high Φ values indicating that additional interactions between α-helix 2 and 4 play a role in stabilizing the transition state, whilst in bovine ACBP, there were a different three mutants which showed high Φ values which indicated that there were more substantial interactions between α-helix 1 and 4 in this case (*60*).

4. Immunity Proteins

The Radford group have studied the folding of a different class of four-helix proteins which do not form a bundle structure, the immunity (Im) proteins, in some detail, extensively applying a range of protein engineering techniques. Three different Im proteins which have a reasonable degree of sequence similarity—Im2, Im7, and Im9 have been examined. In early work, the folding kinetics of Im2 and Im9 and a series of chimeric variants in which regions of secondary structure were combined from the two parent proteins were characterized (*61*). Some chimeras showed behavior similar to the parent whilst others did not. From this study, it was concluded that, whereas topology does play an important role in folding, specific interactions can also modulate the energy landscape (*61*). Whilst wild-type Im2 and Im9 both fold with two-state kinetics, Im7 was found to fold with three-state kinetics and populate an intermediate state during folding (*62*). A comprehensive Φ-value analysis of the intermediate state of Im7 using 29 point mutations established that three out of the four helices have native-like structure and these are packed around a

hydrophobic core (62). Unusually, there was also evidence of significant nonnative interactions in the intermediate state, consistent with a misfolded species in which the nonnative interactions have to be broken before the protein can pass over the rate-limiting transition state barrier (62).

A Φ-value analysis of the transition state of folding for Im9 showed very similar results to Im7 (63). The pattern of Φ values indicated that three out of the four helices have native-like structure in the transition state and these were docked around a hydrophobic core. However, the magnitude of the Φ values measured for Im9 were all significantly lower than those found for Im7. The results suggested that the transition state for Im7 is conformationally restricted compared to Im9 and that the population of an intermediate state prior to the rate-limiting transition state can have an effect on the breadth of the transition state ensemble (63).

Using the knowledge obtained on the structure of the intermediate state of Im7, variants of the protein were designed to trap it in the intermediate state by removing interactions between α-helix 3 and the rest of the protein. Characterization of these variants showed that they had a substantial amount of helical structure and a well formed hydrophobic core (64). However, in contrast to the native state of Im7, the intermediate state was observed to be flexible enough to undergo some structural rearrangement in response to mutation (64). Additional NMR and small-angle X-ray scattering (SAXS) measurements on engineered variants of Im7 which were trapped in the intermediate state even under equilibrium conditions showed that the intermediate is much more conformationally dynamic than the native state (65). Chemical shift data indicated that α-helices 1 and 4 are formed, α-helix 2 is partially formed but there is little structure in α-helix 3, consistent with earlier protein engineering studies. Backbone dynamic measurements established that the hydrophobic core in the intermediate state is not uniquely structured and, despite the presence of extensive secondary structure, the hydrophobic core is relatively fluid (65).

In a similar study, Im9 was engineered with a set of rationally designed mutations to stabilize the intermediate state such that its kinetics changed from two-state to three-state (66). A Φ-value analysis on the intermediate state showed that it had structure very similar to that observed for Im7. The three-state variant of Im9 folded 20 times faster than the two-state wild-type Im9 showing that the formation of an intermediate state can speed up the folding of even small proteins and that nonnative interactions can play an important role in stabilizing intermediate states (66).

Further protein engineering studies on Im9 investigated the role of helical propensity and hydrophobicity on the folding pathway (67). Whereas increasing the helical propensity of residues in solvent-exposed positions in α-helices 1, 2, or 4, resulted in a switch from two- to three-state kinetics, increasing the

hydrophobicity in helices 1 and 4 had no effect. In contrast, increasing the hydrophobicity of residues in α-helix 2 resulted in a stabilization of both the intermediate and transition state, again confirming the role of nonnative interactions in the folding of this class of protein (*67*). Further protein engineering and molecular dynamic simulations probed the role of nonnative interactions on the energy landscape of Im9 and showed that the rate-limiting step involves reorganization of the nonnative interactions as the transition state is traversed (*68*). This study demonstrated that the immunity protein family folds along closely-related micropathways, the exact pathway depending upon the relative strength of native and nonnative interactions (*68*).

The Radford and Imperiali groups have also used semisynthetic methods on Im7 to produce a novel glycoprotein variant (*69*). Native chemical ligation was used to produce an N-linked chitobiose glycoprotein using an engineered cysteine mutant of Im7. The folding of the glycoprotein was found to be similar to the parent protein (*69*).

5. All-α-Spectrin Domains

The all-helical R16 and R17 domains from spectrin are somewhat different in nature to the other all-helical proteins already discussed in that they form more rod-like structures rather than globular structures with three long α-helices. The folding of R16 has been investigated and showed curvature in the unfolding limb of the chevron plot indicative of a sequential mechanism with a high energy intermediate (*70*). A Φ-value analysis showed that, for most mutants, Φ values increased between the early and the late folding transition state consistent with a hierarchical mechanism and consolidation of structure during folding (*71*). One or two residues, however, did have Φ values which were lower in the late transition state, possibly arising because of the need to relax some of the α-helical structure formed early on the final packing of the core (*71*). A similar analysis of the folding pathway of the R17 domain from spectrin showed some similarities and some differences between the two (*72*). For the early transition state, R16 has considerable helical structure in helix-A but rather little structure in the other two helices, in comparison R17 also has structure in helix-A but also shows structure in helix-C and some packing of helix-A with helix-C. Neither R16 nor R17 were found to have significant structure in helix-B in the first transition state. In the late transition state, there has been a consolidation of structure in both proteins; both R16 and R17 have increased interactions between helix-A and helix-C. In R17, helix-B is still largely unstructured, however, in R16 there are some interactions between helix-C and the N-terminus of helix-B (*72*).

C. All-β-Proteins

The folding pathways of many all-β-sheet proteins have now been characterized using protein engineering techniques. This includes studies on proteins with very different topologies: from the smallest possible of stable β-structures of the WW domains which have just three antiparallel β-strands, through the small β-barrel type structures adopted by cold shock proteins and SH3 domains, to the larger and more complex structures of the β-sandwich proteins with Greek key motifs.

1. SH3 Domains

SH3 domains are small, ubiquitous domains associated with a large number of proteins including many kinases which form a superfamily in the SCOP database (28). They are all-β proteins where the β-strands form an open β-barrel structure. As small, fast folding proteins, they have been model systems for studying folding pathways and in particular for using protein engineering and Φ-value analysis based approaches for the determination of structure in the transition state. Three groups have studied SH3 domains from different proteins—the Baker group have studied the SH3 domain from src, the Serrano group have studied the SH3 domain from α-spectrin in addition to studying the structural homologue Sso 7d, and the Davidson group studied the SH3 domain from fyn.

Early protein engineering studies on src SH3 domain established that structure in the transition state ensemble is highly polarized for this small protein with the hydrogen bonding network associated with two β-turns and an adjacent hydrophobic cluster formed, the rest of the protein being largely unstructured (73). The characterization of an additional set of mutants showed that the folding nucleus formed in the transition state was even more polarized than had originally been proposed (74). A double mutant analysis confirmed that the distal β-hairpin and the diverging turn are formed in the transition state and that all conformations in the transition state ensemble have the central three-stranded β-sheet formed (75).

Protein engineering studies on the SH3 domain from α-spectrin, published shortly after the first src SH3 study, showed a very similar pattern of Φ values with high Φ values observed in the distal loop β-hairpin and 3_{10}-helix (76). This study was conducted at various pHs where the protein had very different thermodynamic stabilities, however, the pattern of Φ values remained the same suggesting that there is little conformational variability in the transition state ensemble of the α-spectrin SH3 domain (76). The similarity in results between src and α-spectrin SH3 domains suggests that folding pathways of proteins may be evolutionarily conserved, and that topology may play an important role in determining the folding pathway of this structure, as had

been proposed by the Baker group (*73*). The Serrano group followed up their initial study with extended protein engineering studies to probe the folding pathway in further detail—a set of Tyr→Phe substitutions were made which established that hydrophobic residues outside of the folding nucleus could contribute to the stabilization of the transition state, suggesting a role for nonnative interactions (*77*). In addition, a series of core Val→Thr mutations were used to investigate the role of desolvation during folding. In this case, the folding kinetics of the mutants were found to vary with position of the mutated residue indicating that there are different degrees of desolvation in the transition state in the core of the protein (*78*). A protein engineering study of Sso 7d, a protein with a similar topology to the SH3 domain but which is not a member of the SH3 family, showed that it had very different structure in the transition state compared with either the src or α-spectrin SH3 proteins demonstrating that sequence features underlying topology can also play an important role in folding (*79*).

Forty substitutions of both large and small amino acids were made to residues in the hydrophobic core of fyn SH3 in order to investigate how tightly packed the hydrophobic core of this protein is in the transition state ensemble (*80*). In this case, mutations at three positions, which were designated as the folding nucleus, showed the largest decreases in folding rate and high Φ values consistent with this region being critical in stabilizing the transition state. However, mutation of residues further away from the nucleus to larger hydrophobic residues, which were thought to be in a more loosely packed region in the transition state, generally accelerated the folding rate despite destabilizing the native state of the protein leading to nonclassical Φ values (*80*). This study was followed up with another protein engineering analysis taking protein folding kinetics beyond the classical Φ-value approach. The effects of multiple amino acid substitutions at two positions in the structure on the rate of folding was assessed (*81*). The results supported earlier studies on src and α-spectrin SH3 that the diverging turn and the distal loop region play an important role in directing the folding of SH3 domains (*81*). In addition, they showed that the transition state ensemble is generally tolerant to amino-acid substitutions. The highly conserved Gly48 in fyn SH3 was investigated with another series of multiple amino-acid substitutions at this single position. Similar to the original study on hydrophobic core mutants, some of the mutants generated accelerated folding whilst destabilizing the native state (*82*). A strong correlation between folding rate and *β*-sheet propensity was observed indicating a possible nonnative *β*-strand conformation at position 48 in the transition state, in contrast to the conformation this residue adopts in the native structure (*82*).

For the fyn SH3 domain, the extensive protein engineering and folding studies by the Davidson group has lead to the creation of mutants with ideal properties for novel NMR experiments, which have expanded the experimental

techniques available for characterizing protein folding pathways. In this study, a lowly populated intermediate of a double mutant of fyn SH3 was characterized by relaxation–dispersion NMR methods (*83*). Subsequently, the technique has been used to establish that there are nonnative long-range interactions in a collapsed intermediate state of fyn SH3 (*84*) and even a Φ-value analysis of the three-state folding pathway of fyn SH3 has been undertaken using these NMR methods (*85*).

2. Cold Shock Protein

Cold shock protein B (CspB) is another small, rapidly folding protein for which a Φ-value analysis of the transition state for folding has been carried out (*86*). The structure of CspB consists of a three-stranded (*β*-1–*β*-3) and two-stranded (*β*-4 and *β*-5) mini *β*-sheet which pack together to form a closed *β*-barrel structure, and the transition state was found to be strongly polarized with only a few residues, particularly those in *β*-strand 1 and the subsequent turn, having high Φ values (*86*). Despite a high value for *β*-Tanford (which is a measure of the compactness of the transition state relative to the denatured and native states), most of the Φ values measured were low. Together, the results showed that there is some kind of energetic linkage between *β*-strands 1 and 4 in the transition state; thus, interactions between the two *β*-sheets have begun to form (*86*). Two of the residues which were shown to have high Φ values were partially exposed lysines located in *β*-strand 1. An additional study focused on these two residues and used double-mutant cycles with ionic strength dependence of folding rates to elucidate the role of electrostatic interactions on the folding pathway (*87*). For CspB, the Schmid group showed that long-range Coulombic interactions were important for organizing and stabilizing native-like structure early on the folding pathway (*87*).

3. WW Domains

WW domains are one of the smallest and fastest folding systems studied by protein engineering techniques. These three-stranded, antiparallel *β*-sheet structures range in size typically between 28 and 44 residues in length and have been studied extensively by the Fersht and Kelly groups. Early work on the folding of a wild-type and mutant WW domain established that the position of the transition state on the reaction coordinate was variable and temperature dependent (*88*). A Φ-value analysis of the single mutant investigated (W39F, a probe of the hydrophobic interface between *β*-2 and *β*-3) showed that the transition state moved from an early to a late position with a change in temperature (*88*). This tuning of the folding energy surface was also observed in a study of a different WW domain from Formin-binding protein 28 (FBP) (*89*). Two- or three-state kinetics were observed under different conditions induced by temperature, a C-terminal truncation of the protein, or by mutation (*89*).

A full Φ-value analysis was performed on the WW domain from Pin and 16 mutants characterized (*90*). Φ values were found to vary smoothly with sequence suggesting that local interactions dominated in the transition state, and loop 1 was found to be structured in the transition state. As with the other studies, a temperature dependence study on the effect of destabilizing mutations in loop 2 or in the hydrophobic cluster showed they could cause a switch from an early to a late transition state (*90*). The role of loop 1 in stabilizing the transition state and influencing the folding of WW domains was investigated in further detail by an additional study in which this six-residue loop was replaced with a shorter sequence that had a high propensity for forming a type-1 β-turn (*91*). Interestingly, the engineered variant folded faster and was more stable than the wild type demonstrating that the protein was not at its speed limit, however, the mutant was not functional (*91*). This result supported the idea that for many proteins there is a trade-off between fast folding/stability and function (*92*).

Protein engineering usually involves the substitution of one amino acid for another at a single, or sometimes multiple, positions within a protein. This is an incredibly powerful technique; however, it is limited to side chains—the backbone remaining unchanged in all of the mutants. The Kelly group, in collaboration with the Dawson and Gruebele groups, has used a completely different approach to study the effect of changes to the backbone of a protein on folding (*93*). WW domains are sufficiently small to enable chemical synthesis techniques to be used to make the polypeptide chain. For the Pin WW domain, a series of amide to ester substitutions were made at different positions along the backbone. This substitution eliminates a hydrogen bond as an NH group is replaced by an O; in addition, the hydrogen bond acceptor is weakened by replacing an amide carbonyl with an ester carbonyl. Characterization of 19 amides to ester mutants of the WW domain showed that native-like secondary structure forms in one loop in the transition state, but the rest of the backbone is less ordered (*93*). Highly complementary to the information that can be obtained from traditional protein engineering methods, these studies provided a detailed picture of how these small, fast folding domains fold.

Another chemical synthesis strategy for producing unusual variants of a WW domain was employed by the Fersht group (*94*). In this case, a thioether linkage was incorporated into the WW domain from yes kinase-associated protein (YAP65). The highly flexible thio-ether linkage alters the hydrogen bonding and dynamics of the loop and thereby acts as a probe of folding. Studies of the variants produced showed that the first β-turn has extensive structure in the transition state whilst the second turn/loop is unstructured (*94*). Despite the marginal stability of WW domains, a Φ-value analysis of the folding pathway of the FBP28 WW domain has been undertaken (*95*). The Φ values determined were highly polarized with many values being zero or one.

As with the other WW domains, β-turn 1 in the FBP28 structure has native-like structure in the transition state whilst the rest of the protein is relatively unstructured (*95*). Together with the results on the WW domains from Pin and YAP65, these results suggest that the folding pathways of this type of structure are highly conserved.

4. Greek Key—β-Sandwich Proteins

β-sandwich proteins that have a Greek key topology are formed by the packing of two antiparallel β-sheets to form a hydrophobic core. The core is always formed by the packing of the four central strands—B, C, E, and F, the number and position of the edge strands being quite variable. This type of structure is very common and found in a large number of proteins with different sequences and functions. The Clarke group have used this superfamily of proteins and developed a "fold" approach to study protein folding pathways. The strategy uses protein engineering techniques to characterize the folding pathways of different proteins within this superfamily, which are evolutionarily unrelated with little sequence similarity, but which have the same structure. The first β-sandwich protein to be characterized in detail using Φ-value analysis was TNfn3, the third fibronectin domain from human tenascin (*96*). 48 mutants which probed structure in the transition state at 32 different positions within the structure were characterized. High Φ values were found in the four central strands—B, C, E, and F, with residues with close contacts which formed the folding nucleus. High Φ values were also observed in C′ region and the EF loop, but residues further away from the critical core had decreasing Φ values (*96*). The results were consistent with a nucleation–condensation mechanism in which the folding nucleus was formed by a common core or ring of residues in the four central strands.

This study was quickly followed by that on another protein, the tenth fibronectin type III domain from human fibronectin, FNfn10 (*97*). Again, a large number of mutant proteins were studied in order to get a detailed picture of the folding pathway and structure of the transition state. In this case, Φ values of all the residues probed were fractional, with many residues in the central strands of the structure forming a large core of interactions in the transition state. However, local interactions in the turns and loops in FNfn10 are much less important. FNfn10 also folds by a nucleation-condensation mechanism but in contrast to TNfn3, FNfn10 has a more extended folding nucleus.

Extending the study to other members of the superfamily, the immunoglobulin-like domain I27 of human cardiac titin was also subjected to Φ-value analysis (*98*). In this case, the Φ values were generally found to be high, and few residues had Φ values of zero, indicating a very native-like transition state. Analysis of the mutational data was somewhat more complex than for

TNfn3 or FNfn10 as many mutations caused the transition state to become less native-like, that is, showed anti-Hammond behavior. However, residues involved in nucleating the folding of I27 were found to be structurally equivalent to those that formed the nucleus in TNfn3 or FNfn10 suggesting a common folding mechanism for all proteins which adopt this structure (*98*).

The role of loop length in the folding of an Ig domain, I27, was investigated by creation of a pseudo wild-type protein which had a five glycine insert engineered into the B–C loop (*99*). A Φ-value analysis on the pseudo wild-type I27 showed that the folding pathway was unchanged by altering the intersheet loop length. Thus, loop length is not critical in the formation of a folding nucleus made from long-range interactions (*99*).

Studies on the TNfn3, FNfn10, and I27 Greek key topology β-sandwich structures showed that within the Ig-like superfamily of proteins residues involved in forming the folding nucleus are highly conserved. These residues are slightly less conserved in the fnIII superfamily but four hydrophobic residues play a critical role. However, some members of the family, including CAfn2, lack one of these four critical residues (*100*). A Φ-value analysis of the folding pathway of CAfn2 showed that whereas the folding mechanism was very similar to that observed for the other β-sandwich proteins, the position of the folding nucleus within the hydrophobic core has moved (*100*).

Additional studies on hybrid proteins created by core and surface swapping between TNfn3 and FNfn10 have also revealed important aspects of the stability and folding of these types of structures (*101*). The hybrid proteins show structures very similar to the parent molecules, however, unexpected cross-talk between the surface and the core residues was observed (*101*).

CD2.d1 is another β-sandwich protein whose folding pathway has been characterized using different protein engineering strategies. CD2.d1 folds with three-state kinetics and populates an intermediate state during folding. First, a set of chemically conservative mutations was made within the core of the protein and the effect of these mutations on the energies of the intermediate, transition and native state were determined (*102*). Whilst the mutations all destabilized the native state, as expected as they all removed buried hydrophobic surface area from the core, many were observed to stabilize the partially structured intermediate state. A correlation between increased stability of the intermediate state and the intrinsic propensity of the amino acid to form β-structure was observed. It was concluded that although the side-chain interactions are weak in the intermediate state, the β-strand backbone is formed. In addition, the rate-limiting transition state is formed by a tightly localized nucleus of hydrophobic residues (*102*). A subsequent full thermodynamic analysis of these mutations on intermediate, transition, and native states reported that the mutations affect both the enthalpy and entropy of the different states, but that there are entropic compensations which lead to overall

small changes in the free energy of the system (*103*). The effects on the enthalpy and entropy were all lower in the intermediate and transition state compared to the native state (*103*). Hydrophobic groups have also been engineered at surface sites in CD2.d1 and their effect on folding measured. Interestingly, many of the mutations, which have little effect on the native state, are found to stabilize both the intermediate state and the rate-limiting transition state. Thus proving that nonnative interactions can play a role in the folding of small proteins (*104*). The fact that mutations that stabilize the intermediate state also stabilize the transition state and lead to faster folding suggested that, for CD2.d1, the intermediate state is on- and not off-pathway (*104*).

IV. Ψ-Value Analysis

In practice, many Φ values are fractional and these can arise from partial structure formation at the site of mutation in the transition state ensemble. However, they may also result from multiple folding pathways in which the site of mutation is structured in the transition state of some of the pathways but unstructured in others (*105, 106*). Using a series of mutations at a single site and either a Bronsted or Leffler analysis, these two situations can be distinguished (*105, 106*). For example, a Leffler plot of data obtained from a large number of mutations at position Glu24 in the fyn SH3 domain, which have a wide range of stabilities, is linear. Such behavior is indicative of a partially structured site in the transition state rather than a fully formed structure within just a subpopulation of the transition state ensemble (*81*). However, in some cases, it may be difficult or impossible to perform such an analysis.

An alternative protein engineering approach to the characterization of folding transition states which overcomes some of these problems has been developed by the Sosnick group (*107*). In this method, pairs of closely spaced histidine residues are engineered into solvent exposed sites in the protein structure such that they can bind divalent metal ions such as Co^{2+}. The metal ion binding stabilizes the native state of the protein and is equivalent to the energetic perturbation of the system by mutation used in Φ-value analysis. The thermodynamics and kinetics of folding of the engineered bihistidine mutants in the absence and presence of varying concentrations of metal ion are measured and the relative effects of metal-ion binding on the stability of the transition and native states calculated through use of a Ψ value. The Ψ values are similar to the Φ values described above. The advantages of this approach is that by measuring the effects over a wide range of metal-ion concentrations rather than a single concentration (which is effectively what is done for a single site-specific mutation in Φ-value analysis), the analysis is able

to evaluate the shift in the transition state ensemble resulting from the stabilization and thus able to distinguish between multiple folding pathways and a broad transition state ensemble (*107*).

Ψ-value analysis has been applied to much fewer proteins than Φ-value analysis and is, therefore, in some respects less well tested. It has been used on the GCN4 coiled coil (*108*) and ubiquitin (*109, 110*). For the dimeric version of GCN4, there was already evidence from protein engineering studies and Φ-value analysis that fractional Φ values were likely due to multiple pathways (*111*). A Ψ-value analysis, in conjunction with Ala to Gly mutations, provided further evidence for heterogeneity in folding pathways and the transition state ensemble (*108*). In contrast, a Ψ-value analysis of an effectively monomeric, cross-linked version of GCN4 showed that the folding pathway was much more homogeneous and that nucleation occurred at whichever end of the protein was cross-linked. A Ψ-value study on the folding of ubiquitin used 14 bihistidine pairs to characterize the transition state ensemble and found that much of the native-state topology was present in the transition state with four out of the five β-strands and the α-helix having some structure (*109*). However, six bi-His sites gave rise to fractional Ψ values suggesting these were in regions which only had partial structure in the transition state. These results implied that there was more extensive structure in the transition state ensemble than had been characterized using Φ-value analysis, where the C-terminal β-strands were not observed to have any extensive structure (*26*). Some criticism of the analysis methods used in this paper were made (*112*), however, these were later addressed (*110*). For ubiquitin, the original Φ-value analysis was consistent with many previous studies on peptides from the sequence which showed that the C-terminal region of the protein did not have any tendency to form structure in the absence of the rest of the protein in contrast to peptides corresponding to the N-terminal region of the protein which were shown to be structured under a variety of different conditions (*25*).

Although the two different protein engineering methods produced slightly different sets of results for ubiquitin, they can easily be rationalized—it appears that within the transition state ensemble, the N-terminal region of ubiquitin has robust structure which is present in all the species present, whilst the C-terminal structure is much less stable and easily destabilized and abolished by mutation.

Both Φ- and Ψ-value methods provide valuable information on the nature of folding transition states, and the information obtained is somewhat complementary in nature. Φ-Value analysis has been used extensively to probe the formation of the hydrophobic core in folding transition states, something which Ψ-value analysis can not do. Both techniques can probe the formation of secondary structure—both α-helix and β-sheet—through the use of carefully designed and positioned mutations. Although it is somewhat more

straightforward with Ψ-value analysis to establish heterogeneity in pathways and the transition state ensemble, this is also possible with Φ-value using multiple mutations and Bronsted/Leffler plots. As with any strategy which introduces mutations into a protein which perturb not only the energy of the system but also potentially the structure, care must be taken both in the design of mutants and interpretation of results. In addition, it has become clear that with any protein engineering method only when a large number of mutations are used is a detailed and reliable picture of the transition state obtained.

V. Proline Isomerization

It has long been recognized that proline isomerization can result in slow phases in refolding reactions due to the *cis-trans* isomerization of proline residues in the denatured states of proteins (*113–115*). For proteins that contain many proline residues, such as collagen (*116*), or which have a proline peptide bond in a *cis* configuration in the native state (*117, 118*), then proline isomerization can dominate the folding kinetics. Protein engineering methods have played an important role in characterizing the slow, proline isomerization processes associated with protein folding, enabling the assignment of slow, proline-isomerization events to specific proline residues within a structure, thereby addressing many questions on the nature of the isomerization event and the degree to which it is coupled with folding. In the simplest case, substitution of a single proline within a structure to any other amino acid should result in the disappearance of a slow phase in the refolding reaction which is, therefore, assigned to the isomerization of that particular proline. For some proteins, this has been observed—for example, ribonuclease A has two *cis* prolines (Pro93 and Pro114), the double mutant P93A, P114G showed only single exponential refolding kinetics and no evidence of proline isomerization limited processes (*119*), and mutation of the *cis* proline at position 110 in CheY resulted in the disappearance of the major slow refolding phase (*120*).

In other cases, however, replacement of a proline with another amino acid does not simply remove an apparent proline-isomerization folding event. In CRABP-I substitution of Pro85 with an alanine resulted in the disappearance of a slow folding phase, however, substitution with valine had little effect on the refolding kinetics suggesting that the loss of a refolding phase upon mutation of a proline is not always diagnostic of a proline-isomerization limited phase (*121, 122*). For pectate lyase C, which contains many proline residues including Pro220, which is in a *cis* conformation in the native state, mutation of Pro220 to alanine resulted in the loss of one of the slow folding phases, but mutation of the other 11 proline phases all showed slow phases indicating that a single proline residue is not responsible for the second slow refolding phase (*123*).

Other studies have shown dramatically different effects of mutating different proline residues illustrating that the effect of prolines is highly context dependent (*124*).

Prolines can also sometimes be found in both *cis* and *trans* conformations even in the native states of proteins leading to heterogeneity and complex unfolding behavior. Again, protein engineering of these prolines has, in some cases, lead to a simplification of the unfolding kinetics. This has been shown for staphylococcal nuclease (*125*) and suc 1 (*126*).

VI. Repeat Proteins

Complementing studies on the folding of globular proteins or domains, various protein engineering techniques have been used to investigate the folding mechanisms of topologically simple repeat proteins. These typically consist of 20–40 residue tandem repeat motifs that stack together in an approximately linear fashion to form nonglobular, elongated, and super-helical structures, as shown in Fig. 4 (*127, 128*). Repeat proteins are ubiquitous, with 20% of proteins coded for in the human genome likely to contain multiple repeats (*129*), and they are frequently involved in a large variety of protein–protein interactions, mediating molecular recognition in numerous biological processes. They can be subdivided into classes according to the repeating structural unit, for example, ankyrin (ANK) repeats contain tandem pairs of antiparallel α-helices, tetratricopeptide (TPR) repeats have a α-helix-turn-α-helix motif, and the leucine-rich repeat (LRR) forms a β-strand-loop-helix structure, as shown in Fig. 4 (*127, 128*).

Due to their unique structural properties, repeat proteins represent an interesting folding problem. Unlike the numerous long-range contacts found in globular-protein topologies, they contain mostly regular, short-range interactions, and lack an obvious hydrophobic core as well as the stabilizing tertiary contacts between residues that are distant in sequence. The interactions that define the functional protein can, therefore, be divided into two main types: the packing within each repeat motif and the stacking interactions between adjacent repeat motifs. This allows them to be effectively "dissected," it being relatively straightforward to separate out the local energetic contributions of a particular residue (*129*). Repeat proteins can also be extended in size by the addition of a neighboring domain, permitting the effect of the number of repeats on folding and stability to be examined. Knowledge about the folding of single repeats is useful for designing stable and highly active scaffolds. Additionally, a number of repeat proteins have been associated with disease states (*130, 131*), the cause of which could be related to their folding

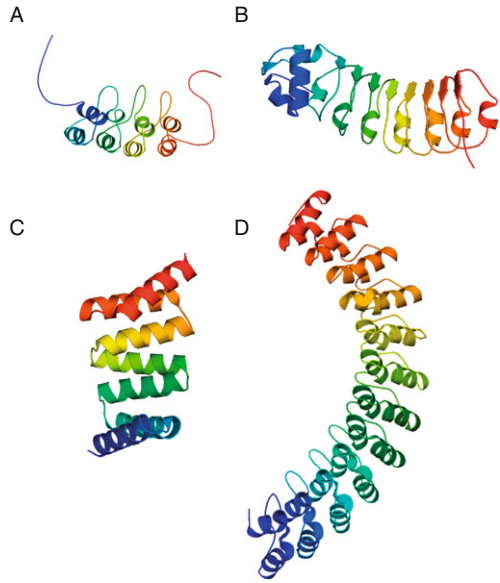


FIG. 4. Examples of common repeat protein motifs: (A) The tumor suppressor protein p16 containing four ANK repeats, (B) The leucine-rich repeats and α-helical capping domain of InIB, (C) A designed repeat protein containing 3.5 consensus TPR motifs, (D) D34, a 426-residue domain consisting of the last 12 ANK repeats of Ankyrin R and the largest repeat motif to have its folding properties characterized to date. Proteins are colored blue to red from amino to carboxy terminus. Ribbon representations were generated using PyMOL (www.pymol.org).

properties. To probe or alter the folding mechanisms of repeat proteins, a variety of protein engineering strategies from single-site mutations to the addition or deletion of whole repeats have been undertaken.

Proteins containing naturally occurring ANK motifs have been the most extensively studied in terms of their folding properties. This repeat is one of the most common structural motifs found in proteins and, like other repeats, forms the scaffold for specific, high-affinity molecular interactions (*132*). Each ANK repeat comprises 33 amino acid residues encompassing a pair of antiparallel helices that are connected to the preceding ANK domain by a β-loop. Usually, only four to seven repeats stack to form an elongated structure, but up to 29 repeats have been found in a single protein (*132*). The first ankyrin-repeat-containing protein to be characterized in detail was the tumor suppressor p16, which consists of four ANK repeats and is involved in the growth control pathway in eukaryotic cells (Fig. 4) (*131*). p16 is both thermodynamically and kinetically unstable, and therefore particularly susceptible to inactivation by

the single-point mutations associated with cancer (*131*). Itzhaki and colleagues analyzed the folding pathway of p16 using Φ-value analysis and found that, unlike many globular proteins, the folding mechanism is highly polarized with the repeats folding sequentially. The two C-terminal repeats, thought to be important for maintaining the structural scaffold, are fully formed in the rate-determining transition state, whereas the N-terminal repeats remained largely unstructured (*131*). Simulation studies performed later gave results consistent with this observation (*133*). The related five ANK repeat protein p19 has been shown to fold *via* a similar folding mechanism, elucidated using truncated variants of the protein, and an on-pathway intermediate in which only the last three C-terminal repeats are structured is populated during folding (*130*). Interestingly, p19 is more stable than p16; the extra stability is thought to arise from the additional fifth repeat (*130*). Lowe and Itzhaki using Φ-value analysis have mapped the folding pathway of the four-ankyrin repeat protein myotropin in detail (*134*). Like p16 and p19, the folding of wild-type myotropin is initiated at the C-terminal repeats. However, the folding kinetics of some mutants revealed a parallel folding mechanism, indicating that the wild-type folding pathway can be easily perturbed by single-site mutations. An alternative folding mechanism in which the transition state structure is polarized at the other end of the molecule was determined; the authors concluded that destabilizing mutations in the C-terminal repeats reduce the flux through the wild-type pathway and consequently folding is initiated at the N-terminal repeats (*134*). Therefore, in contrast to globular proteins that tend to have a well-defined path between native and denatured states, the symmetrical structure of myotropin leads to more potential for folding to be initiated at multiple sites. Lowe and Itzhaki describe this as "folding on a fulcrum"; specifically, when one end of the molecule is destabilized, folding is shifted to a different nucleation site (*134*). More recently the folding of the leucine-rich repeat domain of Internalin B (InIB), which is composed of seven LLR units and an N-terminal capping motif, has been studied by examining the effects of individual point mutations on the folding pathway (*135, 136*). This Φ-value analysis showed that the folding of InIB nucleates in the N-terminal helical-capping domain and propagates towards the C-terminus, thus providing evidence to suggest the cap serves as a nucleus onto which the folding of nearby β-sheet LRR motifs can propagate.

ANK proteins have been manipulated at the genetic level to examine the effect of deleting terminal repeats (*137–139*). Studies on peptide fragments of p16 indicate that a polypeptide consisting of the two C-terminal ANK repeats, the same part of the protein that is formed in the folding transition state (*131*), is cooperatively folded (*138*); this appears to be the minimum independent folding unit, as fragments of a size less than this remain unfolded. Polypeptides that contain different numbers of repeats from the *D. melanogaster* Notch

ankyrin repeat region, a 268-residue domain containing seven ankyrin repeats, have also been characterized thermodynamically (*139*). The inclusion of the C-terminal seventh repeat leads to a significant increase in stability, suggesting that this repeat might act as a "capping" motif to interact favorably with the sixth repeat and solvent. In an extension of this study, the stability of a series of engineered Notch proteins with one or more of the seven ANK repeats deleted was examined (*137*). Barrick and co-workers determined from the stabilities of each construct that cooperative folding of the protein arises as each repeat is intrinsically unstable but strongly stabilized by repeat-repeat interactions. The effect of duplicating and deleting internal Notch ANK repeats has also been examined by the construction of a series of variants where the repeat number is altered by the duplication or removal of internal repeats (*140*). Unlike the wild-type construct which folds in a cooperative manner, multistate folding was observed for proteins with more than one duplicated internal ANK repeat. Deletions of internal repeats caused a large destabilization, more than that resulting from deletion of an end repeat. The authors suggested that this was due to unfavorable interactions between non-native interfaces, and therefore highlighted the importance of repeat-repeat interactions in the folding mechanism (*140*).

The origin of cooperativity in repeat proteins has been investigated. Barrick and co-workers have made use of the nonglobular, modular structure of Notch to examine its folding energy landscape by measuring the distribution of stability throughout the repeats (*141*). Destabilizing mutations were introduced to cause an uneven folding energy landscape; they found that when the distribution of stability over the protein is not uniform, cooperative folding no longer occurs. Simulation experiments performed on natural ankyrin repeat domains suggest that as the number of ANK repeats increases, folding is no longer cooperative but decoupled due to the ever-increasing entropic advantage of a broken repeat-repeat interface (*142*). Results from studies on D34, a 426-residue domain consisting of the last 12 ANK repeats of Ankyrin R and the largest repeat protein to have its folding properties examined to date, are consistent with this theory (*143*). Unlike most smaller ANK repeat proteins that unfold in a two-state manner when under thermodynamic control, the equilibrium unfolding of D34 involves an intermediate. Werbeck and Itzhaki engineered a set of mutants involving single-site mutations throughout the protein and found that, depending on the position of the perturbation, the number of repeats that are unfolded in the intermediate was altered. Thus, the folding energy landscape for domains containing a large number of repeats appears very rough, and, unlike globular proteins, small destabilizations in one part of the protein can cause a different folding intermediate to be populated (*143*).

The simple modular nature of repeat proteins makes them useful models for protein design, a more ambitious protein-engineering project compared to the single point mutations used in a Φ-value analysis. Ideal "consensus" repeats have been designed to investigate the role of the amino acid sequence in specifying the topology, stability and folding of repeat proteins (*144*). The high abundance of repeat motifs has allowed a statistical analysis of many sequences to design a consensus repeat, permitting the key residues that code for either fold or function within each repeated module to be identified. Consensus sequences for ANK, TPR, LRR, and HEAT motifs have all been successfully designed (*145–151*). In several studies, it was found that a critical number of repeats was required in order to obtain a folded repeat protein (*147, 148*). N- and C-capping repeats or helices have been incorporated in some designs to increase the solubility of the designed proteins by avoiding solvent exposure of hydrophobic residues at the end of the repeat (*145, 147*). In general, designed repeats are more thermodynamically stable than their natural equivalents (*150, 152*). They can also be added to naturally occurring repeat domains to increase their stability (as has been shown with Notch (*153*)), or, in the case of myotrophin and the ankyrin domain from IκBα, residues in a natural repeat motif can be mutated to those of the consensus (*134, 154*). Interestingly, the consensus repeats of the all-helical TPR proteins designed by Main and co-workers all fold rapidly with rate constants that agree with those predicted from their contact order, in contrast to what is often observed with natural repeats (*155*).

VII. Circular Permutants

The use of protein engineering to manufacture a library of nondisruptive, single point mutations strategically placed throughout a protein to analyze its folding provides important information about the interactions of the side chains of individual residues in the transition state (*18*). A more extreme protein modification involves rearranging the order of secondary structure elements using the technique of circular permutation. This requires joining the natural N- and C- termini with a new covalent bond while cleaving the original polypeptide chain at an appropriate position to create new termini (Fig. 5) (*156, 157*). It follows that this adjustment is suited to proteins with natural termini that are in close proximity. Circularly permuted mutants can yield information about protein-folding mechanisms complementary to that obtained from Φ-value analysis; permutants can probe the role of the natural

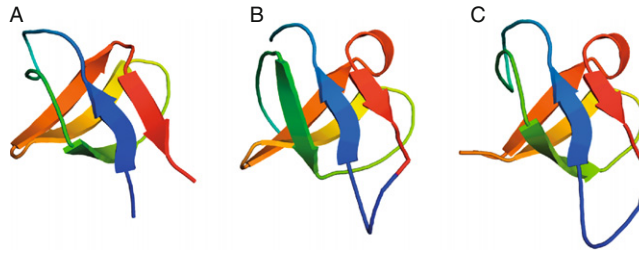


FIG. 5. Crystal structures of circular permutants (A) wild-type α-spectrin SH3 domain and two circular permutants, (B) S19-P20s, and (C) N47-D48s, PDB codes 1SHG, 1TUC and 1TUD, respectively. Wild-type SH3 is colored from blue to red from N- to C-terminus, and equivalent structural regions in each circular permutant are colored to match. Ribbon representations were generated using PyMOL (www.pymol.org).

N- and C-terminal regions, can be used to examine the role of topology in protein folding reactions, and how the order of secondary structure elements and local interactions affects the folding process.

The first reported circular permutant was of bovine pancreatic trypsin inhibitor; this was created post-translationally using chemical condensation to join the original termini, followed by trypsin cleavage to generate new termini between residues Lys15 and Ala16 (*156*). Luger and co-workers later developed a more adaptable method by circularly permuting the corresponding gene at the genetic level, successfully utilizing this approach to create active permuted forms of phosphoribosyl anthranilate isomerase (*157*). Using this method, the effect of altering the connectivity of secondary structural elements in a systematic manner by generating a series of circular permutants of the same protein was examined (*157*).

Circular permutation has been used to examine the folding mechanisms of several proteins in detail including the Src-homology domain (SH3) from α-spectrin, chymotrypsin inhibitor 2 (CI2), DHFR, and the ribosomal protein S6. Folding studies on circularly permuted forms of α-spectrin SH3, a small, 62-residue single domain protein consisting of an orthogonal β-sandwich with three β-hairpins were the first to exploit this type of protein modification (Fig. 5A) (*158*). Viguera and co-workers disrupted the connectivity of all of the β-hairpins in the SH3 structure by constructing permutants with new N- and C- termini in β-turns or loops. This effectively converted some local interactions present in the native structure of the wild-type protein to long-range interactions in the permutants, and, therefore, probed the importance of local interactions between any of the β-hairpin strands in guiding the SH3 folding process. That all the SH3 permutants were able to fold to a native conformation similar to the wild-type protein indicated that neither the order

of secondary-structure elements, nor local interactions caused by the presence of any of the β-hairpins, were necessary for the protein to fold (*158*). Changes in the folding rates of the permutants, however, suggested that they were folding with a different mechanism to the wild type. Evidence to support this hypothesis came from a further study that examined the effect of single point mutations made at equivalent positions throughout two of the SH3 circular permutants, S19-P20s, with new termini in a long, irregular loop between residues 19 and 20, and N47-D48s, which has a cut in the regular distal hairpin between residues 47 and 48 (Fig. 5A) (*159*). Analysis of the kinetics of these mutants suggested that, unlike the native structures, the transition states of wild-type, S19–20s and N47-D48s SH3 are notably different. The consequence of differing loop lengths to join the original termini of the same two SH3 permutants was also investigated (*160*). Elongated forms of the original permuted proteins S19-P20s and N47-D48s were constructed with one, three, and five extra glycine residues added to the loop joining the natural SH3 termini. This extension resulted in a small destabilization that the authors attributed to an entropic effect. In agreement with previous work, Eyring plot analysis of the permutants suggested that the thermodynamic nature of their transition state structures was different to the wild-type SH3 domain (*159, 160*).

The studies described above performed on circular permutants of an SH3 domain indicate that the transition state structure of the protein is dependent on the connectivity of the secondary-structure elements. This, together with a protein engineering analysis that determined the transition state structure of the α-spectrin SH3 domain under different stability conditions and compared it to that of the evolutionary related Src SH3 domain, led to the proposal that the folding transition state of SH3 domains is evolutionary conserved (*75, 76*). Martinez and co-workers suggested that this is likely due to the topological constraints of the SH3 fold, and the apparent importance of chain connectivity in determining the folding mechanism of SH3 domains (*76*).

To test the hypothesis that adjusting the connectivity of secondary structural elements in a protein can alter its transition-state structure, a study characterizing the folding pathway of a permuted version of the 64-residue protein CI2 was undertaken by Otzen and Fersht (*161*). CI2 was permuted post-translationally by disulfide bond formation of engineered cysteine residues at positions 3 and 63, followed by cleavage of the scissile bond between Met40 and Glu41. Analysis of the kinetics of eleven mutants tactically placed in the CI2 circular permutant showed that, despite altering the connectivity of the protein, the folding nucleus was retained. The authors suggested that the different response of CI2 and SH3 to permutation arises from the fact that the SH3 permutants were cleaved in a position within the protein folding nucleus whilst the loop containing Met40-Glu41 in CI2 is not structured in the transition state; it is interesting to note that CI2 permutants that involve

breaking bonds other than Met40-Glu41 did not fold. Therefore, the effect of permutation on the folding nucleus is likely to be protein specific, and depend on its amenability to permutation (*161*).

In an alternative approach, Iwakura and co-workers performed a systematic circular permutation analysis on DHFR, a monomeric, two-domain protein, to identify regions of the sequence that are essential for the protein to fold (*162*). They sequentially broke every pair of residues in the protein to create a library of permutants. Cleavage at certain positions resulted in the inability of DHFR to fold; and when this occurred, the authors concluded that their connectivity is essential for the folding process. They called these areas "folding elements"—polypeptide segments crucial for the protein to fold. Ten such folding elements were found in DHFR, and it was proposed that these have an important role in the early stages of folding (*162, 163*). A similar systematic circular permutation analysis was performed on the 189–residue, monomeric disulfide oxidoreductase DsbA to identify regions of the protein essential for successful folding and stability (*164*).

More recently, investigations on circular permuted forms of the small 101-residue $\beta-\alpha-\beta$ ribosomal protein S6 from *Thermus thermophilus* (Fig. 5D) have revealed some important features of its folding mechanism. Both the transition state of S6 and the relationship between the folding rate and the average sequence separation between contacting residues (contact order) can be altered by circular mutation (*165–167*). Oliveberg and co-workers examined the effect of permutation on the transition state structure of S6 in some detail (*165*). Wild-type S6 has strong interactions in the transition state between residues that are distant in sequence; this entropic driving force causes all parts of the protein to fold together with a diffuse transition-state structure. In contrast, the transition state of a circular permutant cleaved between residues 13 and 14, designed to have strong contacts between residues that are close in sequence, is locally condensed and polarized towards the linker between the original wild-type N- and C-termini. The authors concluded from these data that the diffuse transition state structure of wild-type S6 may have been selected for a biological advantage as it is not a requirement for successful folding (*165*).

In a more extensive study involving a transition-state analysis of four topological variants of S6 from *T. thermophilus*, Lindberg and co-workers showed that the folding pathway of the protein can be altered systematically by circular permutation; changes in the Φ values of the transition-state structure can be directly related to the change in sequence separation between the interacting residues (*168*). Intriguingly, the folding nuclei of all the S6 permutants studied retained a common structural motif of one helix docking against two β-strands (*168*). This nucleation motif is also seen in the transition state structure of the evolutionary divergent S6 protein from *Aquifex aeolicus* (*169*).

These results have led to the concept of "foldons"–independent, cooperative structural units involved in folding that, unlike the folding elements proposed by Iwakura and co-workers from their studies on DHFR (*170*), need not necessarily be contiguous in sequence. In the case of S6, the foldon is a two-strand-helix motif (*168*). Lindberg and Oliveberg have proposed that a protein can be composed of competing foldons, and the mechanism by which it folds will depend on the number of these within its structure and the amount of overlap between them (*171*).

VIII. Multidomain Proteins

Protein domains can be described as modular units with distinct structural, functional, or evolutionary properties (*172*). They may exist as small, single-domain, monomeric proteins, but more often are found as part of larger proteins composed of numerous domains. Indeed, analysis of sequenced genomes suggests that at least two thirds of eukaryotic proteins contain more than one domain (*173*). A multidomain protein can fold from a single polypeptide chain, or exist as an oligomeric protein composed of domains from different chains that associate. In comparison to single domain proteins, multidomain proteins additionally contain one or more interdomain interface, the effect of which is important when considering their folding properties.

Presently, the majority of protein-folding studies have focused on easily manipulated, single-domain monomeric proteins as they represent simple folding systems (*174*). Such studies have led to the development of many models for the different mechanisms by which small proteins fold (*175–177*). It is often assumed that these models will also be applicable to the folding of multidomain proteins, however, the effect of a neighboring domain on a protein's folding properties may be significant. It is, therefore, of importance to establish the degree to which folding mechanisms generated for small, single-domain proteins can be applied to the folding of larger, multidomain proteins (*172*).

Studying the folding of single-chain, multidomain proteins allows additional questions about folding mechanisms that cannot be answered by examining single domains alone to be probed (*172*). Some issues that have been addressed are the effect of a domain on the stability, folding rate and pathway of a neighbor domain, and the possible biological implications of this. To investigate the effects of a neighboring domain, a single domain must be characterized thermodynamically and kinetically in isolation as well as in the context of the multidomain protein (*172*). Manipulation at the genetic level using protein engineering allows a protein domain to be expressed in any number of different arrays, either on its own or as part of a natural or artificial multidomain

construct (Fig. 6) (*178*). The biophysical characterization of multidomain proteins can be challenging, however, as their kinetics and thermodynamics are often complicated (*172, 179, 180*). Furthermore, when considering a domain in isolation, it can be difficult to define where it ends and another begins; it has been shown that unnatural shortening of domains can lead to a decrease in stability or different folding properties (*181*). Finally, separate domains are often insoluble when expressed individually due to exposure of hydrophobic domain–domain interfaces (*172*).

To date, there are limited examples of monomeric multidomain proteins that have been studied comprehensively. In general, systems that have been characterized kinetically and thermodynamically in single and multidomain forms fall into two categories: those where interdomain interactions have an effect on folding and those whose domains fold independently (*172*). The folding of chicken brain α-spectin is one of the most extensively characterized by the Clarke group. Spectrin repeats are composed of a 106-residue, three-helix bundle with domains connected by an uninterrupted helix causing the C-terminal helix of one domain to be contiguous with the N-terminal helix of

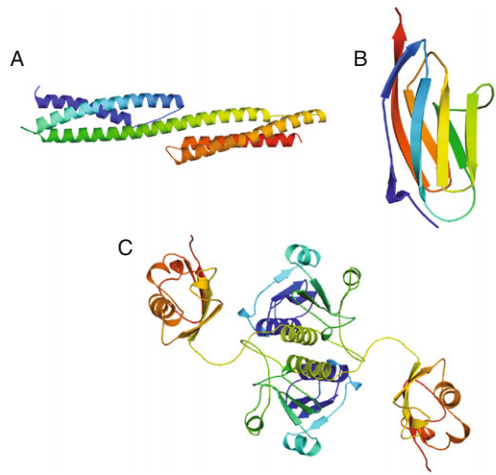


FIG. 6. Structures of multidomain proteins. (A) The two-domain spectrin fragment R1516 (PDB ID 1U5P). The domains are connected by an uninterrupted α-helix, (B) A titin domain (PDB ID 1TIT). There is no structure of the multidomain protein, but NMR and SAXS experiments suggest that domains have no significant interface and are connected by a flexible linker (*253*), (C) An artificial multidomain protein of the dimeric knotted protein YibK-ThiS, engineered to investigate the effect of an additional domain on its folding properties. The structure was determined using SAXS measurements (*254*). Proteins are colored blue to red from N- to C-terminus. Ribbon representations were generated using PyMOL (www.pymol.org).

the next (Fig. 6A). Interdomain interactions appear to affect both the stability and folding kinetics of the individual spectrin domains, specifically they stabilize the native state and slow down the rate of unfolding (*179, 182, 183*). Examples of other multidomain proteins where this has been observed include yeast phosphoglycerate kinase (*184*), protein S (*185*), and the scFv fragment (*186*). Interestingly, all these proteins exhibit densely packed domain interfaces (*172*).

The folding of pairs of α-spectrin domains have been investigated by Clarke and co-workers, specifically, they have compared the folding of the 15th and 16th repeats (R1516) and the 16th and 17th repeats (R1617) to that of the individual domains R15, R16, and R17. Equilibrium studies have demonstrated that spectrin domains are stabilized by their neighbors (*179, 182*). For example, in R1617, the R17 domain is stabilized by some 3 kcal mol^{-1} by a folded R16 domain and likewise the R16 domain is more stable in the presence of a neighboring folded or unfolded R17 domain (*182*). The folding kinetics of R1617 are complex, and studies show that domain–domain interactions have significant effects on the kinetic behavior of the individual domains (*183*). In particular, the presence of a folded N-terminal R16 domain notably speeds up the folding rate of the following R17 domain. Similarly, the presence of a folded R15 domain speeds up the folding of an R16 domain six-fold (*179*). The authors suggest that this may be to promote more efficient cotranslational folding, and prevent misfolding in the cell (*183*). In a study to compare the folding properties of R1516 and R1617, Batey and Clarke demonstrate that to accurately predict the effect of one domain on its neighbor both equilibrium and kinetic data are required; the relative rates at which the constituent domains fold dictates their behavior and apparent cooperativity under thermodynamic control (*179*). Interestingly, when the transition-state structure of the individual R16 domain was examined in isolation and when part of the R1516 spectrin-repeat pair using Φ-value analysis, the folding pathway appeared the same (*187*). This result suggests that although interdomain interactions affect the folding rates and the cooperativity of folding of spectrin domains, they do not affect the folding mechanism. Thus, folding principles gained from the study of single domains may be applicable to multidomain systems (*187*).

In contrast to spectrin, there are some examples of single-chain, multidomain proteins whose component domains appear to fold independently, and the kinetic and thermodynamic properties of a particular domain are not altered by the presence of its neighbor. All *β*-sheet immunoglobulin domains in the giant muscle protein titin (Fig. 6B) behave in this manner, and the protein is said to behave as "the sum of its parts" (*188*). This is thought to contribute favorably towards the elastic properties of the protein, which require it to unfold to fulfill its function, as the unfolding and potential aggregation of adjacent domains would be less likely. Other independently folding domains

have been observed in fibronectin type III (*189*) and SH3-spectrin protein (*190*). As noted by Han and co-workers, these are all examples of structural proteins with small interdomain interfaces (*172*).

The above studies highlight how the use of protein engineering techniques to analyze the effect of neighboring domains in multidomain proteins has yielded some important results. The relevance of single domain studies to understanding the general principles that control the mechanisms of protein folding depends on whether the stability or folding pathway change when a domain is attached to its neighbors. Experiments on spectrin demonstrate that not all multidomain proteins are a simple "sum of their parts" as the fully folded spectrin protein is more stable thermodynamically and kinetically than the isolated domains (*179, 182, 183*). The ability of domains in multidomain proteins to fold independently appears to be dictated by the size of the interdomain interface; domains of multidomain proteins fold independently when the interaction between the domains is limited, which in turn is likely to be related to their functional role (*172*).

Protein engineering can be used to determine folding mechanisms of multidomain proteins that are oligomeric in nature consisting of two or more domains from separate polypeptide chains. A number of studies have focused on the folding of oligomeric protein systems, the simplest of which are dimeric proteins that require both folding and association to occur. Proteins that exist as dimers have been shown to fold with a variety of different mechanisms (*191*). One of the first oligomeric proteins to have its folding pathway characterized was the 53-residue P22 Arc Repressor dimer; it folds with a cooperative, two-state mechanism (*192, 193*). The simple folding kinetics of the protein allowed Milla and Sauer to generate a library of mutant proteins to probe the transition state structure, similar to a Φ-value analysis performed on a monomeric protein (*192*). However, it is often the situation that the folding kinetics of dimeric proteins are complex, mainly due to separate, non-cooperative, folding and association events. For example, SecA from *Escherichia coli* (*194*) and the knotted protein YibK from *Haemophilus influenzae* (*195*) are both dimers whose complex folding kinetics demonstrate that there are at least three intermediates on the folding pathway. It is often useful to make rational mutations aimed at disrupting the association of a protein–protein interface and produce a monomeric variant of the dimeric protein. The folding kinetics of the monomer is often simplified, and thus any association steps in the folding mechanism can be identified by their absence in the monomer kinetics. Examples of the successful application of this methodology include the construction of monomeric forms of YibK (*196*) and tryptophan repressor (*197*).

Lastly, it is possible to construct unnatural multidomain proteins to probe the folding of a natural protein domain (Fig. 6C). An additional domain can be fused to the N- or C-terminus of the protein of interest. The effect of this

additional domain on the natural protein's stability and folding kinetics can then be monitored. This technique was used by Mallam and Jackson to examine how protein knot formation occurs during the folding of the knotted proteins YibK and YbeA, both of which contain a deep trefoil knot formed by the threading of the polypeptide backbone. The small protein ThiS was fused onto both termini of the knotted proteins and the effect on the folding kinetics monitored to show that knotting of the polypeptide chain is not rate limiting during folding (*198*). Additionally, Randles and co-workers constructed spectrin-titin pairs to show that nonnatural neighboring domains can cause a spectrin domain to be stabilized and alter its kinetic properties (*199*). In both cases, the additional domains ThiS and titin were chosen due to their resistance to urea denaturation such that they remain folded under the conditions required to unfold the domain of interest, thus allowing its folding to be monitored exclusively.

IX. Engineering Probes of Folding into Proteins

A number of studies have engineered tryptophan residues into a protein structure in order to probe the folding of the protein in that region or simply to act as a probe of the global folding/unfolding of the protein—tryptophan fluorescence usually being considerably stronger than tyrosine fluorescence and a sensitive probe of the state of the protein. One of the first examples was ubiquitin, the F45W mutant produced being comprehensively characterized (*200*) and subsequently extensively used as a pseudo wild-type for many folding studies, see Sections III and IV. Very recent studies that have engineered tryptophan into other sites in the structure of ubiquitin have generated intriguing results (*201*). Whilst the F45W mutant shows simple two-state kinetics under most conditions (*202, 255*), some of the other tryptophan mutants, particularly double tryptophan mutants, show much more complex folding behavior indicative of the presence of intermediate states (*201*). In this case, it is not known whether this could be a general method of detecting transient, lowly populated intermediate states or whether the introduction of large tryptophan residues in a small protein distorts the folding energy landscape.

A tryptophan has been engineered into the monomeric form of the λ repressor to act as a fluorescent reporter of folding. Kinetic experiments on this variant confirmed previous results from NMR line shape analysis that established that this small α-helical protein folds very fast with two-state kinetics (*203*). The Oas group have also used this approach with another ultra-fast folding protein, the B-domain from protein A (BdpA) (*58, 204*).

Wild-type L-lactate dehydrogenase (LDH) has three tryptophan residues and all these were first mutated to tyrosine before reintroducing single tryptophans at specific positions throughout the structure of the protein to incorporate different structural probes of folding. The unfolding of nine single Trp mutants of LDH was used to identify four intermediate states (*205*). In a separate study, phosphoglycerate kinase (PGK) was also engineered with two tryptophan groups to report on folding (*206*).

A similar approach has been taken for the cellular retinoic acid binding protein I (CRABPI) where the wild-type protein contains three tryptophan residues. Mutants of CRABPI have been made which contain only one of the three tryptophans, and these mutants have been used to investigate the folding of different regions of this predominantly β-sheet protein (*207, 208*). In this case, a folding mechanism in which there is an initial rapid hydrophobic collapse, followed by the formation of specific interactions which restrict the arrangement of the chain topology, in particular between the N- and the C-termini and which result in burial of Trp7, was proposed. The development of native interactions occurs only late during folding (*207, 208*). The same approach has been used for other proteins that adopt this β-clam structure—rat intestinal fatty acid binding protein (IFABP) and rat ileal lipid binding protein (ILBP) (*209–211*). A similar folding mechanism to CRABPI has been proposed—an initial collapse of the polypeptide chain around a hydrophobic core, formation of part of the β-sheet structure by propagation of this core, and finally formation of the rest of the β-sheet structure (*211*). Certainly results on IFABP show that structure around Trp82 forms very early, prior to formation of any native-like structure in the periphery of the protein (*209*). Other studies in which residues in the hydrophobic core were replaced with hydrophilic residues have shown that hydrophobic interactions are critical in the intermediates formed during the folding of IFABP and ILBP (*210*).

The engineering of cysteine residues into proteins for the subsequent labeling of the thiol moiety with fluorescent dyes has become widely used and applied in many protein folding studies. For example, the β-clam protein IFABP contains no cysteine residues in the wild-type sequence. Two cysteine mutants were engineered and then labeled with fluorescein for fluorescence correlation spectroscopy (FCS) experiments which monitored unfolding at low pH (*212*).

Engineering disulfide bonds in or out of proteins: two different strategies have been employed whereby the manipulation of disulfide bonds in proteins has been used to either (i) investigate the oxidative refolding of proteins which naturally contain disulfide bonds, and (ii) investigate the folding pathways of proteins which do not normally contain disulfide bridges. Protein engineering has not been as extensively used in the characterization of the oxidative refolding of disulfide-bonded proteins compared to non-disulfide linked

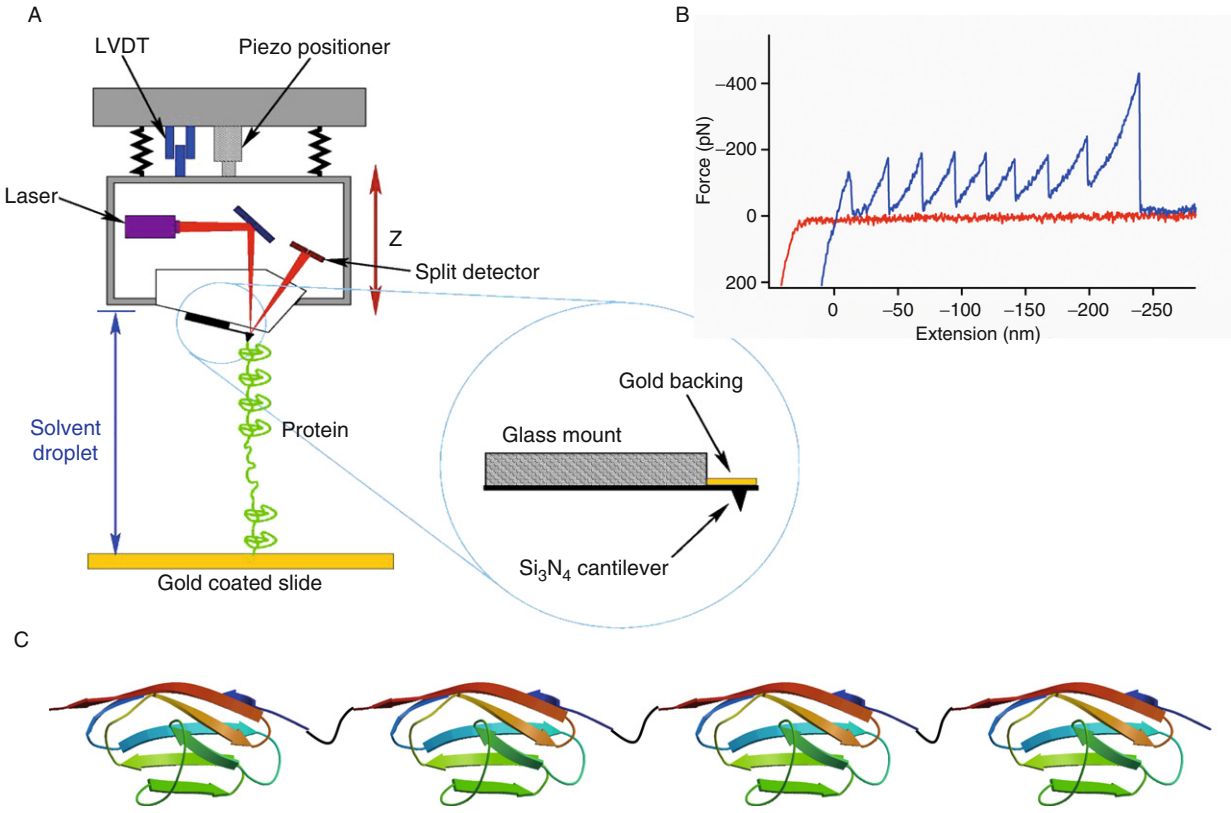
proteins, however, some important studies have used this approach (*213–216*). In other cases, disulfide bonds have been engineered into proteins and two of the best characterized systems are barnase and CD2.d1. CD2.d1 has a β-sandwich structure and populates an intermediate state during folding. Single disulfide cross-links have been engineered at different sites on the surface of the protein to link different β-strands (*217*). Results showed that the linking of β-strands that were distant in sequence stabilized the intermediate state, whereas cross-linking local β-hairpins mainly affected the rate-limiting late transition state (*217*). A follow-up study in which disulfide cross-links were introduced into many of the β-hairpin structures, demonstrated that the effect of the cross-link was very dependent upon the position of the β-hairpin (*218*). Interestingly, a double disulfide bridged mutant folded through a very different pathway from the wild-type and single cross-linked mutants illustrating that large perturbations of proteins can result in deformation of the folding energy landscape (*218*). The results also proved that the intermediate and transition state could be stabilized by nonnative interactions.

X. Single-Molecule Studies of Protein Folding Pathways

Two distinct types of single-molecule techniques have been developed to study protein unfolding and/or folding at the single-molecule level: first mechanical techniques such as atomic force microscopy (AFM) or optical tweezers, where the force is used to unfold single proteins and second, unfolding of individual molecules using chemical denaturants and fluorescence spectroscopy, frequently fluorescence resonance energy transfer (FRET). In both cases, the proteins require engineering for the single-molecule experiments.

For the AFM experiments, a multidomain construct of the protein or proteins under investigation has to be created using molecular biology techniques, as shown in Fig. 7. A versatile cloning system for the production of the multidomain constructs has been described (*178*), and these and alternative methods (*219*) have been used to study the mechanical unfolding of a wide range of proteins. A detailed discussion of all of these studies is beyond the scope of this review, interested readers are directed towards more specific reviews in this area (*220–224*).

A protein engineering approach has been used by the Clarke group to investigate the mechanical stability and unfolding of proteins. The I27 domain from titin is known to show high resistance to force, and a Φ-value analysis was used to investigate the mechanical unfolding transition state (*225*). This showed extensive structure, and that mechanical strength was associated with the interactions formed between β-strands A′ and G. Significant differences between the unfolding transition state with no force applied (by extrapolation


A
LVDT
Piezo positioner
Laser
Z
Split detector
Solvent droplet
Protein
Gold coated slide
Gold backing
Glass mount
Si_3N_4 cantilever
B
Force (pN)
−400
−200
0
200
0
−50
−100
−150
−200
−250
Extension (nm)
C

of experiments in chemical denaturants) and that under force were observed. Despite this, the region of the protein responsible for kinetic stability was found to be the same under both conditions (225). In a similar study on the mechanical unfolding of TNfn3 from human tenascin, the hydrogen bonds and hydrophobic interactions between β-strands A′ and G were also found to be critical for mechanical stability (226). However, there were additional effects particularly rearrangements throughout the core of TNfn3 which also contributed.

Fluorescence-based single-molecule studies using either FRET techniques or two-color coincidence spectroscopy (TCCD) have also been applied to the study of protein folding reactions. In these experiments, a protein is typically labeled with two different dyes—an acceptor and donor pair which can undergo FRET. FRET is used as a measure of the distance between the two fluorophores and allows detection of native, denatured, and potentially intermediate states. For most proteins, the dyes are covalently attached to the protein using thiol chemistry and engineered cysteine residues (227), however, other methods have also been developed (228, 229). Protein engineering techniques are used to incorporate cysteines at specific sites in the protein thereby creating probes of different elements of structure. So far, this powerful technique has mainly be applied to studying the folding/unfolding of proteins under equilibrium conditions, on proteins such as CI2 (230), Csp*Tm* (231), RNase H (232–234), Im9(235), adenylate kinase (236), and protein A (237).

Despite the rapidly increasing number of single molecule protein unfolding studies, there have been relatively few kinetic studies of protein unfolding/folding under nonequilibrium conditions. Recently, a number of such studies have been published: the Eaton group have developed a microfluidic mixing device with which they have studied the folding of Csp*Tm* using dual-labeled protein and FRET techniques (238, 239). Other groups have also designed novel techniques for studying folding processes at the single molecule level including using coaxial mixing devices and capillary flow cells (240, 241). An alternative approach was taken by Chirico, Baldini and co-workers who

FIG. 7. The use of AFM to study protein unfolding. (A) Schematic diagram of the AFM apparatus showing how the multidomain protein construct is attached at one end to a gold surface using engineered cysteine residues and thiol chemistry, and at the other end to a microfabricated silicon nitride cantilever by nonspecific adsorption, (B) Typical force extension curves of a polyprotein containing multiple copies of a single protein domain, (C) Typical multidomain protein construct used in the AFM experiments, the polyprotein is synthesized as a single chain in bacteria with multiple copies of a gene encoding a single domain repeated head-to-tail. The figures show the I27 domain from the giant muscle protein titin. Figure adapted from reference (221).

found evidence for several folded substates and unfolding pathways for green fluorescent protein (GFP) by studying the denaturant induced unfolding of the protein in wet nano-porous gels (*242, 243*). Single-molecule FRET and TCCD techniques have also been used in conjunction with chemical denaturation and a nano-pipette developed by the Klenerman group to study the unfolding of GFP (*244*). In this case, and in contrast to the single-molecule studies on other proteins, an intermediate state on the unfolding pathway was directly observed and the protein found to unfold along two parallel pathways (*244*).

Single-molecule fluorescence techniques and their application to studying protein folding pathways are reviewed in a number of recent publications (*239, 245–247*).

XI. Summary

Over the past 25 years, protein engineering techniques have been used extensively in the study of folding pathways and enabled the determination of folding mechanisms at high resolution. The experimental data obtained from these studies has also had a significant impact in computational studies of folding, either acting as critical benchmarks with which to test simulations (*248–251*), or more recently as restraints in simulations which generate ever more detailed pictures of the energy landscape for folding (*252*). Protein engineering techniques have advanced over the last two decades and new developments such as the use of larger and more diverse protein libraries and selection methods, and the incorporation of novel amino acids into proteins using engineered and expanded genetic codes or the combination of semisynthetic methods and protein engineering techniques, increases the experimental possibilities for studying folding pathways. Together with many advances in instrumentation, particularly for the study of fast folding reactions, means that we now have a formidable array of techniques available to us. One of the next big challenges for both the experimental and computational folding communities is to gain the same level of understanding of how large proteins with complex structures fold as we currently have on small "model" systems. Protein engineering-techniques will undoubtedly continue to play an essential role in these studies.

Acknowledgments

We would like to thank many people in the protein folding community for providing reprints and PDFs and for many fruitful discussions. ALM is funded through a Research Fellowship at St John's College, Cambridge.

References

1. Matthews, C. R., Crisanti, M. M., Manz, J. T., and Gepner, G. L. (1983). Effect of a single amino-acid substitution on the folding of the alpha-subunit of tryptophan synthase. *Biochemistry* **22,** 1445–1452.
2. Beasty, A. M., Hurle, M. R., Manz, J. T., Stackhouse, T., Onuffer, J. J., and Matthews, C. R. (1986). Effects of the phenylalanine-22-]leucine, glutamic acid-49-]methionine, glycine-234-] aspartic acid, and glycine-234-]lysine mutations on the folding and stability of the alpha-subunit of tryptophan synthase from *Escherichia coli*. *Biochemistry* **25,** 2965–2974.
3. Hurle, M. R., Tweedy, N. B., and Matthews, C. R. (1986). Synergism in folding of a double mutant of the alpha-subunit of tryptophan synthase. *Biochemistry* **25,** 6356–6360.
4. Tweedy, N. B., Hurle, M. R., Chrunyk, B. A., and Matthews, C. R. (1990). Multiple replacements at position-211 in the alpha-subunit of tryptophan synthase as a probe of the folding unit association reaction. *Biochemistry* **29,** 1539–1545.
5. Beasty, A. M., Hurle, M. R., Manz, J. T., Stackhouse, T., and Matthews, C. R. (1987). Mutagenesis as a probe of protein folding and stability. *In* "Protein Engineering." Alan R. Liss, New York, pp. 91–102.
6. Matthews, C. R. (1987). Effect of point mutations on the folding of globular-proteins. *Methods Enzymol.* **154,** 498–511.
7. Matthews, C. R., and Hurle, M. R. (1987). Mutant sequences as probes of protein folding mechanisms. *Bioessays* **6,** 254–257.
8. Perry, K. M., Onuffer, J. J., Touchette, N. A., Herndon, C. S., Gittelman, M. S., Matthews, C. R., Chen, J. T., Mayer, R. J., Taira, K., Benkovic, S. J., Howell, E. E., and Kraut, J. (1987). Effect of single amino-acid replacements on the folding and stability of dihydrofolate-reductase from *Escherichia coli*. *Biochemistry* **26,** 2674–2682.
9. Garvey, E. P., and Matthews, C. R. (1989). Effects of multiple replacements at a single position on the folding and stability of dihydrofolate-reductase from *Escherichia coli*. *Biochemistry* **28,** 2083–2093.
10. Perry, K. M., Onuffer, J. J., Gittelman, M. S., Barmat, L., and Matthews, C. R. (1989). Long-range electrostatic interactions can influence the folding, stability, and cooperativity of dihydrofolate-reductase. *Biochemistry* **28,** 7961–7968.
11. Crisanti, M. M., and Matthews, C. R. (1981). Characterization of the slow steps in the folding of the alpha-subunit of tryptophan synthase. *Biochemistry* **20,** 2700–2706.
12. Matthews, C. R., and Crisanti, M. M. (1981). Urea-induced unfolding of the alpha-subunit of tryptophan synthase—Evidence for a multistate process. *Biochemistry* **20,** 784–792.
13. Touchette, N. A., Perry, K. M., and Matthews, C. R. (1986). Folding of dihydrofolate-reductase from *Escherichia coli*. *Biochemistry* **25,** 5445–5452.
14. Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M., and Fersht, A. R. (1990). Transient folding intermediates characterized by protein engineering. *Nature* **346,** 440–445.
15. Matouschek, A., Kellis, J. T., Serrano, L., and Fersht, A. R. (1989). Mapping the transition-state and pathway of protein folding by protein engineering. *Nature* **340,** 122–126.
16. Serrano, L., Matouschek, A., and Fersht, A. R. (1992). The folding of an enzyme. III. Structure of the transition-state for unfolding of barnase analyzed by a protein engineering procedure. *J. Mol. Biol.* **224,** 805–818.
17. Matouschek, A., and Fersht, A. R. (1991). Protein engineering in analysis of protein folding pathways and stability. *Methods Enzymol.* **202,** 82–112.
18. Fersht, A. R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **224,** 771–782.

19. Fersht, A. R., Kellis, J. T., Matouschek, A., and Serrano, L. (1990). Folding pathway enigma—Reply. *Nature* **343,** 602.
20. Jackson, S. E., and Fersht, A. R. (1991). Folding of chymotrypsin inhibitor-2.1. Evidence for a 2-state transition. *Biochemistry* **30,** 10428–10435.
21. Jackson, S. E., Elmasry, N., and Fersht, A. R. (1993). Structure of the hydrophobic core in the transition-state for folding of chymotrypsin inhibitor-2—A critical test of the protein engineering method of analysis. *Biochemistry* **32,** 11270–11278.
22. Otzen, D. E., Itzhaki, L. S., Elmasry, N. F., Jackson, S. E., and Fersht, A. R. (1994). Structure of the transition-state for the folding/unfolding of the barley chymotrypsin inhibitor-2 and its implications for mechanisms of protein-folding. *Proc. Natl. Acad. Sci. USA* **91,** 10422–10425.
23. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995). The structure of the transition-state for folding of chymotrypsin inhibitor-2 analyzed by protein engineering methods—Evidence for a nucleation–condensation mechanism for protein-folding. *J. Mol. Biol.* **254,** 260–288.
24. Khorasanizadeh, S., Peters, I. D., and Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat. Struct. Biol.* **3,** 193–205.
25. Jackson, S. E. (2006). Ubiquitin: A small protein folding paradigm. *Org. Biomol. Chem.* **4,** 1845–1853.
26. Went, H. M., and Jackson, S. E. (2005). Ubiquitin folds through a highly polarized transition state. *Protein Eng. Des. Sel.* **18,** 229–237.
27. Campbell-Valois, F. X., and Michnick, S. W. (2007). The transition state of the ras binding domain of raf is structurally polarized based on phi-values but is energetically diffuse. *J. Mol. Biol.* **365,** 1559–1577.
28. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995). Scop—A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247,** 536–540.
29. Gu, H. D., Kim, D., and Baker, D. (1997). Contrasting roles for symmetrically disposed beta-turns in the folding of a small protein. *J. Mol. Biol.* **274,** 588–596.
30. Kim, D. E., Yi, Q., Gladwin, S. T., Goldberg, J. M., and Baker, D. (1998). The single helix in protein L is largely disrupted at the rate-limiting step in folding. *J. Mol. Biol.* **284,** 807–815.
31. Gu, H., Doshi, N., Kim, D. E., Simons, K. T., Santiago, J. V., Nauli, S., and Baker, D. (1999). Robustness of protein folding kinetics to surface hydrophobic substitutions. *Protein Sci.* **8,** 2734–2741.
32. Kim, D. E., Fisher, C., and Baker, D. (2000). A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.* **298,** 971–984.
33. McCallister, E. L., Alm, E., and Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nat. Struct. Biol.* **7,** 669–673.
34. Nauli, S., Kuhlman, B., and Baker, D. (2001). Computer-based redesign of a protein folding pathway. *Nat. Struct. Biol.* **8,** 602–605.
35. Schymkowitz, J. W. H., Rousseau, F., Irvine, L. R., and Itzhaki, L. S. (2000). The folding pathway of the cell-cycle regulatory protein p13suc1: Clues for the mechanism of domain swapping. *Structure* **8,** 89–100.
36. Rousseau, F., Schymkowitz, J. W. H., Wilkinson, H. R., and Itzhaki, L. S. (2002). The structure of the transition state for folding of domain-swapped dimeric p13suc1. *Structure* **10,** 649–657.
37. Seeliger, M. A., Breward, S. E., and Itzhaki, L. S. (2003). Weak cooperativity in the core causes a switch in folding mechanism between two proteins of the cks family. *J. Mol. Biol.* **325,** 189–199.
38. Viguera, A. R., Villegas, V., Aviles, F. X., and Serrano, L. (1997). Favourable native-like helical local interactions can accelerate protein folding. *Fold. Des.* **2,** 23–33.

39. Villegas, V., Martinez, J. C., Aviles, F. X., and Serrano, L. (1998). Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.* **283,** 1027–1036.
40. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M., and Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nat. Struct. Biol.* **6,** 1005–1009.
41. Ternstrom, T., Mayor, U., Akke, M., and Oliveberg, M. (1999). From snapshot to movie: Phi analysis of protein folding transition states taken one step further. *Proc. Natl. Acad. Sci. USA* **96,** 14854–14859.
42. Otzen, D. E., and Oliveberg, M. (2002). Conformational plasticity in folding of the split beta-alpha-beta protein S6: Evidence for burst-phase disruption of the native state. *J. Mol. Biol.* **317,** 613–627.
43. Hedberg, L., and Oliveberg, M. (2004). Scattered Hammond plots reveal second level of site-specific information in protein folding: $\varphi'(\beta^{\ddagger})$. *Proc. Natl. Acad. Sci. USA* **101,** 7606–7611.
44. Raschke, T. M., and Marqusee, S. (1997). The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat. Struct. Biol.* **4,** 298–304.
45. Raschke, T. M., Kho, J., and Marqusee, S. (1999). Confirmation of the hierarchical folding of RNase H: A protein engineering study. *Nat. Struct. Biol.* **6,** 825–831.
46. Spudich, G., and Marqusee, S. (2000). A change in the apparent m value reveals a populated intermediate under equilibrium conditions in *Escherichia coli* ribonuclease HI. *Biochemistry* **39,** 11677–11683.
47. Spudich, G. M., Miller, E. J., and Marqusee, S. (2004). Destabilization of the *Escherichia coli* RNase H kinetic intermediate: Switching between a two-state and three-state folding mechanism. *J. Mol. Biol.* **335,** 609–618.
48. Mayor, U., Johnson, C. M., Daggett, V., and Fersht, A. R. (2000). Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. USA* **97,** 13518–13522.
49. Mayor, U., Grossmann, J. G., Foster, N. W., Freund, S. M. V., and Fersht, A. R. (2003). The denatured state of engrailed homeodomain under denaturing and native conditions. *J. Mol. Biol.* **333,** 977–991.
50. Religa, T. L., Johnson, C. M., Vu, D. M., Brewer, S. H., Dyer, R. B., and Fersht, A. R. (2007). The helix-turn-helix motif as an ultrafast independently folding domain: The pathway of folding of Engrailed homeodomain. *Proc. Natl. Acad. Sci. USA* **104,** 9272–9277.
51. Mayor, U. *et al*. (2003). The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421,** 863–867.
52. Huang, G. S., and Oas, T. G. (1995). Submillisecond folding of monomeric lambda-repressor. *Proc. Natl. Acad. Sci. USA* **92,** 6878–6882.
53. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997). The energy landscape of a fast-folding protein mapped by Ala→Gly substitutions. *Nat. Struct. Biol.* **4,** 305–310.
54. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W., and Oas, T. G. (1996). Microsecond protein folding through a compact transition state. *J. Mol. Biol.* **263,** 311–322.
55. Burton, R. E., Myers, J. K., and Oas, T. G. (1998). Protein folding dynamics: Quantitative comparison between theory and experiment. *Biochemistry* **37,** 5337–5343.
56. Myers, J. K., and Oas, T. G. (1999). Contribution of a buried hydrogen bond to lambda repressor folding kinetics. *Biochemistry* **38,** 6761–6768.
57. Kapp, G. T., Richardson, J. S., and Oas, T. G. (2004). Kinetic role of helix caps in protein folding is context-dependent. *Biochemistry* **43,** 3814–3823.
58. Arora, P., Oas, T. G., and Myers, J. K. (2004). Fast and faster: A designed variant of the B-domain of protein A folds in 3 mu sec. *Protein Sci.* **13,** 847–853.

59. Kragelund, B. B., Osmark, P., Neergaard, T. B., Schiodt, J., Kristiansen, K., Knudsen, J., and Poulsen, F. M. (1999). The formation of a native-like structure containing eight conserved hydrophobic residues is rate limiting in two-state protein folding of ACBP. *Nat. Struct. Biol.* **6,** 594–601.
60. Teilum, K., Thormann, T., Caterer, N. R., Poulsen, H. I., Jensen, P. H., Knudsen, J., Kragelund, B. B., and Poulsen, F. M. (2005). Different secondary structure elements as scaffolds for protein folding transition states of two homologous four-helix bundles. *Protein Struct. Funct. Bioinform.* **59,** 80–90.
61. Ferguson, N., Li, W., Capaldi, A. P., Kleanthous, C., and Radford, S. E. (2001). Using chimeric immunity proteins to explore the energy landscape for alpha-helical protein folding. *J. Mol. Biol.* **307,** 393–405.
62. Capaldi, A. P., Kleanthous, C., and Radford, S. E. (2002). Im7 folding mechanism: Misfolding on a path to the native state. *Nat. Struct. Biol.* **9,** 209–216.
63. Friel, C. T., Capaldi, A. P., and Radford, S. E. (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.
64. Spence, G. R., Capaldi, A. P., and Radford, S. E. (2004). Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J. Mol. Biol.* **341,** 215–226.
65. Whittaker, S. B. M., Spence, G. R., Grossmann, J. G., Radford, S. E., and Moore, G. R. (2007). NMR analysis of the conformational properties of the trapped on-pathway folding intermediate of the bacterial immunity protein Im7. *J. Mol. Biol.* **366,** 1001–1015.
66. Friel, C. T., Beddard, G. S., and Radford, S. E. (2004). Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J. Mol. Biol.* **342,** 261–273.
67. Cranz-Mileva, S., Friel, C. T., and Radford, S. E. (2005). Helix stability and hydrophobicity in the folding mechanism of the bacterial immunity protein Im9. *Protein Eng. Des. Sel.* **18,** 41–50.
68. Morton, V. L., Friel, C. T., Allen, L. R., Paci, E., and Radford, S. E. (2007). The effect of increasing the stability of non-native interactions on the folding landscape of the bacterial immunity protein Im9. *J. Mol. Biol.* **371,** 554–568.
69. Hackenberger, C. P. R., Friel, C. T., Radford, S. E., and Imperiali, B. (2005). Semisynthesis of a glycosylated lm7 analogue for protein folding studies. *J. Am. Chem. Soc.* **127,** 12882–12889.
70. Scott, K. A., and Clarke, J. (2005). Spectrin R16: Broad energy barrier or sequential transition states? *Protein Sci.* **14,** 1617–1629.
71. Scott, K. A., Randles, L. G., and Clarke, J. (2004). The folding of spectrin domains II: Phi-value analysis of R16. *J. Mol. Biol.* **344,** 207–221.
72. Scott, K. A., Randles, L. G., Moran, S. J., Daggett, V., and Clarke, J. (2006). The folding pathway of spectrin R17 from experiment and simulation: Using experimentally validated MD simulations to characterize states hinted at by experiment. *J. Mol. Biol.* **359,** 159–173.
73. Grantcharova, V. P., Riddle, D. S., Santiago, J. V., and Baker, D. (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat. Struct. Biol.* **5,** 714–720.
74. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999). Experiment and theory highlight role of native state topology in SH3 folding. *Nat. Struct. Biol.* **6,** 1016–1024.
75. Grantcharova, V. P., Riddle, D. S., and Baker, D. (2000). Long-range order in the src SH3 folding transition state. *Proc. Natl. Acad. Sci. USA* **97,** 7084–7089.
76. Martinez, J. C., and Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* **6,** 1010–1016.

77. Viguera, A. R., Vega, C., and Serrano, L. (2002). Unspecific hydrophobic stabilization of folding transition states. *Proc. Natl. Acad. Sci. USA* **99,** 5349–5354.
78. Fernandez-Escamilla, A. M., Cheung, M. S., Vega, M. C., Wilmanns, M., Onuchic, J. N., and Serrano, L. (2004). Solvation in protein folding analysis: Combination of theoretical and experimental approaches. *Proc. Natl. Acad. Sci. USA* **101,** 2834–2839.
79. Guerois, R., and Serrano, L. (2000). The SH3-fold family: Experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* **304,** 967–982.
80. Northey, J. G. B., Di Nardo, A. A., and Davidson, A. R. (2002). Hydrophobic core packing in the SH3 domain folding transition state. *Nat. Struct. Biol.* **9,** 126–130.
81. Northey, J. G. B., Maxwell, K. L., and Davidson, A. R. (2002). Protein folding kinetics beyond the Phi value: Using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state. *J. Mol. Biol.* **320,** 389–402.
82. Di Nardo, A. A., Korzhnev, D. M., Stogios, P. J., Zarrine-Afsar, A., Kay, L. E., and Davidson, A. R. (2004). Dramatic acceleration of protein folding by stabilization of a non-native backbone conformation. *Proc. Natl. Acad. Sci. USA* **101,** 7954–7959.
83. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M., and Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430,** 586–590.
84. Neudecker, P., Zarrine-Afsar, A., Choy, W. Y., Muhandiram, D. R., Davidson, A. R., and Kay, L. E. (2006). Identification of a collapsed intermediate with non-native long-range interactions on the folding pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy. *J. Mol. Biol.* **363,** 958–976.
85. Neudecker, P., Zarrine-Afsar, A., Davidson, A. R., and Kay, L. E. (2007). Phi-Value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy. *Proc. Natl. Acad. Sci. USA* **104,** 15717–15722.
86. Garcia-Mira, M. M., Boehringer, D., and Schmid, F. X. (2004). The folding transition state of the cold shock protein is strongly polarized. *J. Mol. Biol.* **339,** 555–569.
87. Garcia-Mira, M. M., and Schmid, F. X. (2006). Key role of coulombic interactions for the folding transition state of the cold shock protein. *J. Mol. Biol.* **364,** 458–468.
88. Crane, J. C., Koepf, E. K., Kelly, J. W., and Gruebele, M. (2000). Mapping the transition state of the WW domain beta-sheet. *J. Mol. Biol.* **298,** 283–292.
89. Nguyen, H., Jager, M., Kelly, J. W., and Gruebele, M. (2005). Engineering beta-sheet protein toward the folding speed limit. *J. Phys. Chem. B* **109,** 15182–15186.
90. Jager, M., Nguyen, H., Crane, J. C., Kelly, J. W., and Gruebele, M. (2001). The folding mechanism of a beta-sheet: The WW domain. *J. Mol. Biol.* **311,** 373–393.
91. Jager, M. *et al.* (2006). Structure-function-folding relationship in a WW domain. *Proc. Natl. Acad. Sci. USA* **103,** 10648–10653.
92. Schreiber, C., Buckle, A. M., and Fersht, A. R. (1994). Stability and function—Two constraints in the evolution of barstar and other proteins. *Structure* **2,** 945–951.
93. Deechongkit, S., Nguyen, H., Powers, E. T., Dawson, P. E., Gruebele, M., and Kelly, J. W. (2004). Context-dependent contributions of backbone hydrogen bonding to beta-sheet folding energetics. *Nature* **430,** 101–105.
94. Ferguson, N. *et al.* (2001). Using flexible loop mimetics to extend Phi-value analysis to secondary structure interactions. *Proc. Natl. Acad. Sci. USA* **98,** 13008–13013.
95. Petrovich, M., Jonsson, A. L., Ferguson, N., Daggett, V., and Fersht, A. R. (2006). phi-Analysis at the experimental limits: Mechanism of beta-hairpin formation. *J. Mol. Biol.* **360,** 865–881.
96. Hamill, S. J., Steward, A., and Clarke, J. (2000). The folding of an immunoglobulin-like Greek key protein is defined by a common-core nucleus and regions constrained by topology. *J. Mol. Biol.* **297,** 165–178.

97. Cota, E., Steward, A., Fowler, S. B., and Clarke, J. (2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. *J. Mol. Biol.* **305,** 1185–1194.
98. Fowler, S. B., and Clarke, J. (2001). Mapping the folding pathway of an immunoglobulin domain: Structural detail from phi value analysis and movement of the transition state. *Structure* **9,** 355–366.
99. Wright, C. F., Christodoulou, J., Dobson, C. M., and Clarke, J. (2004). The importance of loop length in the folding of an immunoglobulin domain. *Protein Eng. Des. Sel.* **17,** 443–453.
100. Lappalainen, I., Hurley, M. G., and Clarke, J. (2008). Plasticity within the obligatory folding nucleus of an immunoglobulin-like domain. *J. Mol. Biol.* **375,** 547–559.
101. Billings, K. S., Best, R. B., Rutherford, T. J., and Clarke, J. (2008). Crosstalk between the protein surface and hydrophobic core in a core-swapped fibronectin type III domain. *J. Mol. Biol.* **375,** 560–571.
102. Lorch, M., Mason, J. M., Clarke, A. R., and Parker, M. J. (1999). Effects of core mutations on the folding of a beta-sheet protein: Implications for backbone organization in the I-State. *Biochemistry* **38,** 1377–1385.
103. Lorch, M., Mason, J. M., Sessions, R. B., and Clarke, A. R. (2000). Effects of mutations on the thermodynamics of a protein folding reaction: Implications for the mechanism of formation of the intermediate and transition states. *Biochemistry* **39,** 3480–3485.
104. Poso, D., Sessions, R. B., Lorch, M., and Clarke, A. R. (2000). Progressive stabilization of intermediate and transition states in protein folding reactions by introducing surface hydrophobic residues. *J. Biol. Chem.* **275,** 35723–35726.
105. Fersht, A. R. (2004). Relationship of Leffler (Bronsted) alpha values and protein folding Phi values to position of transition-state structures on reaction coordinates. *Proc. Natl. Acad. Sci. USA* **101,** 14338–14342.
106. Fersht, A. R., Itzhaki, L. S., El Masry, N. F., Matthews, J. M., and Otzen, D. E. (1994). Single versus parallel pathways of protein-folding and fractional formation of structure in the transition-state. *Proc. Natl. Acad. Sci. USA* **91,** 10426–10429.
107. Sosnick, T. R., Krantz, B. A., Dothager, R. S., and Baxa, M. (2006). Characterizing the protein folding transition state using psi analysis. *Chem. Rev.* **106,** 1862–1876.
108. Krantz, B. A., and Sosnick, T. R. (2001). Engineered metal binding sites map the heterogeneous folding landscape of a coiled coil. *Nat. Struct. Biol.* **8,** 1042–1047.
109. Krantz, B. A., Dothager, R. S., and Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J. Mol. Biol.* **337,** 463–475.
110. Krantz, B. A., Dothager, R. S., and Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J. Mol. Biol.* **347,** 463–475.
111. Moran, L. B., Schneider, J. P., Kentsis, A., Reddy, G. A., and Sosnick, T. R. (1999). Transition state heterogeneity in GCN4 coiled coil folding studied by using multisite mutations and crosslinking. *Proc. Natl. Acad. Sci. USA* **96,** 10699–10704.
112. Fersht, A. R. (2004). Phi value versus psi analysis. *Proc. Natl. Acad. Sci. USA* **101,** 17327–17328.
113. Lin, L. N., and Brandts, J. F. (1978). Further evidence suggesting that slow phase in protein unfolding and refolding is due to proline isomerization—Kinetic study of carp parvalbumins. *Biochemistry* **17,** 4102–4110.
114. Brandts, J. F., Brennan, M., and Lin, L. N. (1977). Unfolding and refolding occur much faster for a proline-free protein than for most proline-containing proteins. *Proc. Natl. Acad. Sci. USA* **74,** 4178–4181.

115. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975). Consideration of possibility that slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. *Biochemistry* **14,** 4953–4963.

116. Engel, J., and Bachinger, H. P. (2005). Structure, stability and folding of the collagen triple helix. *Collagen* **247,** 7–33.

117. Kiefhaber, T., Grunert, H. P., Hahn, U., and Schmid, F. X. (1990). Replacement of a cis proline simplifies the mechanism of ribonuclease-T1 folding. *Biochemistry* **29,** 6475–6480.

118. Schreiber, G., and Fersht, A. R. (1993). The refolding of cis-peptidylprolyl and trans-peptidylprolyl isomers of barstar. *Biochemistry* **32,** 11195–11203.

119. Schultz, D. A., Schmid, F. X., and Baldwin, R. L. (1992). Cis proline mutants of ribonuclease-a.2. Elimination of the slow-folding forms by mutation. *Protein Sci.* **1,** 917–924.

120. Munoz, V., Lopez, E. M., Jager, M., and Serrano, L. (1994). Kinetic characterization of the chemotactic protein from *Escherichia coli*, chey—Kinetic-analysis of the inverse hydrophobic effect. *Biochemistry* **33,** 5858–5866.

121. Eyles, S. J., and Gierasch, L. M. (2000). Multiple roles of prolyl residues in structure and folding. *J. Mol. Biol.* **301,** 737–747.

122. Burns-Hamuro, L. L., Dalessio, P. M., and Ropson, I. J. (2004). Replacement of proline with valine does not remove an apparent proline isomerization-dependent folding event in CRABP I. *Protein Sci.* **13,** 1670–1676.

123. Kamen, D. E., and Woody, R. W. (2002). Identification of proline residues responsible for the slow folding kinetics in pectate lyase C by mutagenesis. *Biochemistry* **41,** 4724–4732.

124. Maki, K., Ikura, T., Hayano, T., Takahashi, N., and Kuwajima, K. (1999). Effects of proline mutations on the folding of staphylococcal nuclease. *Biochemistry* **38,** 2213–2223.

125. Kuwajima, K., Okayama, N., Yamamoto, K., Ishihara, T., and Sugai, S. (1991). The Pro117 to glycine mutation of staphylococcal nuclease simplifies the unfolding folding kinetics. *FEBS Lett.* **290,** 135–138.

126. Rousseau, F., Schymkowitz, J. W. H., del Pino, M. S., and Itzhaki, L. S. (1998). Stability and folding of the cell cycle regulatory protein, p13(suc1). *J. Mol. Biol.* **284,** 503–519.

127. Main, E. R., Jackson, S. E., and Regan, L. (2003). The folding and design of repeat proteins: Reaching a consensus. *Curr. Opin. Struct. Biol.* **13,** 482–489.

128. Main, E. R., Lowe, A. R., Mochrie, S. G., Jackson, S. E., and Regan, L. (2005). A recurring theme in protein engineering: The design, stability and folding of repeat proteins. *Curr. Opin. Struct. Biol.* **15,** 464–471.

129. Barrick, D., Ferreiro, D. U., and Komives, E. A. (2008). Folding landscapes of ankyrin repeat proteins: Experiments meet theory. *Curr. Opin. Struct. Biol.* **18,** 27–34.

130. Low, C., Weininger, U., Zeeb, M., Zhang, W., Laue, E. D., Schmid, F. X., and Balbach, J. (2007). Folding mechanism of an ankyrin repeat protein: Scaffold and active site formation of human CDK inhibitor p19(INK4d). *J. Mol. Biol.* **373,** 219–231.

131. Tang, K. S., Fersht, A. R., and Itzhaki, L. S. (2003). Sequential unfolding of ankyrin repeats in tumor suppressor p16. *Structure* **11,** 67–73.

132. Mosavi, L. K., Cammett, T. J., Desrosiers, D. C., and Peng, Z. Y. (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13,** 1435–1448.

133. Interlandi, G., Settanni, G., and Caflisch, A. (2006). Unfolding transition state and intermediates of the tumor suppressor p16INK4a investigated by molecular dynamics simulations. *Proteins* **64,** 178–192.

134. Lowe, A. R., and Itzhaki, L. S. (2007). Rational redesign of the folding pathway of a modular protein. *Proc. Natl. Acad. Sci. USA* **104,** 2679–2684.

135. Truhlar, M. E., and Komives, E. A. (2008). LRR domain folding: Just put a cap on it!. *Structure* **16,** 655–657.

136. Courtemanche, N., and Barrick, D. (2008). The leucine-rich repeat domain of internalin B folds along a polarized N-terminal pathway. *Structure* **16,** 705–714.

137. Mello, C. C., and Barrick, D. (2004). An experimentally determined protein folding energy landscape. *Proc. Natl. Acad. Sci. USA* **101,** 14102–14107.

138. Zhang, B., and Peng, Z. (2000). A minimum folding unit in the ankyrin repeat protein p16(INK4). *J. Mol. Biol.* **299,** 1121–1132.

139. Zweifel, M. E., and Barrick, D. (2001). Studies of the ankyrin repeats of the *Drosophila melanogaster* Notch receptor. II. Solution stability and cooperativity of unfolding. *Biochemistry* **40,** 14357–14367.

140. Tripp, K. W., and Barrick, D. (2004). The tolerance of a modular protein to duplication and deletion of internal repeats. *J. Mol. Biol.* **344,** 169–178.

141. Street, T. O., Bradley, C. M., and Barrick, D. (2007). Predicting coupling limits from an experimentally determined energy landscape. *Proc. Natl. Acad. Sci. USA* **104,** 4907–4912.

142. Ferreiro, D. U., Cho, S. S., Komives, E. A., and Wolynes, P. G. (2005). The energy landscape of modular repeat proteins: Topology determines folding mechanism in the ankyrin family. *J. Mol. Biol.* **354,** 679–692.

143. Werbeck, N. D., and Itzhaki, L. S. (2007). Probing a moving target with a plastic unfolding intermediate of an ankyrin-repeat protein. *Proc. Natl. Acad. Sci. USA* **104,** 7863–7868.

144. Kajander, T., Cortajarena, A. L., and Regan, L. (2006). Consensus design as a tool for engineering repeat proteins. *Methods Mol. Biol.* **340,** 151–170.

145. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P., and Pluckthun, A. (2003). Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332,** 489–503.

146. Kohl, A., Binz, H. K., Forrer, P., Stumpp, M. T., Pluckthun, A., and Grutter, M. G. (2003). Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci. USA* **100,** 1700–1705.

147. Main, E. R., Xiong, Y., Cocco, M. J., D'Andrea, L., and Regan, L. (2003). Design of stable alpha-helical arrays from an idealized TPR motif. *Structure* **11,** 497–508.

148. Mosavi, L. K., Minor, D. L., Jr., and Peng, Z. Y. (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci. USA* **99,** 16029–16034.

149. Stumpp, M. T., Forrer, P., Binz, H. K., and Pluckthun, A. (2003). Designing repeat proteins: Modular leucine-rich repeat protein libraries based on the mammalian ribonuclease inhibitor family. *J. Mol. Biol.* **332,** 471–487.

150. Wetzel, S. K., Settanni, G., Kenig, M., Binz, H. K., and Pluckthun, A. (2008). Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins. *J. Mol. Biol.* **376,** 241–257.

151. Parmeggiani, F., Pellarin, R., Larsen, A. P., Varadamsetty, G., Stumpp, M. T., Zerbe, O., Caflisch, A., and Pluckthun, A. (2008). Designed armadillo repeat proteins as general peptide-binding scaffolds: Consensus design and computational optimization of the hydrophobic core. *J. Mol. Biol.* **376,** 1282–1304.

152. Tripp, K. W., and Barrick, D. (2003). Folding by consensus. *Structure* **11,** 486–487.

153. Tripp, K. W., and Barrick, D. (2007). Enhancing the stability and folding rate of a repeat protein through the addition of consensus repeats. *J. Mol. Biol.* **365,** 1187–1200.

154. Ferreiro, D. U., Cervantes, C. F., Truhlar, S. M., Cho, S. S., Wolynes, P. G., and Komives, E. A. (2007). Stabilizing IkappaBalpha by "consensus" design. *J. Mol. Biol.* **365,** 1201–1216.

155. Main, E. R., Stott, K., Jackson, S. E., and Regan, L. (2005). Local and long-range stability in tandemly arrayed tetratricopeptide repeats. *Proc. Natl. Acad. Sci. USA* **102,** 5721–5726.

156. Goldenberg, D. P., and Creighton, T. E. (1983). Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **165,** 407–413.

157. Luger, K., Hommel, U., Herold, M., Hofsteenge, J., and Kirschner, K. (1989). Correct folding of circularly permuted variants of a beta alpha barrel enzyme *in vivo*. *Science* **243,** 206–210.

158. Viguera, A. R., Blanco, F. J., and Serrano, L. (1995). The order of secondary structure elements does not determine the structure of a protein but does affect its folding kinetics. *J. Mol. Biol.* **247,** 670–681.
159. Viguera, A. R., Serrano, L., and Wilmanns, M. (1996). Different folding transition states may result in the same native structure. *Nat. Struct. Biol.* **3,** 874–880.
160. Martinez, J. C., Viguera, A. R., Berisio, R., Wilmanns, M., Mateo, P. L., Filimonov, V. V., and Serrano, L. (1999). Thermodynamic analysis of alpha-spectrin SH3 and two of its circular permutants with different loop lengths: Discerning the reasons for rapid folding in proteins. *Biochemistry* **38,** 549–559.
161. Otzen, D. E., and Fersht, A. R. (1998). Folding of circular and permuted chymotrypsin inhibitor 2: Retention of the folding nucleus. *Biochemistry* **37,** 8139–8146.
162. Iwakura, M., Nakamura, T., Yamane, C., and Maki, K. (2000). Systematic circular permutation of an entire protein reveals essential folding elements. *Nat. Struct. Biol.* **7,** 580–585.
163. Arai, M., Maki, K., Takahashi, H., and Iwakura, M. (2003). Testing the relationship between foldability and the early folding events of dihydrofolate reductase from *Escherichia coli*. *J. Mol. Biol.* **328,** 273–288.
164. Hennecke, J., Sebbel, P., and Glockshuber, R. (1999). Random circular permutation of DsbA reveals segments that are essential for protein folding and stability. *J. Mol. Biol.* **286,** 1197–1215.
165. 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.
166. Lindberg, M. O., Tangrot, J., Otzen, D. E., Dolgikh, D. A., Finkelstein, A. V., and Oliveberg, M. (2001). Folding of circular permutants with decreased contact order: General trend balanced by protein stability. *J. Mol. Biol.* **314,** 891–900.
167. Miller, E. J., Fischer, K. F., and Marqusee, S. (2002). Experimental evaluation of topological parameters determining protein-folding rates. *Proc. Natl. Acad. Sci. USA* **99,** 10359–10363.
168. Lindberg, M. O., Haglund, E., Hubner, I. A., Shakhnovich, E. I., and Oliveberg, M. (2006). Identification of the minimal protein-folding nucleus through loop-entropy perturbations. *Proc. Natl. Acad. Sci. USA* **103,** 4083–4088.
169. Olofsson, M., Hansson, S., Hedberg, L., Logan, D. T., and Oliveberg, M. (2007). Folding of S6 structures with divergent amino acid composition: Pathway flexibility within partly overlapping foldons. *J. Mol. Biol.* **365,** 237–248.
170. Iwakura, M., and Nakamura, T. (1998). Effects of the length of a glycine linker connecting the N- and C-termini of a circularly permuted dihydrofolate reductase. *Protein Eng.* **11,** 707–713.
171. 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.
172. Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A., and Clarke, J. (2007). The folding and evolution of multidomain proteins. *Nat. Rev. Mol. Cell. Biol.* **8,** 319–330.
173. Ekman, D., Bjorklund, A. K., Frey-Skott, J., and Elofsson, A. (2005). Multi-domain proteins in the three kingdoms of life: Orphan domains and other unassigned regions. *J. Mol. Biol.* **348,** 231–243.
174. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3,** R81–R91.
175. Daggett, V., and Fersht, A. (2003). The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell. Biol.* **4,** 497–502.
176. Daggett, V., and Fersht, A. R. (2003). Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **28,** 18–25.
177. Zarrine-Afsar, A., Larson, S. M., and Davidson, A. R. (2005). The family feud: Do proteins with similar structures fold via the same pathway? *Curr. Opin. Struct. Biol.* **15,** 42–49.
178. Steward, A., Toca-Herrera, J. L., and Clarke, J. (2002). Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci.* **11,** 2179–2183.

179. Batey, S., and Clarke, J. (2006). Apparent cooperativity in the folding of multidomain proteins depends on the relative rates of folding of the constituent domains. *Proc. Natl. Acad. Sci. USA* **103,** 18113–18118.

180. Arora, P., Hammes, G. G., and Oas, T. G. (2006). Folding mechanism of a multiple independently-folding domain protein: Double B domain of protein A. *Biochemistry* **45,** 12312–12324.

181. Hamill, S. J., Meekhof, A. E., and Clarke, J. (1998). The effect of boundary selection on the stability and folding of the third fibronectin type III domain from human tenascin. *Biochemistry* **37,** 8071–8079.

182. Batey, S., Randles, L. G., Steward, A., and Clarke, J. (2005). Cooperative folding in a multidomain protein. *J. Mol. Biol.* **349,** 1045–1059.

183. Batey, S., Scott, K. A., and Clarke, J. (2006). Complex folding kinetics of a multidomain protein. *Biophys. J.* **90,** 2120–2130.

184. Osvath, S., Kohler, G., Zavodszky, P., and Fidy, J. (2005). Asymmetric effect of domain interactions on the kinetics of folding in yeast phosphoglycerate kinase. *Protein Sci.* **14,** 1609–1616.

185. Wenk, M., Jaenicke, R., and Mayr, E. M. (1998). Kinetic stabilisation of a modular protein by domain interactions. *FEBS Lett.* **438,** 127–130.

186. Jager, M., Gehrig, P., and Pluckthun, A. (2001). The scFv fragment of the antibody hu4D5–8: Evidence for early premature domain interaction in refolding. *J. Mol. Biol.* **305,** 1111–1129.

187. Batey, S., and Clarke, J. (2008). The folding pathway of a single domain in a multidomain protein is not affected by its neighbouring domain. *J. Mol. Biol.* **378**(2), 297–301.

188. Scott, K. A., Steward, A., Fowler, S. B., and Clarke, J. (2002). Titin; a multidomain protein that behaves as the sum of its parts. *J. Mol. Biol.* **315,** 819–829.

189. Steward, A., Adhya, S., and Clarke, J. (2002). Sequence conservation in Ig-like domains: The role of highly conserved proline residues in the fibronectin type III superfamily. *J. Mol. Biol.* **318,** 935–940.

190. Robertsson, J., Petzold, K., Lofvenberg, L., and Backman, L. (2005). Folding of spectrin's SH3 domain in the presence of spectrin repeats. *Cell. Mol. Biol. Lett.* **10,** 595–612.

191. Jaenicke, R., and Lilie, H. (2000). Folding and association of oligomeric and multimeric proteins. *Adv. Protein Chem.* **53,** 329–401.

192. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995). P22 Arc repressor: Transition state properties inferred from mutational effects on the rates of protein unfolding and refolding. *Biochemistry* **34,** 13914–13919.

193. Milla, M. E., and Sauer, R. T. (1994). P22 Arc repressor: Folding kinetics of a single-domain, dimeric protein. *Biochemistry* **33,** 1125–1133.

194. Doyle, S. M., Bilsel, O., and Teschke, C. M. (2004). SecA folding kinetics: A large dimeric protein rapidly forms multiple native states. *J. Mol. Biol.* **341,** 199–214.

195. Mallam, A. L., and Jackson, S. E. (2006). Probing Nature's knots: The folding pathway of a knotted homodimeric protein. *J. Mol. Biol.* **359,** 1420–1436.

196. Mallam, A. L., and Jackson, S. E. (2007). The dimerization of an alpha/beta-knotted protein is essential for structure and function. *Structure* **15,** 111–122.

197. Shao, X., Hensley, P., and Matthews, C. R. (1997). Construction and characterization of monomeric tryptophan repressor: A model for an early intermediate in the folding of a dimeric protein. *Biochemistry* **36,** 9941–9949.

198. Mallam, A. L., Onuoha, S. C., Grossmann, J. G., and Jackson, S. E. (2008). Knotted fusion proteins reveal unexpected possibilities in protein folding. *Mol. Cell* **30**(5), 642–648.

199. Randles, L. G., Batey, S., Steward, A., and Clarke, J. (2008). Distinguishing specific and nonspecific interdomain interactions in multidomain proteins. *Biophys. J.* **94,** 622–628.

200. Khorasanizadeh, S., Peters, I. D., Butt, T. R., and Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **32,** 7054–7063.

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

202. Krantz, B. A., and Sosnick, T. R. (2000). Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* **39,** 11696–11701.

203. Ghaemmaghami, S., Word, J. M., Burton, R. E., Richardson, J. S., and Oas, T. G. (1998). Folding kinetics of a fluorescent variant of monomeric lambda repressor. *Biochemistry* **37,** 9179–9185.

204. Dimitriadis, G., Drysdale, A., Myers, J. K., Arora, P., Radford, S. E., Oas, T. G., and Smith, D. A. (2004). Microsecond folding dynamics of the F13W G29A mutant of the B domain of staphylococcal protein A by laser-induced temperature jump. *Proc. Natl. Acad. Sci. USA* **101,** 3809–3814.

205. Smith, C. J., Clarke, A. R., Chia, W. N., Irons, L. I., Atkinson, T., and Holbrook, J. J. (1991). Detection and characterization of intermediates in the folding of large proteins by the use of genetically inserted tryptophan probes. *Biochemistry* **30,** 1028–1036.

206. Staniforth, R. A., Burston, S. G., Smith, C. J., Jackson, G. S., Badcoe, I. G., Atkinson, T., Holbrook, J. J., and Clarke, A. R. (1993). The energetics and cooperativity of protein folding—A simple experimental-analysis based upon the solvation of internal residues. *Biochemistry* **32,** 3842–3851.

207. Clark, P. L., Weston, B. F., and Gierasch, L. M. (1998). Probing the folding pathway of a beta-clam protein with single-tryptophan constructs. *Fold. Des.* **3,** 401–412.

208. Clark, P. L., Liu, Z. P., Zhang, J. H., and Gierasch, L. M. (1996). Intrinsic tryptophans of CRABPI as probes of structure and folding. *Protein Sci.* **5,** 1108–1117.

209. Dalessio, P. M., Fromholt, S. E., and Ropson, I. J. (2005). The role of Trp-82 in the folding of intestinal fatty acid binding protein. *Protein Struct. Funct. Bioinform.* **61,** 176–183.

210. Dalessio, P. M., Boyer, J. A., McGettigan, J. L., and Ropson, I. J. (2005). Swapping core residues in homologous proteins swaps folding mechanism. *Biochemistry* **44,** 3082–3090.

211. Yeh, S. R., Ropson, I. J., and Rousseau, D. L. (2001). Hierarchical folding of intestinal fatty acid binding protein. *Biochemistry* **40,** 4205–4210.

212. Chattopadhyay, K., Saffarian, S., Elson, E. L., and Frieden, C. (2002). Measurement of microsecond dynamic motion in the intestinal fatty acid binding protein by using fluorescence correlation spectroscopy. *Proc. Natl. Acad. Sci. USA* **99,** 14171–14176.

213. Scheraga, H. A., Wedemeyer, W. J., and Welker, E. (2001). Bovine pancreatic ribonuclease A: Oxidative and conformational folding studies. *Ribonucleases, Pt A* **341,** 189–221.

214. Narayan, M., Welker, E., Wedemeyer, W. J., and Scheraga, H. A. (2000). Oxidative folding of proteins. *Acc. Chem. Res.* **33,** 805–812.

215. Wedemeyer, W. J., Welker, E., Narayan, M., and Scheraga, H. A. (2000). Disulfide bonds and protein folding. *Biochemistry* **39,** 4207–4216.

216. Goldenberg, D. P. (1992). Native and nonnative intermediates in the Bpti folding pathway. *Trends Biochem. Sci.* **17,** 257–261.

217. Mason, J. M., Gibbs, N., Sessions, R. B., and Clarke, A. R. (2002). The influence of intramolecular bridges on the dynamics of a protein folding reaction. *Biochemistry* **41,** 12093–12099.

218. Mason, J. M., Cliff, M. J., Sessions, R. B., and Clarke, A. R. (2005). Low energy pathways and non-native interactions—The influence of artificial disulfide bridges on the mechanism of folding. *J. Biol. Chem.* **280,** 40494–40499.

219. Kufer, S. K., Dietz, H., Albrecht, C., Blank, K., Kardinal, A., Rief, M., and Gaub, H. E. (2005). Covalent immobilization of recombinant fusion proteins with hAGT for single molecule force spectroscopy. *Eur. Biophys. J.* **35,** 72–78.

220. Best, R. B., Brockwell, D. J., Toca-Herrera, J. L., Blake, A. W., Smith, D. A., Radford, S. E., and Clarke, J. (2003). Force mode atomic force microscopy as a tool for protein folding studies. *Anal. Chim. Acta* **479,** 87–105.
221. Forman, J. R., and Clarke, J. (2007). Mechanical unfolding of proteins: Insights into biology, structure and folding. *Curr. Opin. Struct. Biol.* **17,** 58–66.
222. Fisher, T. E., Marszalek, P. E., and Fernandez, J. M. (2000). Stretching single molecules into novel conformations using the atomic force microscope. *Nat. Struct. Biol.* **7,** 719–724.
223. Fisher, T. E., Oberhauser, A. F., Carrion-Vazquez, M., Marszalek, P. E., and Fernandez, J. M. (1999). The study of protein mechanics with the atomic force microscope. *Trends Biochem. Sci.* **24,** 379–384.
224. Carrion-Vazquez, M., Oberhauser, A. F., Fisher, T. E., Marszalek, P. E., Li, H. B., and Fernandez, J. M. (2000). Mechanical design of proteins-studied by single-molecule force spectroscopy and protein engineering. *Prog. Biophys. Mol. Biol.* **74,** 63–91.
225. Best, R. B., Fowler, S. B., Toca-Herrera, J. L., and Clarke, J. (2002). A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc. Natl. Acad. Sci. USA* **99,** 12143–12148.
226. Ng, S. P., Rounsevell, R. W. S., Steward, A., Geierhaas, C. D., Williams, P. M., Paci, E., and Clarke, J. (2005). Mechanical unfolding of TNfn3: The unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *J. Mol. Biol.* **350,** 776–789.
227. Kim, Y. G., Ho, S. O., Gassman, N. R., Korlann, Y., Landorf, E. V., Collart, F. R., and Weiss, S. (2008). Efficient site-specific labeling of proteins via cysteines. *Bioconjug. Chem.* **19,** 786–791.
228. Jager, M., Nir, E., and Weiss, S. (2006). Site-specific labeling of proteins for single-molecule FRET by combining chemical and enzymatic modification. *Protein Sci.* **15,** 640–646.
229. Jager, M., Michalet, X., and Weiss, S. (2005). Protein–protein interactions as a tool for site-specific labeling of proteins. *Protein Sci.* **14,** 2059–2068.
230. Deniz, A. A. *et al*. (2000). Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA* **97,** 5179–5184.
231. Schuler, B., Lipman, E. A., and Eaton, W. A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419,** 743–747.
232. Kuzmenkina, E. V., Heyes, C. D., and Nienhaus, G. U. (2006). Single-molecule FRET study of denaturant induced unfolding of RNase H. *J. Mol. Biol.* **357,** 313–324.
233. Kuzmenkina, E. V., Heyes, C. D., and Nienhaus, G. U. (2005). Single-molecule Forster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci. USA* **102,** 15471–15476.
234. Groll, J., Amirgoulova, E. V., Ameringer, T., Heyes, C. D., Rocker, C., Nienhaus, G. U., and Moller, M. (2004). Biofunctionalized, ultrathin coatings of cross-linked star-shaped poly (ethylene oxide) allow reversible folding of immobilized proteins. *J. Am. Chem. Soc.* **126,** 4234–4239.
235. Tezuka-Kawakami, T., Gell, C., Brockwell, D. J., Radford, S. E., and Smith, D. A. (2006). Urea-induced unfolding of the immunity protein Im9 monitored by spFRET. *Biophys. J.* **91,** L42–L44.
236. Rhoades, E., Gussakovsky, E., and Haran, G. (2003). Watching proteins fold one molecule at a time. *Proc. Natl. Acad. Sci. USA* **100,** 3197–3202.
237. Huang, F., Sato, S., Sharpe, T. D., Ying, L. M., and Fersht, A. R. (2007). Distinguishing between cooperative and unimodal downhill protein folding. *Proc. Natl. Acad. Sci. USA* **104,** 123–127.
238. Lipman, E. A., Schuler, B., Bakajin, O., and Eaton, W. A. (2003). Single-molecule measurement of protein folding kinetics. *Science* **301,** 1233–1235.

239. Schuler, B., and Eaton, W. A. (2008). Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **18,** 16–26.
240. Hamadani, K. M., and Weiss, S. (2008). Nonequilibrium single molecule protein folding in a coaxial mixer. *Biophysic. J.* **95,** 352–365.
241. Kinoshita, M., Kamagata, K., Maeda, A., Goto, Y., Komatsuzaki, T., and Takahashi, S. (2007). Development of a technique for the investigation of folding dynamics of single proteins for extended time periods. *Proc. Natl. Acad. Sci. USA* **104,** 10453–10458.
242. Chirico, G., Cannone, F., and Diaspro, A. (2006). Unfolding time distribution of GFP by single molecule fluorescence spectroscopy. *Eur. Biophys. J.* **35,** 663–674.
243. Baldini, G., Cannone, F., Chirico, G., Collini, M., Campanini, B., Bettati, S., and Mozzarelli, A. (2007). Evidence of discrete substates and unfolding pathways in green fluorescent protein. *Biophysic. J.* **92,** 1724–1731.
244. Orte, A., Craggs, T. D., White, S. S., Jackson, S. E., and Klenerman, D. (2008). Evidence of an intermediate and parallel pathways in protein unfolding from single-molecule fluorescence. *J. Am. Chem. Soc.* **130,** 7898–7907.
245. Michalet, X., Weiss, S., and Jager, M. (2006). Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem. Rev.* **106,** 1785–1813.
246. Nienhaus, G. U. (2006). Exploring protein structure and dynamics under denaturing conditions by single-molecule FRET analysis. *Macromol. Biosci.* **6,** 907–922.
247. Borgia, A., Williams, P. M., and Clarke, J. (2008). Single-molecule studies of protein folding. *Annu. Rev. Biochem.* **77,** 101–125.
248. Sato, S., Religa, T. L., Daggett, V., and Fersht, A. R. (2004). Testing protein-folding simulations by experiment: B domain of protein A. *Proc. Natl. Acad. Sci. USA* **101,** 6952–6956.
249. Schaeffer, R. D., Fersht, A., and Daggett, V. (2008). Combining experiment and simulation in protein folding: Closing the gap for small model systems. *Curr. Opin. Struct. Biol.* **18,** 4–9.
250. Clementi, C. (2008). Coarse-grained models of protein folding: Toy models or predictive tools? *Curr. Opin. Struct. Biol.* **18,** 10–15.
251. Matysiak, S., and Clementi, C. (2008). Mapping folding energy landscapes with theory and experiment. *Arch. Biochem. Biophys.* **469,** 29–33.
252. Vendruscolo, M., and Paci, E. (2003). Protein folding: Bringing theory and experiment closer together. *Curr. Opin. Struct. Biol.* **13,** 82–87.
253. Improta, S., Krueger, J. K., Gautel, M., Atkinson, R. A., Lefevre, J. F., Moulton, S., Trewhella, J., and Pastore, A. (1998). The assembly of immunoglobulin-like modules in titin: Implications for muscle elasticity. *J. Mol. Biol.* **284,** 761–777.
254. Mallam, A. L., Onuoha, S. C., Grossmann, J. G., and Jackson, S. E. (2008). Knotted fusion proteins reveal unexpected possibilities in protein folding. *Mol. Cell* **30,** 642–648.
255. Went, H. M., Benitez-Cardoza, C. G., and Jackson, S. E. (2004). Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Letters* **567,** 333–338.